

Arguments from Expert Opinion and Persistent Bias

Moti Mizrahi

Florida Institute of Technology

Abstract: Accounts of arguments from expert opinion take it for granted that expert judgments count as (defeasible) evidence for propositions, and so an argument that proceeds from premises about what an expert judges to a conclusion that the expert is probably right is a strong argument. In Mizrahi (2013), I consider a potential justification for this assumption, namely, that expert judgments are significantly more likely to be true than novice judgments, and find it wanting because of empirical evidence suggesting that expert judgments under uncertainty are not significantly more likely to be true than novice judgments or even chance. In this paper, I consider another potential justification for this assumption, namely, that expert judgments are not influenced by the cognitive biases novice judgments are influenced by, and find it wanting, too, because of empirical evidence suggesting that experts are vulnerable to pretty much the same cognitive biases that novices are vulnerable to. If this is correct, then the basic assumption at the core of accounts of arguments from expert opinion, namely, that expert judgments count as (defeasible) evidence for propositions, remains unjustified.

Keywords: arguments from expert opinion; cognitive bias; decision heuristics; expert performance; persistent bias

1. Introduction

In Mizrahi (2013), I argue that arguments from expert opinion are weak arguments. An argument from expert opinion is an argument one makes “under conditions of uncertainty” (Mizrahi 2016, p. 246), i.e., when the truth-value of p is unknown and there is no reason to believe that p is the case other than the fact that an expert judges that p is the case. To appeal to expert opinion is to take the expert’s judgment that p is the case as (defeasible) evidence that (probably) p . For if there is evidence that p is the case independent of the fact that an expert judges that p , then why appeal to expert opinion at all? We can “cut the middleman” and appeal directly to the evidence that supports p (Mizrahi 2013, p. 71). To say that an argument from expert opinion is weak is to say that an expert’s judgment that p does not make it significantly more likely that p is the case. This claim is supported by experimental studies on expert performance. Such studies show that expert performance is often no better than the performance of novices or even chance.

Some argumentation theorists have responded to my argument in Mizrahi (2013), but I think that their responses miss the main question that my argument raises. That is to say, instead of addressing the question of *how to justify the assumption that expert judgments are a reliable source of evidence*, which accounts of arguments from expert opinion simply take for granted, those who responded to Mizrahi (2013) have chosen to focus on my formulation of arguments from expert opinion and accuse me of “refusing to countenance the possibility that other premises of the form of the argument from expert opinion need to be taken into account” (Walton 2014, p. 142),¹ on my examples and accuse me of being a “radical sceptic about

¹ See also Walton (2016), pp. 117-144.

expertise” (Seidel 2014, p. 215),² and even on objections raised against my argument (rather than my own argument) and accuse me of being “confused” (Hinton 2015, p. 542).

However, as I show in Mizrahi (2013, pp. 67-72), virtually all formulations of arguments from expert opinion, including Walton’s (2006, p. 750), take it for granted that an expert’s judgment that *p* counts as (defeasible) evidence for *p*. For example, in Walton et al’s (2008, p. 20) argument scheme for Appeal to Expert Opinion (Version IV), it is the Conditional Premise that captures the basic assumption that expert judgments count as (defeasible) evidence for propositions:

Conditional Premise: If source *E* is an expert in a subject domain *S* containing proposition *A*, and *E* asserts that proposition *A* is true (false), then *A* may plausibly be taken to be true (false).

Similarly, in Wagemans’ (2011, p. 337) “scheme for argumentation from expert opinion,” this basic assumption is stated as follows: “Accepting that *O* is asserted by *E* renders acceptable that *O* is true or acceptable.” No matter how many premises an argumentation scheme for arguments from expert opinion contains, or how many critical questions are added to that argumentation scheme (cf. Walton et al 2008, pp. 14-15), the basic assumption is that expert judgments count as (defeasible) evidence for propositions. After all, that is precisely what an argument from expert opinion is supposed to be; that is, an argument from expert opinion is supposed to be an inference from what an expert judges to be the case to the conclusion that what the expert judges to be the case probably is the case.³ The question that my argument in Mizrahi (2013) argument raises, then, is this: *Why assume that an expert’s judgment that p is (defeasible) evidence for p?* In other words, according to Goodwin (2011, p. 293):

the appeal to expert authority is a blackmail and bond transaction (Goodwin 2001) that brings *the background norm of respect for expertise* to bear in a particular situation. The appeal gives citizens some means to assess the expert and some reason to trust what he says; it also forces them to heed his views. The appeal also requires the expert to take responsibility for his views, putting himself in a position to be held accountable if he turns out to be wrong (emphasis added).⁴

Goodwin (2011, p. 293) calls this “‘he would not risk...so I can trust’ reasoning” and says that it “is widespread in other situations where people are trying to communicate in the face of information asymmetries.” As I point out in Mizrahi (2013, p. 72), however, the fact that this sort of “he would not risk...so I can trust” reasoning is widespread does not make it good reasoning. The tendency to think in terms of stereotypes is also widespread, perhaps even hardwired into the human brain (Cohen 2009, pp. 164-165). But stereotypical thinking is clearly not a good way to reason. So the question is this: why should we respect expertise at all (let alone accept such

² Cf. Jackson (2015), pp. 227-243.

³ See also Linker (2014) and Gelfert (2011) who take the “authority” of expert opinion for granted.

⁴ It is important to note that the kind of authority in question here is what Walton (1992, p. 48) calls “cognitive authority,” i.e., the kind of authority that “looks for the auditor to take what is stated as true and believe it” (Goodwin 1998, pp. 268-269), as opposed to what Walton (1992, p. 48) calls “administrative authority,” i.e., the kind of authority that “looks for the auditor to take what is commanded and make it true” (Goodwin 1998, pp. 268-269). See also Mizrahi (2010). This distinction will become important later on in Section 3.

respect as a “background norm”)? In a recent reply to critics (Mizrahi 2016, p. 241), I put this question as follows:

The question, then, is whether arguments from expert opinion are good *ampliative* arguments, i.e., whether we can gain new knowledge by arguing on the basis of expert opinion (emphasis in original).

In other words, when we appeal to expert opinion, we often need experts to weigh in on questions whose answers are unknown, and we expect experts to be able to provide new answers that are significantly more likely to be correct than not. For instance, we turn to political pundits and analysts for insight into the electoral process. But if it turns out that the judgments of these pundits and analysts are unreliable, as seems to have been the case in the recent Brexit vote in the UK and the recent presidential elections in the United States,⁵ then why trust expert judgments at all? If they are indeed unreliable, then expert judgments cannot be a trustworthy source of evidence or knowledge.

In Mizrahi (2013), I consider the following response to this question concerning the justification for the “background norm of respect for expertise”:

The assumption that an expert’s judgment that *p* is (defeasible) evidence for *p* is warranted only if expert judgments are more reliable (i.e., significantly more likely to be true) than novice judgments.

In other words, we should respect expertise, or treat expert opinion as a trustworthy source of evidence, because experts are significantly more likely to get things right than novices are. As I argue in Mizrahi (2013, pp. 63-65), however, this response is inadequate because expert judgments under uncertainty are not significantly more likely to be true than novice judgments are. Since the question whether expert judgments are reliable or not is an empirical question, I looked at experimental studies aimed at testing expert performance. I (Mizrahi 2013, pp. 63-65) observed that many experimental studies on expert performance show that experts often fail to perform better than novices on cognitive tasks such as decision-making, forecasting, diagnosing, and the like. In fact, expert judgments under uncertainty are often no more likely to be true than false, which means that, statistically speaking, they are not significantly better than guessing. Based on such empirical evidence derived from experimental studies on expert performance, in Mizrahi (2013, pp. 58-59), I argue as follows:

1. An expert’s judgment that *p* is (defeasible) evidence for *p* only if expert judgments are reliable.
 2. It is not the case that expert judgments are reliable.
- Therefore,
3. It is not the case that an expert’s judgment that *p* is (defeasible) evidence for *p*.

⁵ See, e.g., this article in *The Guardian* on how “Experts get it wrong again by failing to predict Trump victory” (Blanchflower 2016): <https://www.theguardian.com/business/2016/nov/09/experts-trump-victory-economic-political-forecasters-recession>.

If cogent, this argument shows that the “background norm of respect for expertise” cannot be justified by claiming that expert judgments are reliable (either significantly more likely to be true than novice judgments or significantly more likely to be true than false). The first premise of this argument states that reliability is a *necessary condition* for being a trustworthy source of evidence, hence the “only if.” Being unreliable disqualifies a source from being a trustworthy source of evidence. As Devitt (2015, p. 687) puts the point with respect to intuitive judgments:

intuitions are a source of evidence *because they are reliable*. And if they really are reliable, which does not of course require them to be infallible, then they are indeed a source of evidence. (And, one might add, this is true of *any* judgment, whether intuitive or not.) (emphasis in original).

In other words, only reliable sources can be trustworthy sources of evidence. To use my own example from Mizrahi (2013, p. 65):

Would you trust a watch that gets the time right 55% of the time? Would you trust a thermometer that gets the temperature right 55% of the time? I suspect the answer to these questions is “no.” Similarly, a method of reasoning, such as appealing to expert opinion, is trustworthy only if expert opinion is [reliable].

Accordingly, if there were evidence that expert judgments are reliable, either significantly more likely to be true than false or at least significantly more likely to be true than novice judgments are, then that would have been a strong reason to think that expert judgments are reliable, and hence that they can be a trustworthy source of (defeasible) evidence for the truth of such judgments. Since there is no evidence that expert judgments are reliable (in fact, there is evidence to the contrary, i.e., that expert judgments are not more reliable than novice judgments), there is no reason to think that expert judgments are a trustworthy source of (defeasible) evidence for the truth of such judgments.

If this is correct, then the justification for the “background norm of respect for expertise” cannot be that expert judgments are reliable, since experimental studies suggest that expert judgments are not more reliable than novice judgments. In fact, there is empirical evidence that expert judgments are often not significantly more likely to be true than false.

In this paper, I will consider another potential response to the question concerning the justification for the “background norm of respect for expertise.” This response goes as follows:

The assumption that an expert’s judgment that p is (defeasible) evidence for p is warranted only if experts are not susceptible to the kinds of cognitive biases that novices are susceptible to.

In other words, we should respect expertise, or treat expert opinion as a trustworthy source of evidence, because, unlike novices, experts are immune to cognitive biases, such as confirmation bias, framing effects, order effects, and the like. Since the question whether or not experts are immune to cognitive biases is an empirical question, I will look at experimental studies aimed at testing for cognitive biases among experts and novices. I will argue that this response, too, is

inadequate, since experimental studies on cognitive biases show that experts are vulnerable to pretty much the same cognitive biases that novices are vulnerable to. If this is correct, then the “background norm of respect for expertise” cannot be justified by claiming that experts are immune to cognitive biases, and thus this basic assumption at the core of arguments from expert opinion, according to which expert judgments are a trustworthy source of evidence, remains unjustified.

2. Are experts immune to cognitive biases?

For present purposes, I take it as uncontroversial that, from an epistemic point of view, we should not trust or respect sources of evidence that are not reliable. My critics would probably agree with that. For example, Seidel (2014, p. 195) writes that he has “no objection to this latter claim [namely, that “arguments from expert opinion are weak arguments unless the fact that E says that p makes it significantly more likely that p is true” (Mizrahi 2013, p. 58)] since it just is a formulation of the close connection between the reliability and the epistemic trustworthiness of an epistemic source.” In Mizrahi (2013), I cite experimental studies that provide empirical evidence against the claim that expert judgments are reliable (either significantly more likely to be true than novice judgments or significantly more likely to be true than false).

The aforementioned empirical evidence notwithstanding, some might still think that we should respect expertise, or treat expert opinion as a trustworthy source of evidence, because, unlike the judgments of novices, the judgments of experts are not subject to cognitive biases. In other words, if expert judgments are *unbiased* judgments, then they should be trusted. For example, confirmation bias is “a tendency to search out and pay special attention to information that supports one’s beliefs, while ignoring information that contradicts a belief” (Goodwin 2010, p. 8). If we were to find out that, unlike novices, experts are immune to confirmation bias, then that would provide a strong reason to think that expert judgments are reliable, and thus that we should respect expertise. The question, then, is whether or not experts are as susceptible to confirmation bias (as well as other cognitive biases) as novices are. Since this is an empirical question, we should look at experimental studies aimed at testing for cognitive biases among experts and novices.

Here are a few examples of experimental findings about the cognitive biases novices and experts are susceptible to in a variety of domains:

- **Driving:** Waylen et al (2004, p. 323) “compared expert police drivers with novice police drivers,” and found that “[d]espite their extensive additional training and experience, experts still appear to be as susceptible to illusions of superiority as everyone else.”
- **Law:** In a series of experimental studies, Guthrie, Rachlinski, and Wistrich (2001, 2007, 2009) “demonstrated that anchoring, hindsight, framing, and other documented biases influence the way in which judges [even “judges who specialize in a specific area of law”] analyze legal vignettes” (Teichman and Zamir 2014, p. 691).
- **Medicine:** The results of several experimental studies (e.g., McNeil et al 1982) show that decisions made by patients, medical students, and physicians are all subject to framing

effects (e.g., whether the same treatment is presented as a potential loss or as a potential gain).⁶

- **Philosophy:** A growing body of experimental work suggests that, just like the judgments of non-philosophers, the judgments of professional philosophers are influenced by cognitive biases. For instance, Schwitzgebel and Cushman (2012) show that the moral judgments of professional philosophers, just like the moral judgments of non-philosophers, are affected by the order in which hypothetical scenarios are presented. In a follow-up study, Schwitzgebel and Cushman (2015) show that framing and order effects persist despite high levels of academic expertise and familiarity with the hypothetical scenarios (e.g., the trolley problem) that were presented to the professional philosophers who participated in the study.
- **Software engineering:** Leventahl et al (1994) found that expert programmers are susceptible to confirmation bias, or what they call “positive test bias,” just as novices are. The results of their experimental study show evidence of positive test bias, i.e., of the tendency of software testers to test software against data that supports their judgments about that software regardless of the subjects’ level of expertise. (See also Calikli and Bener 2015.)

Accordingly, the results of many experimental studies do not support the claim that experts are not susceptible to the kinds of cognitive biases that novices are susceptible to. In fact, such experimental studies provide empirical evidence to the contrary; that is, they show that experts are susceptible to cognitive biases just like everyone else. As Eisenstein and Lodish (2002, p. 437) put it, the results of experimental studies like those mentioned above suggest that “[e]xperts can fall prey to the same array of cognitive biases that affect novices, resulting in sub-optimal performance and unreliability.”

In addition to the aforementioned cognitive biases, there is empirical evidence suggesting that there are several decision heuristics that both experts and novices tend to use. These unreliable decision heuristics have the following characteristics:

First, the errors in judgments attributable to the heuristic are systematic and directional; that is, they always act in the same way and in the same direction. Second, they are general and nontransferable; that is, *all humans are susceptible to the errors and knowledge of how they act does not immunize us against them*. Third, they are independent of intelligence and education; that is, *experts make the same mistakes as novices* (Lash et al 2009, p. 6; emphasis added).

Two of these decision heuristics are:

- **Anchoring and adjustment:** a decision-making process in which “people make estimates by starting from an initial value that is adjusted to yield a final answer” (Tversky and Kahneman 1974, p. 1128). These “adjustments are typically insufficient” (Tversky and Kahneman 1974, p. 1128). For example, Northcraft and Neale (1987) conducted a study whose results show that both novices (undergraduate students) and

⁶ For more on cognitive biases in clinical decision-making, see Groopman (2007).

experts (real estate agents) use anchoring and adjustment in information-rich, real-world settings (as opposed to a laboratory setting).

- **Overconfidence:** “people are often more confident in their judgments than is warranted by the facts” (Griffin and Tversky 1992, p. 411). Numerous experimental studies show evidence of overconfidence in expert judgments of physicians (Lusted 1977), clinical psychologists (Oskamp 1965), lawyers (Wagenaar and Keren 1986), negotiators (Neale and Bazerman 1991), engineers (Kidd 1970), and security analysts (Staël von Holstein 1972). As Griffin and Tversky (1992, p. 412) report: “As one critic described expert prediction, ‘often wrong but rarely in doubt’.”

In a study that looked at both anchoring and overconfidence, English et al (2005) show that, just like novices, criminal law experts are susceptible to anchoring but tend to be very confident about their biased judgments. As English et al (2005, p. 718) put it, the results of their experimental study suggest that “expertise does not protect against assimilative sentencing bias.” More generally, “experts, like lay citizens, tend to draw on various heuristics—even when such usage does not fit context” (Perez 2015, p. 119).

To avoid confirmation bias, and to make sure that the aforementioned studies on cognitive biases among experts and novices are representative, I have conducted a more systematic review of the literature on cognitive biases among experts and novices. In particular, I have searched the Web of Science for articles containing the terms ‘expert’, ‘novice’, ‘bias’, and ‘judgment’. This search turned up 21 results. From the 21 search results, I have excluded studies that are not specifically about cognitive biases among experts and novices. Four articles did not meet this selection criterion. The first article presents the results of a psychometric (as opposed to a “bare-bone”) meta-analysis of judgmental achievement (i.e., accuracy of judgments) and decision-making. The “results indicated that students reached a slightly higher judgmental achievement than experts” (Kaufmann 2013, p. 13). After “separating expertise within domains,” and applying the psychometric method, “judgmental achievement among business experts dropped to a low value,” and “experts’ judgments in other research domains decreased [as well]” (Kaufmann 2013, p. 13). The second article discusses the use of decision heuristics and diagnostic aids by organizational consultants, but does not provide new empirical evidence on cognitive biases among experts and novices (Armenakis et al 1990). The third article presents the results of a study of novice performance in evaluating e-learning environments. The results show that novice evaluators perform better when they use pattern-based inspection (Lanzilotti 2011). The fourth article presents the results of a study on gender recognition. The results show that “conductors were more often judged to be male” (Wöllner 2013, p. 84).

Of the remaining search results (17), the following studies report that expert judgments exhibit cognitive biases that novice judgments do not:

- Experts are susceptible to “*the curse of knowledge*, a bias whereby knowledgeable people are unable to ignore information they hold that others do not” (Hinds 1999, p. 206; emphasis in original).

The following studies report that expert and novice judgments exhibit similar cognitive biases but to varying degrees or in different directions:

- “Perceptual judgments of aesthetics in gymnastics appear to be subject to specific memory influences regardless of the judge’s expertise” (Ste-Marie and Lee 1991, p. 129).
- “No significant difference between the magnitude of experts’ [experienced police drivers] and novices’ [inexperienced police drivers] aggregate biases [namely, self-enhancement bias and illusions of superiority] were found” (Waylen et al 2004, p. 329).
- “[N]on-experts were more likely to exhibit focal brand favorableness bias than experts” (Posavac et al 2005, p. 93).
- “[T]o the extent that framing bias occurs, the bias seems to occur in both expert clinicians and novice students” (Brailey et al 2001, p. 74); “there was no evidence that students were more likely than clinicians to exhibit a decision-consistent framing bias” (Brailey et al 2001, p. 74); “expert clinicians are less affected by confirmatory bias than students” (Brailey et al 2001, p. 75).
- “[I]rrespective of the level of expertise in teaching, classroom teachers and university students misjudged a tutee’s understanding” (Herppich et al 2013, p. 256).
- Art experts are not “immune from verbalization bias [i.e., “an increase in the evaluative significance of easily verbalized stimulus attributes brought about by the evaluator’s efforts to verbally encode a stimulus” (McGlone 2005, p. 242)] in their evaluation of the abstract works [of art]” (McGlone 2005, p. 249), but “experts were better able to articulate the reasons for their judgments than novices” (McGlone 2005, pp. 250-251).
- The judgments of “both novices and experts were affected [by self-assessment]; it was little effected by warning that self-assessment is unreliable” (Chen and Kemp 2012, p. 587).
- “[T]he salesperson selections made by novice and expert sales managers were equally biased, albeit in different directions, with novices outweighing and expert underweighting historical performance trends” (DeCarlo et al 2015, p. 1484).

The following studies report a tendency among either experts or novices to make a particular judgment rather than another as a “response bias”:

- “[G]oalkeepers were significantly biased towards deception decisions, whereas neither the field players nor the novices showed this bias” (Canal-Bruland and Schmidt 2009, p. 239).
- “A contextual bias effect was observed for novice and expert participants when comparing and assessing fingerprints. That bias effect was stronger for novice participants. Experts were more resistant to bias suggestions towards individualization and less so to suggestions towards inconclusive and exclusion” (Langenburg et al 2009, p. 581).
- “Reading the text [on baseball] reduced confidence bias scores more for the higher knowledge than for the lower knowledge participants” (Griffin 2009, p. 1008).
- “[T]he law enforcement group [compared to a group of students] showed a bias to respond ‘lie’ and not greater discrimination of message veracity [in deception detection]” (Bond 2008, p. 343).
- “[B]oth experts’ and novices’ judgments are biased in that they assign excessively heavier weight to more pathological information and that this bias is stronger among experts than among novices” (Ganzach 1997, pp. 956-957).

- Unlike amateur soccer players (the novices), “experts [i.e., assistant referees] showed a higher response bias toward not responding to the stimulus. [...] [I]t appears that with years of practice and increased task-specific expertise, international assistant referees have biased their response criterion into a more conservative direction” (Put 2013, p. 581).⁷
- Among groups of Norwegian detectives, Norwegian novices, English detectives, and English novices, “all groups, except the English [detectives], were clearly biased toward investigative hypotheses involving the abduction or killing of the girl by family members. [...] Although the identity of the body had not been established, all groups, including the English [detectives], displayed a clear bias toward investigative hypotheses involving murder” (Fahsing and Ask 2016, p. 215).
- “[W]hen recalling locations in scenes of geological relevance, geology experts’ conceptual knowledge resulted in different patterns of bias compared to experts from unrelated fields” (Holden et al 2016, p. 450).

Of the 17 search results, none report that expert judgments are immune to cognitive biases or that expertise provides protection against cognitive biases. However, as mentioned above, one study did find that “expert clinicians are *less* affected by confirmatory bias than students” (Brailey et al 2001, p. 75; emphasis added). So I have looked beyond the sample of articles produced by my search through the Web of Science for additional studies on cognitive biases among experts and novices that report similar results. I did manage to find one study whose results could be construed as showing that expertise might be able to provide at least some measure of protection against confirmation bias. Krems and Zierer (1994) argue that “high-domain knowledge experts have *less* of a confirmation bias than intermediates or novices” (emphasis added). This means that, like intermediates and novices, experts are still susceptible to confirmation bias, but not as susceptible as intermediates and novices are. The results of this study are taken to show that expert are less susceptible to confirmation bias than novices are because the latter take more time to correct erroneous assumptions than the former (cf. Mendel et al 2011).⁸ Accordingly, the judgments of experts are still biased, but experts correct their biased decisions faster than novices do.

⁷ Assistant referees are not better at making offside decisions than amateur soccer players are, but the former do take more time to make their decisions than the latter do. The researchers think that this result shows that assistant referees make offside decisions on a cognitive level, as opposed to a merely perceptual level (Put et al 2013). In general, however, studies show that “[i]ntelligence and experience do not make a person immune to cognitive biases” (Cooper and Frain 2017, p. 26). Accordingly, even if the fact that assistant referees take more time to make offside decisions than amateur soccer players do suggests that they think about it more carefully, it doesn’t necessarily follow that they are immune to *cognitive* illusions, as opposed to *perceptual* illusions, which is what this study focused on.

⁸ An alternative hypothesis is that it is not expertise *per se* that makes (medical) experts less susceptible to confirmation bias than novices but rather awareness of the problem of cognitive bias itself. In a feature article in *Nature*, Nuzzo (2015) quotes Robert MacCoun who says that “When crises like this issue of reproducibility come along, it’s a good opportunity to advance our scientific tools.” For example, “when scientists in the mid-twentieth century realized that experimenters and subjects often unconsciously changed their behaviour to match expectations” [...] the double-blind standard was born” (p. 183). If this is correct, then it’s not being an expert in a particular field of study that makes one less susceptible to confirmation bias, but rather it’s knowing about confirmation biases themselves and how they can affect one’s judgments.

To sum up, then, most of the empirical evidence on cognitive biases among experts and novices suggests that “expert judgment is subject to various biases, such as the hindsight bias,” to varying degrees, and that “experts are not immune to these biases” (Perez 2015, p. 119). In the words of Kahneman (1991, p. 144), “there is much evidence that experts are not immune to the cognitive illusions that affect other people.” If we should not trust *biased* judgments, and expert judgments are *biased* judgments, much like novice judgments are, then it follows that we should not trust expert judgments.

3. Objections and replies

I have argued that the “background norm of respect for expertise,” i.e., the assumption that an expert’s judgment that p is (defeasible) evidence for p , cannot be justified by claiming that experts are immune to the sort of cognitive biases that novices are susceptible to. This is because empirical evidence suggests that experts are susceptible to cognitive biases just like everyone else. My overall argument can be stated as follows:

1. An expert’s judgment that p is (defeasible) evidence for p only if experts are not susceptible to the sort of cognitive biases that novices are susceptible to.
2. It is not the case that experts are not susceptible to the sort of cognitive biases that novices are susceptible to.

Therefore,

3. It is not the case that an expert’s judgment that p is (defeasible) evidence for p .

If cogent, this argument shows that the “background norm of respect for expertise” cannot be justified by claiming that experts are immune to cognitive biases. In other words, if there were evidence that experts are immune to the sort of cognitive biases that novices are susceptible to, then that would have been a strong reason to think that expert judgments are reliable, and hence that they can be a trustworthy source of (defeasible) evidence for the truth of such judgments. Since there is no evidence that experts are immune to the sort of cognitive biases that novices are susceptible to (in fact, there is empirical evidence to the contrary; that is, that experts are vulnerable to pretty much the same cognitive biases as novices), there is no reason to think that expert judgments are a trustworthy source of (defeasible) evidence for the truth of such judgments.

If this is correct, then the justification for the “background norm of respect for expertise” cannot be that experts are immune to the sort of cognitive biases that novices are susceptible to, since experimental studies on cognitive biases provide empirical evidence that expert judgments are susceptible to pretty much the same cognitive biases that novice judgments are susceptible to. Given that it is also not the case that expert judgments are significantly more likely to be true than novice judgments or significantly more likely to be true than false, as I argue in Mizrahi (2013), the justification for the “background norm of respect for expertise” cannot be that expert opinion is reliable (either significantly more likely to be true than novice opinion or significantly more likely to be true than false). If this is correct, then the “background norm of respect for expertise,” the basic assumption at the core of arguments from expert opinion according to which expert judgments are a trustworthy source of evidence, remains unjustified. In this section, I will consider a couple of objections to my overall argument.

Before I do so, however, it is important to acknowledge the limits of my argument. In particular, my argument concerns what Walton (1992, p. 48) calls “cognitive authority,” i.e., “authority based on [professed] knowledge (‘cognitive’, ‘epistemic’)” (Walton 2016, p. 2), as opposed to what Walton (1992, p. 48) calls “administrative authority,” i.e., “authority based on directives (‘administrative’, ‘executive’, ‘deontic’)” (Walton 2016, p. 2). Similarly, expertise can be thought of in terms of “what experts [claim to] *know* that others do not” (Phillips 2008, p. 300; emphasis in original), or in terms of “what experts *can do* that others cannot” (Phillips 2008, p. 300; emphasis in original).⁹ The empirical evidence concerning expert susceptibility to cognitive biases, then, pertains to experts *qua* cognitive or epistemic authorities, not to experts *qua* administrative or executive authorities. From an *epistemic* point of view, a source of evidence is trustworthy just in case it is *reliable* (Devitt 2015, p. 687; Mizrahi 2013, p. 65). In particular, judgments or opinions should be trusted only if they are reliable. From an *epistemic* point of view, if expert judgments are unreliable, they should not be trusted. This argument is about what experts *say* or *judge*, not about what experts *can do*. Accordingly, it could be the case that experts are better than novices are at doing things other than *expressing true opinions* or *making unbiased judgments*. For instance, evidence suggests that experts process visual information faster than novices do (Gegenfurtner et al 2011). However, speed does not entail epistemic reliability (either in terms of getting things right more often than not or in terms of being free from cognitive biases). From an epistemic point of view, then, we need good reasons to think that expertise guarantees epistemic reliability, especially since empirical evidence suggests that “[i]ntelligence and experience do not make a person immune to cognitive biases” (Cooper and Frain 2017, p. 26).

Since critics have claimed that my argument in Mizrahi (2013) is “self-undermining” (Seidel 2014, p. 210), some might be tempted to level a similar charge against my argument here. So, in response, it is important to note that the warrant for the claim that experts are not immune to the cognitive biases that affect all of us is *not* that experts on cognition say so. Rather, the warrant for the claim that experts are not immune to the cognitive biases that affect all of us is *empirical evidence* gathered from experimental studies on cognitive biases among experts and novices. There is a clear difference between accepting a claim because it is judged to be true by an expert and accepting a claim because empirical evidence supports it. In other words, an argument that proceeds from a premise about an expert’s judgment that *p* to the conclusion that *p* is true or probable is clearly different from an argument that proceeds from a premise about an experimental procedure that yields result *r* to the conclusion that *r* is true or probable. While the former is an appeal to a person’s *judgment*, the latter is an appeal to a *procedure* (cf. Mizrahi 2016, p. 248). Indeed, in Walton’s (2016, pp. 130-131) argumentation scheme for arguments from expert opinion, an expert is taken to be a source of evidence (the “major premise”) but there is a critical question about whether or not the expert’s assertion is based on *evidence* (the “backup evidence question”).¹⁰ This is why I argue in Mizrahi (2013, pp. 67-72) that Walton’s argumentation scheme for arguments from expert opinion faces a dilemma: if there is evidence for an expert’s assertion from a source that is more reliable than expert opinion, why rely on expert opinion at all?

⁹ See also Anderson (1983) for a distinction between “declarative knowledge” and “procedural knowledge,” and Ryle (1946) for a distinction between “knowing that” and “knowing how.”

¹⁰ Cf. Wagemans (2011, p. 337): “E is able to provide further evidence for O.”

Of course, to say that “there is a clear difference between the inference from ‘Expert *E* judges that *p*’ to ‘*p*’ (where *p* is an opinion or judgment whose truth value is unknown and the only reason to accept *p* is that *E* asserts it) and the inference from ‘Experiment *X* yields the result that *r*’ to ‘*r*’” (Mizrahi 2016, p. 248) is *not* to say that scientific investigation is devoid of human judgment. Historians (e.g., Kuhn 2000), philosophers (e.g., Douglas 2009), and sociologists of science (e.g., Latour and Woolgar 1986) have pointed out that both epistemic and non-epistemic judgments play an important role in scientific practice. Such judgments, however, have to be “trained judgments,” to use Daston and Galison’s terminology, which are judgments made by scientists who are already working through “mechanical objectivity,” that is, scientists engaging in the pursuit of knowledge “that bears no trace of the knower” (Daston and Galison 2017, p. 17). For example, blinding or masking procedures are designed to eliminate any “trace of the knower” by guaranteeing that the experimenters or the participants (single-blind), or both (double-blind), are unaware of the treatment assignment in a clinical study. In this way, blinding is designed to minimize biases, such as assessment bias and observation bias (Bacchieri and Della Cioppa 2007, pp. 216-217). Indeed, interventions such as blinding are needed in research precisely because “[e]xperts are not immune to these cognitive vulnerabilities and hence often exhibit bias in their conclusions, but might be unaware of it” (MacLean and Dror 2016, p. 13).

The origin of this idea of scientific objectivity can be traced back to the motto of the Royal Society of London: “*Nullius in verba*” or “take nobody’s word for it” (<https://royalsociety.org/about-us/history/>). In adopting this motto, the Fellows of the Royal Society did not mean to suggest that they do not trust *each other* as working scientists (or, more precisely, natural philosophers). Rather, this motto “is an expression of the determination of Fellows *to withstand the domination of authority* and to verify all statements by an appeal to facts determined by experiment” (<https://royalsociety.org/about-us/history/>, emphasis added). In other words, the Fellows’ motto is meant to suggest that they do not trust *testimony* as a source of evidence. The Fellows of the Royal Society probably picked up this idea from Locke who argued “both that other people are a highly unreliable source of information and that, even when they speak truthfully, one cannot gain true knowledge merely by taking someone else’s word” (Huemer 2002, p. 217). As Locke (1975, p. 101) writes:

The floating of other Mens Opinions in our brains makes us not one jot the more knowing, though they happen to be true. What in them was Science, is in us but Opiniatrey, whilst we give up our Assent only to revered Names, and do not, as they did, employ our own Reason to *understand* those Truths, which gave them reputation (emphasis in original).

For Locke, then, as well as for the Fellows of the Royal Society, *testimony* is not a trustworthy source of evidence.¹¹

For similar reasons, in Mizrahi (2016, p. 242), I distinguish “between *epistemic trust in expertise*, i.e., trusting that *p* is true because *E* judges that *p* is true, and what we might call *professional trust*, i.e., trusting that *E* is a professional.” Scientists can trust one another *professionally*, i.e., they trust that their fellow scientists are professionals who can be trusted to

¹¹ For more on Locke on testimony, see Shieber (2009).

follow protocols of experimentation, randomization, blinding, and the like. But that does not mean that they trust each other *epistemically*, i.e., that they “take each other’s word” without question. The same distinction applies outside of the domain of science as well. For instance, to use my example in Mizrahi (2016, p. 242), “I may trust my physician insofar as she is a professional who knows what she is doing (e.g., she knows how to perform physical examinations and make clinical decisions) but I may still mistrust her diagnosis in a particular case and seek a second opinion.” This is why, when our physicians misdiagnose, as they often do (Makary 2016), we do not say that they are not professional doctors; they are still professionals, even when they get things wrong. But if they do get things wrong more often than not, or if their judgments are biased, then their expert judgments are not reliable, and thus not a trustworthy source of evidence or knowledge.

I think that these remarks should also help in terms of pre-empting another objection, namely, that we have to appeal to expert opinion; it is an indispensable source of evidence. After all, the objection goes, what else is there? In reply, I would like to make two points. First, even if it were true that we have to appeal to expert opinion, and that it is an indispensable source of evidence, it would still not follow that expert opinion is a *reliable* or *trustworthy* source of evidence. From the fact that the only tool I have at my disposal is a hammer it does not follow that a hammer is a good tool for fixing faulty smartphones. Likewise, even if it is true that we have to appeal to expert opinion, it does not follow that appealing to expert opinion is rationally justified. I take it that, as argumentation theorists, we should not be content with claiming that expert opinion is an indispensable source of evidence. Rather, we should try to *justify* it, i.e., to show that expert opinion is a trustworthy source of evidence. The question, then, is how the “background norm of respect for expertise” (i.e., that an expert’s judgment that *p* is defensible evidence for *p*) can be justified.

Second, it is in fact *not* true that we have to appeal to expert opinion. Some alternatives to expert opinion include algorithms, Artificial Intelligence (AI), machine learning, and neural networks. In medicine, for instance, where “medical error is the third most common cause of death in the US” (Makary 2016), algorithms are used as decision aids for the purposes of diagnosis and treatment. According to the *Encyclopedia of Medical Decision Making*:

Algorithms are branching-logic pathways that permit the application of carefully defined criteria to the task of identifying or classifying different types of the same entity. Clinical algorithms are often represented as schematic models or flow diagrams of the clinical decision pathway described in a guideline (p. 135).

For example, Ramos et al (2016) present evidence that an ICU admission triage algorithm demonstrated higher interrater reliability than physicians’ judgments and Craig et al (2010, p. 11) developed a “computerized diagnostic model” that “outperformed clinical judgment for the diagnosis of fever in young children.” Such clinical algorithms are now becoming more widely used in medicine (Winters-Miner et al 2014).

At this point, some might be tempted to retort that appealing to expert opinion is still indispensable for it is experts who design and “train” the algorithms in the first place. But this retort fails to take into consideration the types of algorithms, and other decision aids, that are

currently available. Keeping with the theme of examples from the field of medicine, AI and machine learning technology can assist medical professionals in their clinical decision-making. As Kononenko (2001, p. 90) points out:

Machine learning technology is currently well suited for analyzing medical data, and in particular there is a lot of work done in medical diagnosis in small specialized diagnostic problems.

One of the prominent examples of the application of machine learning in medicine is IBM's Watson. Watson's capabilities include natural language processing, hypothesis generation and evaluation, and evidence-based learning, which are put to use in helping medical professional make clinical decisions, such as diagnosis and treatment for cancer patients (IBM 2013).

In addition to AIs like IBM's Watson, there are also neural networks that can learn from experience to the point that they outperform human experts on a particular cognitive task. For example, a deep neural network model called "Watch, Listen, Attend, and Spell (WLAS)," which was developed by researchers from Google and Oxford University, can transcribe videos better than a professional lip reader (Chung et al 2016). Another AI machine that surpasses human experts in pattern recognition tasks is Giraffe, "a chess engine that uses self-play to discover all its domain-specific knowledge, with minimal hand-crafted knowledge given by the programmer" (Lai 2015, p. 2). Ever since IBM's Deep Blue beat the world chess champion, Gary Kasparov, in 1997, machines have become much better than professional chess players, to the point where chess players nowadays "prefer to stay away from computers as opponents" (Siegel 2016). Unlike other chess playing machines, however, Giraffe learns to play chess by evaluating moves rather than calculating every possible move (Lai 2015, p. 16).¹²

The point, then, is that appealing to expert opinion is not indispensable; there are alternative sources of evidence that one could appeal to. Of course, there may be *pragmatic* reasons why we might be wary of AI (see, e.g., Bostrom 2014). From an *epistemic* point of view, however, a source of evidence is trustworthy just in case it is *reliable*. If it turns out that the outputs of neural networks are more reliable (either more likely to be true than false or immune to cognitive biases) than the judgments of human experts, then from a purely epistemic point of view, we should prefer the former to the latter.

In that respect, it is also important to note that conceding that expert opinion is not a reliable source of evidence does not mean that "anything goes" and that people should simply make up their own minds without appealing to any source of evidence. Again, there are sources of evidence other than expert opinion as well as a variety of decision aids, such as algorithms and neural networks. Indeed, given the limitations of our sensory perception, we generally need instruments, such as microscopes and telescopes, to help us in our epistemic pursuits. As Hooke (*Micrographia*, Preface) puts it, scientific instruments, like his microscope, are needed for the "inlargement of the dominion of the senses."¹³ Given our cognitive limitations, then, it should not be surprising that we also need instruments, such as algorithms, to help us in our epistemic pursuits. In fact, just as non-experts may need assistance from decision aids, such as the

¹² For more on neural networks outperforming human players in games, see Mnih et al (2013).

¹³ For more on scientific instruments as "enlarging the empire of the senses," see Shapin and Schaffer (1985), Ch. 2.

Carneades Argumentation System (CAS), to evaluate expert opinion (Walton 2016), experts may need assistance from decision aids, such as algorithms and neural networks, to make more reliable, and less biased, judgments.

4. Conclusion

In this paper, I have argued that the justification for the “background norm of respect for expertise” cannot be that experts are immune to the sort of cognitive biases that novices are susceptible to. This is because experimental studies on cognitive biases provide empirical evidence that expert judgments under uncertainty are susceptible to pretty much the same cognitive biases that novice judgments are susceptible to. Given that it is also not the case that expert judgments are significantly more likely to be true than novice judgments or significantly more likely to be true than false (Mizrahi 2013), the justification for the “background norm of respect for expertise” cannot be that expert opinion is reliable (either significantly more likely to be true than novice judgments or significantly more likely to be true than false).

If this is correct, then the basic assumption at the core of accounts of arguments from expert opinion, according to which expert judgments count as (defeasible) evidence for propositions, remains unjustified. And so the question I have raised in Mizrahi (2013) remains unanswered: what is the justification for the “background norm of respect for expertise,” i.e., the assumption that an expert’s judgment that p is (defeasible) evidence for p ?

Some might think that the argument I have advanced in this paper is mostly “negative,” and that it would be ideal if I could offer a “positive” proposal in the form of a new argumentation scheme for arguments from expert opinion. I am afraid that I do not know how to justify the assumption that expert opinion counts as (defeasible) evidence other than to show that expert opinion is reliable. I think Devitt (2015, p. 687) is right in saying that judgments count as a source of evidence only if they are reliable. (See also Mizrahi 2013, p. 65.) As we have seen, however, there is empirical evidence suggesting that expert opinion is not reliable, both in terms of not being significantly more likely to be true than novice opinion (Mizrahi 2013) and in terms of not being immune to the cognitive biases that affect novice opinion. Since being unreliable disqualifies a source from being a trustworthy source of evidence, at the very least we should suspend judgment about the evidential status of expert opinion until new evidence concerning the reliability of expert opinion comes to light.

However, the brief remarks I have made at the end of Section 3 do point to what I think might be a promising way forward. As I have suggested above, our current epistemic enterprises may be such that they cannot be adequately pursued without the aid of instruments (e.g., telescopes, microscopes, MRI, etc.) that enhance the dominion of perception. Similarly, I submit, perhaps we have reached a point where our epistemic enterprises are such that they cannot be adequately pursued without the aid of instruments that enhance the dominion of cognition. If this is correct, then a promising way forward might be to think of algorithms, neural networks, and other decision aids, not as *replacing* expert opinion, but as *enhancing* expert performance. If it turns out that expert opinions aided by algorithms, neural networks, and other decision aids are more reliable than unaided expert opinion, then, from an epistemic point of view, we should prefer the former to the latter. Just as unaided observations of celestial bodies would not be

trusted nowadays,¹⁴ *unaided* expert opinions should not be trusted as well. If this is correct, then it opens up new avenues of research for argumentation theorists to investigate the ways in which expert opinion can be aided, and by what means, such that it is more reliable and less biased, and how arguments from expert opinion can thereby be made stronger.

Acknowledgments

I am grateful to two anonymous reviewers of *Argumentation* for helpful comments on an earlier draft of this paper.

References

- Anderson, J. R. (1983). *The Architecture of Cognition*. Cambridge, MA: Harvard University Press.
- Armenakis, A., Mossholder, K. W., and Harris, S. G. (1990). Diagnostic bias in organizational consultation. *Omega: International Journal of Management Science* 18, 563-572.
- Bacchieri, A., and Della Cioppa, G. (2007). *Fundamentals of Clinical Research: Bridging Medicine, Statistics and Operations*. Milano: Springer.
- Blanchflower, D. (2016). Experts get it wrong again by failing to predict Trump victory. *The Guardian*, 9 November 2016. <https://www.theguardian.com/business/2016/nov/09/experts-trump-victory-economic-political-forecasters-recession>.
- Bond, G. D. (2008). Deception detection expertise. *Law and Human Behavior* 32, 339-351.
- Bostrom, N. (2014). *Superintelligence: Paths, Dangers, Strategies*. New York: Oxford University Press.
- Brailey, K., Vasterling, J. J., and Franks, J. J. (2001). Memory of psychodiagnostic information: Biases and effects of expertise. *American Journal of Psychology* 114, 55-92.
- Calikli, G., and Bener, A. (2015). Empirical analysis of factors affecting confirmation bias levels of software engineers. *Software Quality Journal* 23, 695-722.
- Canal-Bruland, R., and Schmidt, M. (2009). Response bias in judging deceptive movements. *Acta Psychologica* 130, 235-240.
- Chen, Z., and Kemp, S. (2012). Lie hard: The effect of self-assessments on academic promotion decisions. *Journal of Economic Psychology* 33, 578-589.
- Chung, J. S., Senior, A., Vinyals, O., and Zisserman, A. (2016). Lip reading sentences in the wild. [arXiv:1611.05358](https://arxiv.org/abs/1611.05358) [cs.CV].
- Cohen, E. D. (2009). *Critical Thinking Unleashed*. Lanham, MD: Rowman & Littlefield.
- Cooper, N., and Frain, J. (2017). *ABC of Clinical Reasoning*. Hoboken, NJ: Wiley Blackwell.
- Craig, J. C., Williams, G. J., Jones, M., Codarini, M., Macaskill, P., Hayen, A. et al. (2010). The accuracy of clinical symptoms and signs for the diagnosis of serious bacterial infection in young febrile children: prospective cohort study of 15 781 febrile illnesses. *BMJ* 340, c1594.
- Daston, L., and Galison, P. (2007). *Objectivity*. New York: Zone Books.
- DeCarlo, T., Roy, T., and Barone, M. (2015). How sales manager experience and historical data trends affect decision making. *European Journal of Marketing* 49, 1484-1504.
- Devitt, M. (2015). Relying on intuitions: where Cappelen and Deutsch go wrong. *Inquiry* 58 (7-8): 669-699.

¹⁴ See Kitcher (2001) on how Galileo argued for the reliability of the telescope as an instrument of celestial observation.

- Douglas, H. (2009). *Science, Policy, and the Value-Free Ideal*. Pittsburgh, PA: University of Pittsburgh Press.
- Eisenstein, E. M., and Lodish, L. (2002). Marketing decision support and intelligent systems: precisely worthwhile or vaguely worthless? In B. A. Weitz and R. Wensley (Eds.), *Handbook of Marketing* (pp. 436-455). London: SAGE Publications.
- Englich, B., Mussweiler, T., and Strack, F. (2005). The last word in court—a hidden disadvantage for the defense. *Law and Human Behavior* 29, 705-722.
- Fahsing, I., and Ask, K. (2016). The making of an expert detective: the role of experience in English and Norwegian police officers' investigative decision-making. *Psychology, Crime & Law* 22, 203-223.
- Ganzach, Y. (1997). Theory and configularity in clinical judgments of expert and novice psychologists. *Journal of Applied Psychology* 82, 954-960.
- Gegenfurtner, A., Lehtinen, E. & Säljö, R. (2011). Expertise differences in the comprehension of visualizations: a meta-analysis of eye-tracking research in professional domains. *Educational Psychology Review* 23, 523-552.
- Gelfert, A. (2011). Expertise, argumentation, and the end of inquiry. *Argumentation* 25, 297-312.
- Goodwin, C. J. (2010). *Research in Psychology: Methods and Design*. Sixth Edition. Hoboken, NJ: Wiley.
- Goodwin, J. (1998). Forms of Authority and the Real *Ad Verecundiam*. *Argumentation* 12, 267-280.
- Goodwin, J. (2001). Cicero's authority. *Philosophy and Rhetoric* 34, 38-60.
- Goodwin, J. (2011). Accounting for the appeal to the authority of experts. *Argumentation* 25, 285-296.
- Griffin, T. D., Jee, B. D., and Wiley, J. (2009). The effects of domain knowledge on metacomprehension accuracy. *Memory & Cognition* 37, 1001-1013.
- Griffin, D., and Tversky, A. (1992). The weighing of evidence and the determinants of confidence. *Cognitive Psychology* 24, 411-435.
- Groopman, J. (2007). *How Doctors Think*. New York: Houghton Mifflin Co.
- Guthrie, C., Rachlinski, J., and Wistrich, A. (2001). Inside the judicial mind. *Cornell Law Review* 86, 777-830.
- Guthrie, C., Rachlinski, J., and Wistrich, A. (2007). Blinking on the bench: how judges decide cases. *Cornell Law Review* 93, 1-43.
- Guthrie, C., Rachlinski, J., and Wistrich, A. (2011). Probable cause, probability, and hindsight. *Journal of Empirical Legal Studies* 8, 72-98.
- Herppich, S., Wittwer, J., Nuckles, M., and Renkl, A. (2013). Does it make a Difference? Investigating the Assessment Accuracy of Teacher Tutors and Student Tutors. *Journal of Experimental Education* 81, 242-260.
- Hinds, P. J. (1999). The curse of expertise: the effects of expertise and debiasing methods on predictions of novice performance. *Journal of Experimental Psychology: Applied* 5, 205-221.
- Hinton, M. (2015). Mizrahi and Seidel: experts in confusion. *Informal Logic* 35, 539-554.
- Holden, M. P., Newcombe, N. S., Resnick, I., and Shipley, T. F. (2016). Seeing Like a Geologist: Bayesian Use of Expert Categories in Location Memory. *Cognitive Science* 40, 440-454.
- Hooke, R. (1664). *Micrographia*. London. Available at: <https://ceb.nlm.nih.gov/proj/ttp/flash/hooke/hooke.html>.

- Huemer, M. (2002). Testimony. In M. Huemer (ed.), *Epistemology: Contemporary Readings* (pp. 217-218). London: Routledge.
- IBM. (2013). Memorial Sloan-Kettering Cancer Center: IBM Watson helps fight cancer with evidence-based diagnosis and treatment suggestions. *IBM Corporation* (January 2013): http://www-935.ibm.com/services/multimedia/MSK_Case_Study_IMC14794.pdf.
- Jackson, S. (2015). Deference, distrust, and delegation: three design hypotheses. In F. H. van Eemeren and B. Garssen (Eds.), *Reflections on Theoretical Issues in Argumentation Theory* (pp. 227-243). Cham: Springer.
- Kahneman, D. (1991). Judgment and decision making: a personal view. *Psychological Science* 2, 142-145.
- Kaufmann, E., Reips, U., and Wittmann, W. (2013). A Critical Meta-Analysis of Lens Model Studies in Human Judgment and Decision-Making. *PLOS ONE* 8, e83528.
- Kidd, J. B. (1970). The utilization of subjective probabilities in production planning. *Acta Psychologica* 34, 338-347.
- Kitcher, P. (2001). Real realism: the Galilean strategy. *Philosophical Review* 110, 151-197.
- Kononenko, I. (2001). Machine learning for medical diagnosis: history, state of the art and perspective. *Artificial Intelligence in Medicine* 23, 89-109.
- Krems, J. F., and Zierer, C. (1994). Are experts immune to cognitive bias? The dependence of confirmation bias on specialist knowledge. *Zeitschrift fur experimentelle und angewandte Psychologie* 41, 98-115.
- Kuhn, T. S. (2000). Afterwords. In J. Conant and J. Haugeland (eds.), *The Road since Structure: Philosophical Essays, 1970-1993, with an Autobiographical Interview* (pp. 224-252). Chicago: The University of Chicago Press.
- Lai, M. (2015). Giraffe: Using deep reinforcement learning to play chess. [arXiv:1509.01549](https://arxiv.org/abs/1509.01549) [cs.AI].
- Langenburg, G., Champod, C., and Wertheim, P. (2009). Testing for potential contextual bias effects during the verification stage of the ACE-V methodology when conducting fingerprint comparisons. *Journal of Forensic Sciences* 54, 571-582.
- Lanzilotti, R., Ardito, C., Costabile, M. F., and De Angeli, A. (2011). Do patterns help novice evaluators? A comparative study. *International Journal of Human-Computer Studies* 69, 52-69.
- Lash, T., Fox, M., and Fink, A. (2009). *Applying Quantitative Bias Analysis to Epidemiologic Data*. Dordrecht: Springer.
- Latour, B., and Woolgar, S. (1986). *Laboratory Life: The Construction of Scientific Facts*. 2nd Ed. Princeton, NJ: Princeton University Press.
- Leventhal, L., Teasley, B., and Rohlman, D. (1994). Analyses of factors related to positive test bias in software testing. *International Journal of Human-Computer Studies* 41, 717-749.
- Linker, M. (2014). Epistemic privilege and expertise in the context of meta-debate. *Argumentation* 28, 67-84.
- Locke, J. (1975). *An Essay Concerning Human Understanding*. P. H. Nidditch (ed.). Oxford: Clarendon Press.
- Lusted, L. B. (1977). *A study of the efficacy of diagnostic radiologic procedures: Final report on diagnostic efficacy*. Committee on Efficacy Studies, American College of Radiology.
- MacLean, C. L., and Dror, I. E. (2016). A primer on the psychology of cognitive bias. In C. T. Robertson and A. S. Kesselheim (Eds.), *Blinding as a Solution to Bias* (pp. 13-24). London: Elsevier.

- Makary, M. A. (2016) Medical error—the third leading cause of death in the US. *BMJ* 353, i2139.
- McGlone, M. S., Kobrynowicz, D., and Alexander, R. B. (2005). A certain Je ne sais quoi - Verbalization bias in evaluation. *Human Communication Research* 31, 241-267.
- McNeil, B. J., Pauker, S. J., Sox, H. C., and Tversky, A. (1982). On the elicitation of preferences for alternative therapies. *New England Journal of Medicine* 306, 1259-1262.
- Mendel, R., Traut-Mattausch, E., Jonas, E., Leucht, S., Kane, J. M., Maino, K., Kissling, W., Hamann, J., (2011). Confirmation bias: why psychiatrists stick to wrong preliminary diagnoses. *Psychological Medicine* 41, 2651-2659.
- Mizrahi, M. (2010). Take My Advice—I am not Following It: Ad Hominem Arguments as Legitimate Rebuttals to Appeals to Authority. *Informal Logic* 30, 435-456.
- Mizrahi, M. (2013). Why arguments from expert opinion are weak arguments. *Informal Logic* 33, 57-79.
- Mizrahi, M. (2016). Why arguments from expert opinion are still weak: a reply to Seidel. *Informal Logic* 36, 238-252.
- Mnih, V., Kavukcuoglu, K., Silver, D., Graves, A., Antonoglou, I., Wiestra, D., and Riedmiller, M. (2013). Playing Atari with deep reinforcement learning. [arXiv:1312.5602](https://arxiv.org/abs/1312.5602) [cs.LG].
- Neale, M. A., and Bazerman, M. H. (1991). *Cognition and Rationality in Negotiation*. New York: The Free Press.
- Northcraft, G. B., and Neale, M. A. (1987). Experts, amateurs, and real estate: an anchoring-and-adjustment perspective on property pricing decisions. *Organizational Behavior and Human Decision Processes* 39, 84-97.
- Nuzzo, R. (2015). Fooling ourselves. *Nature* 526, 182-195.
- Oskamp, S. (1965). Overconfidence in case-study judgments. *The Journal of Consulting Psychology* 29, 261-265.
- Perez, O. (2015). Can experts be trusted and what can be done about it? Insights from the biases and heuristics literature. In A. Alemanno and A. Sibony (Eds.), *Nudge and the Law: A European Perspective* (pp. 115-138). Oxford: Hart Publishing.
- Phillips, J. K., Klein, G., and Sieck, W. R. (2008). Expertise in judgment and decision making: a case for training intuitive decision skills. In D. J. Koehler and N. Harvey (Eds.), *Blackwell Handbook of Judgment and Decision Making* (pp. 297-315). Malden, MA: Blackwell Publishing.
- Posavac, S., Kardes, F. R., Sanbonmatsu, D. M., and Fitzsimons, G. J. (2005). Blissful insularity: When brands are judged in isolation from competitors. *Marketing Letters* 16, 87-97.
- Put, K., Baldo, M., Cravo, A M., Wagemans, J., and Helsen, W. F. (2013). Experts in offside decision making learn to compensate for their illusory perceptions. *Journal of Sport & Exercise Psychology* 35, 576-584.
- Ramos, J. G., Perondi, B., Dias, R. D., Miranda, L. C., Cohen, C., Carvalho, C. R., Velasco, I. T., Forte, D. N. (2016). Development of an algorithm to aid triage decisions for intensive care unit admission: a clinical vignette and retrospective cohort study. *Critical Care* 20, 81.
- Ryle, G. (1946). Knowing how and knowing that. *Proceedings of the Aristotelian Society* 46, 1-16.
- Schwitzgebel, E., and Cushman, F. (2012). Expertise in moral reasoning? Order effects on moral judgment in professional philosophers and non-philosophers. *Mind and Language* 27, 135-153.

- Schwitzgebel, E., and Cushman, F. (2015). Philosophers' biased judgments persist despite training, expertise, and reflection. *Cognition* 141, 127-137.
- Seidel, M. (2014). Throwing the baby out with the water: from reasonably scrutinizing authorities to rampant scepticism about expertise. *Informal Logic* 34, 192-218.
- Shaping, S., and Schaffer, S. (1985). *Leviathan and the Air-Pump: Hobbes, Boyle, and the Experimental Life*. Princeton, NJ: Princeton University Press.
- Shieber, J. (2009). Locke on testimony: a reexamination. *History of Philosophy Quarterly* 26 (1): 21-41.
- Siegel, R. (2016). 20 years later, humans still no match for computers on the chessboard. *NPR*, October 24, 2016. <http://www.npr.org/sections/alltechconsidered/2016/10/24/499162905/20-years-later-humans-still-no-match-for-computers-on-the-chessboard>.
- Staël von Holstein, C. A. S. (1972). Probabilistic forecasting: an experiment related to the stock market. *Organizational Behavior and Human Performance* 8, 139-158.
- Ste-Marie, D. M., and Lee, T. D. (1991). Prior processing effects on gymnastic judging. *Journal of Experimental Psychology: Learning Memory and Cognition* 17, 126-136.
- Tversky, A., and Kahneman, D. (1974). Judgment under uncertainty: heuristics and biases. *Science* 185, 1124-1131.
- Teichman, D., and Zamir, E. (2014). Judicial decision-making: a behavioral perspective. In E. Zamir and D. Teichman (Eds.), *The Oxford Handbook of Behavioral Economics and the Law* (pp. 664-702). New York: Oxford University Press.
- Wagemans, J. H. M. (2011). The assessment of argumentation from expert opinion. *Argumentation* 25, 329-339.
- Wagenaar, W. A., and Keren, G. (1986). Does the expert know? The reliability of predictions and confidence ratings of experts. In E. Hollnagel, G. Maneini, and D. Woods (Eds.), *Intelligent Decision Support in Process Environments* (pp. 87-107). Berlin: Springer.
- Walton, D. (1992). *The Place of Emotion in Argument*. University Park, PA: Pennsylvania State University.
- Walton, D. (2006). Examination dialogue: a framework for critically questioning an expert opinion. *Journal of Pragmatics* 38, 745-777.
- Walton, D. Reed, C., and Macagno, F. (2008). *Argumentation Schemes*. New York: Cambridge University Press.
- Walton, D. (2014). On a razor's edge: evaluating arguments from expert opinion. *Argument & Computation* 5, 139-159.
- Walton, D. (2016). *Argument Evaluation and Evidence*. Cham: Springer.
- Waylen, A. E., Horswill, M. S., Alexander, J. L., and McKenna, F. P. (2004). Do expert drivers have a reduced illusion of superiority? *Transportation Research Part F – Traffic Psychology and Behavior* 7, 323-331.
- Winters-Minder, L. A., Bolding, P. S., Hilbe, J. M., Goldstein, M., Hill, T., Nisbet, R., Walton, N., and Miner, G. D. (2015). *Practical Predictive Analytics and Decisioning Systems for Medicine: Informatics Accuracy and Cost-Effectiveness for Healthcare Administration and Delivery Including Medical Research*. San Diego, CA: Elsevier.
- Wöllner, C., and Deconinck, F. J. A. (2013). Gender recognition depends on type of movement and motor skill. Analyzing and perceiving biological motion in musical and nonmusical tasks. *Acta Psychologica* 143, 79-87.