Do people prefer prediction over accommodation? An empirical study

Laura Elaine Strittmatter (estrittmatter@ethz.ch)
Department of Management, Technology, and Economics, ETH Zürich, Weinbergstrasse 56/58, 8092 Zürich, Switzerland

Stephen H. Dewitt (s.dewitt@ucl.ac.uk)
Department of Experimental Psychology, University College London, London WC1H 0AP, United Kingdom

David A. Lagnado (d.lagnado@ucl.ac.uk)
Department of Experimental Psychology, University College London, London WC1H 0AP, United Kingdom

Abstract

Theories can be designed to predict novel evidence or to accommodate known evidence. Despite the lively debate in philosophy of science whether prediction may hold a superior value over accommodation, people's intuitions about this issue have not been empirically examined. Within a medical scenario, we assess individuals' sensitivity to this dilemma. Overall, we find tentative evidence that people favour the predictive account and regard the predicting theorist (i.e., doctor) as more reliable in contrast to their accommodating counterpart. Strikingly, discrepant preference patterns emerged out of their verbal reasoning data echoing the distinct philosophical stances surprisingly well. Possible reasons why people's reasoning systematically diverges despite the general preference for prediction are discussed.

Keywords: prediction; accommodation; probabilistic reasoning; uncertainty; philosophy of science; confirmation bias

Are people more convinced by a theory if it successfully *predicted* novel (unknown) evidence than if it was purely constructed to *accommodate* that observed evidence? What would each of these outcomes say about the theorist? Whilst some philosophers clearly regard prediction as superior to accommodation with respect to the theory's epistemic value (a.k.a. predictionism; e.g., Maher, 1988; Whewell, 1860; White, 2003; Worrall, 1985), others question the prediction advantage thesis (e.g., Achinstein, 1994; Collins, 1994; Fry, 2018; Harker, 2006; Horwich, 1982; Schlesinger; 1987). However, little is known beyond this epistemic philosophical debate. Previous work has not investigated how people intuitively reason about prediction versus accommodation. Thus, the main question to be addressed in this paper is whether people are sensitive to this distinction.

If theory T makes a prediction, and evidence E later reveals that the prediction was in fact true, this is the case of successful *prediction*. If instead T is merely constructed on the basis of observed E, it is a case of successful *accommodation*. One crucial feature of prediction is the order (Persson, 2016): The hypothesis comes prior to the observation. Still, what distinguishes prediction from accommodation is not only the temporal order by which the theory is constructed, but more importantly the causal relation (White, 2003). The creator of the theory has no control over the observational outcome - allowing the theory to be judged on grounds that are independent from its creator (Popper, 2002). According to White (2003), if T is proposed to predict unknown E, our confidence in T should be strengthened given its independent evidential support.

Prediction over accommodation as gold standard has long been praised among influential philosophers of science. This differentiates 'scientific' theories such as Einstein's gravitational theory that allows for testable predictions, and 'pseudoscientific' theories such as Freud's theory of psychoanalysis that Karl Popper claimed could be used to explain almost any outcome, and so cannot be falsified (Popper, 1963). Put simply, a theory that is constructed to explain the evidence at hand is guaranteed to fit (and confirms what is already known), whereas a theory that makes predictions is exposed to the possibility of refutation. Prediction, in contrast to accommodation, is supposed to protect from the risk of overfitting the theory to known evidence (but see also Hitchcock & Sober, 2004). In contrast, accommodation is often characterised by a complex and overfitting theory that must be altered in the light of new data (Lipton, 1991).

Nonetheless, predictions contain varying degrees of riskiness (Mayo, 2018). Successfully predicting that one will observe at least one red car when watching a busy road for hours is not very impressive and fails to indicate superior background knowledge about how the world works. Test severity refers to the riskiness of a hypothesis being refuted (Mayo, 2018). The severity criterion requires a high likelihood that the test would not be passed, if the hypothesis is false (Mayo, 1991). Thus, we should be more persuaded by a theory's successful prediction that is highly risky than a vague one (see Meehl, 1992).

The same distinction has also been discussed and further developed among legal scholars. If a defendant offers a scenario that produces risky predictions e.g. about forensic evidence, which become verified, this can be a legitimate reason to ascribe higher veracity and reliability to the defendant's predictive statement than if it was only grounded on the defendant's ad hoc explanations about the forensic evidence at hand (Mackor, 2017). Yet, given jurists rely heavily on commonsense interpretations of evidence (Lagnado, 2021), it remains to be established whether (and if so, how) they update their beliefs when testimony is framed as prediction or accommodation.

In M. Goldwater, F. K. Anggoro, B. K. Hayes, & D. C. Ong (Eds.), *Proceedings of the 45th Annual Conference of the Cognitive Science Society.*

In both legal and medical contexts scenarios and alternative scenarios have to be evaluated at the level of theory, hypothesis, and evidence. In the clinical milieu, physicians and other medical experts commonly deal with diagnostic decision-making for suspected cases under high uncertainty and fragmentary knowledge, which may account for the considerable number of diagnostic errors (Graber, 2013; Makary & Daniel, 2016) – largely due to cognitive errors (e.g, Nendaz & Perrier, 2012; Oaksford & Chater, 2020; Thammasitboon & Cutrer, 2013). Thus, physicians' different 'fact-finding' approaches in terms of prediction or accommodation are of interest.

Consider the following example: patient Mr. Smith has several health symptoms and seeks medical advice at hospital to rule out a serious disease. Now imagine how the scenario distinctively unfolds for (a) a prediction case and (b) an accommodation case:

(a) The patient Mr. Smith sees Doctor A and complains about all his symptoms. Doctor A says he thinks Mr. Smith has the disease "Ebrosis". At this hospital all patients go through a full set of routine tests. Doctor A *predicts* that in the full set of routine tests (test 1-20), test 2, test 7, and test 11 will come back positive. Mr. Smith undergoes the full set of routine tests. As Doctor A *predicted*, test 2, test 7, and test 11 come back positive.

(b) In order to obtain a second opinion about his state of health from another physician, Mr. Smith sees Doctor B (at the same hospital) and complains about all of his symptoms. Doctor B has *not received* Doctor A's medical report and hence does not know that Doctor A thinks Mr. Smith has "Ebrosis". However, Doctor B has *received the results* of the full set of routine tests and sees that test 2, test 7, and test 11 are positive. Doctor B says he thinks Mr. Smith has the disease "Morinus".

We utilise mixed methods to examine people's sensitivity to and reasoning about prediction versus accommodation in the hypothetical medical case described above. Our study is not aimed at the debate about the prescriptive question of how one should reason in a 'rational' way based on the alleged relation between the truth of theory and its predictive success. Instead, we adopt an explorative approach that is not wedded to a strong philosophical standpoint although people's reasoning may still be in line with some of the dominant philosophical views. Under the assumption that the prediction cue is recognized, we expect that (1) individuals regard the theory (about the disease), which correctly *predicts* the evidence (test results), as being more likely to be true than the theory, which merely *accommodates* that evidence, ceteris paribus. Likewise, (2) people may judge the predicting source (i.e., doctor) as being more reliable than the accommodating analogue.

Method

Participants A total of 101 participants (59 female, 42 male) were recruited via Prolific Academic and completed the study online (www.prolific.ac.uk). The median age was 34 years (SD_{age} = 13.25) ranging from age 18 to 69. All participants identified as native English speakers, gave informed consent, and were compensated for their time £0.72.

Design A 2 × 2 mixed design was adopted, with Diagnosis (Prediction vs Accommodation) as within-subjects factor and Order (Prediction first vs Accommodation first) as between-subjects factor. The dependent variables were participants' probability judgments of the predicted and accommodated diagnoses, and participants' perceived reliability for each of the doctors. Participants were also asked to provide an explanation about their numeric responses for both ratings (i.e., qualitative judgments).

Materials & Procedure All materials and data can be found at https://osf.io/vrt48/. Participants received onscreen instructions that their task is to reason with information about a fictitious medical scenario as outlined above. They were then presented with the scenario. Participants completed both, the Prediction and Accommodation condition that were combined within one scenario. They were randomly allocated to a scenario with a predicted diagnosis first and an accommodated diagnosis last or the reverse. The two reversely ordered scenarios differed minimally for the sake of a plausible and sound storyline.

Dependent measures were taken using an end-of-sequence method (see Hogarth & Einhorn, 1992). Throughout the elicitation stage, participants had access to the accumulated storyline. Participants were asked to indicate the probability estimates of the likelihood that the patient suffers from the disease Ebrosis and Morinus, respectively, on a scale (0 [impossible] – 100 [absolutely certain]). They were then asked to write down their reasoning of their previous numeric responses in an open text box (without word or time limit): "Please explain your reasoning behind your responses above in as much detail as possible". Subsequently, participants' perceived reliability of Doctor A and Doctor B, respectively, was measured (0 [not reliable at all] – 100 [extremely reliable]). This was again followed by a written explanation of their previous response. The procedure ended with participants providing basic demographics (age, gender).

Results

Quantitative

As stated in our preregistration we contrasted prediction and accommodation (https://aspredicted.org/HWF_XGJ). To control for order-effects we used a repeated measures analysis of variance (ANOVA). Separate 2 (Diagnosis: Prediction vs. Accommodation) × 2 (Order: Prediction first vs. Accommodation first) ANOVA's were calculated for each dependent variable (i.e., probability estimates, reliability judgments). There were no main effects of order for the probability estimates (p = 0.49) or reliability estimates (p = 0.59) and no interactions between order and diagnosis for the probability estimates (p = 0.54) or reliability estimates (p = 0.33). However, we found that the normality assumptions were violated (all Shapiro-Wilk tests, p <.01)

and reliability measures were moderately skewed. Therefore, we additionally conducted nonparametric Wilcoxon signed-rank tests. For transparency we report both parametric and nonparametric test results.

The data of participants' probability estimates can be seen in Figure 1. Participants estimated the probability for the predicted diagnosis higher than for the accommodated diagnosis, which was marginally significant on a two-tailed parametric test, $F(1,99) = 3.40$, $p = .068$, , $\eta_p^2 = .033$ and significant on a two-tailed nonparametric test, $z = 2.35$, $p = .019$, $r_{rb} = .36$. Note that a large number of the probability judgments are at the 50% mark, namely 37 responses in the prediction condition and 48 responses in the accommodation condition (additional plots at OSF). Participants' reliability judgments are displayed in Figure 2. The predicting doctor was rated as more reliable than the accommodating doctor on the parametric test $F(1,99) = 9.165$, $p = .003$, $\eta_p^2 = .085$ and the nonparametric test, $z = 2.96$, $p = .003$, $r_{rb} = .45$, two-tailed.

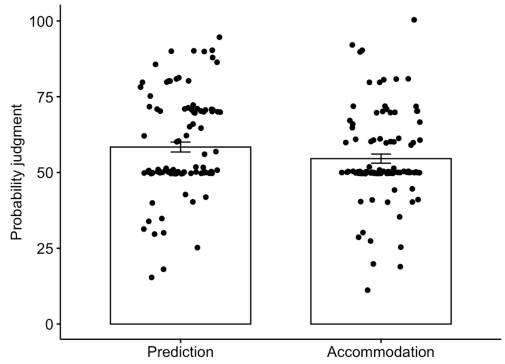

Figure 1. Mean probability judgments by Diagnosis. Error bars represent ±1 standard error of mean (S.E.M.).

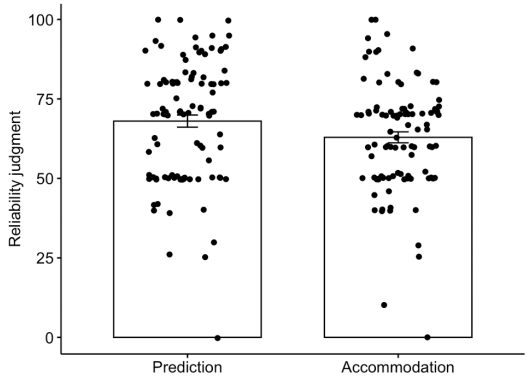

Figure 2. Mean reliability judgments by Diagnosis. Error bars represent ±1 standard error of mean (S.E.M.).

Qualitative

All Participants' open text responses were analysed by the first author. Responses were first segmented into three 'Preference' categories and each participant was allocated to one of these: preferring prediction over accommodation (i.e., Prediction > Accommodation); preferring accommodation over prediction (i.e., Prediction < Accommodation), and having no preference (i.e., Indifferent; see Table 1). Within each preference, participants' reason for their preference was also coded. 'Reason' codes were not mutually exclusive: some participants gave multiple reasons and were therefore assigned multiple codes. Participants gave two open text responses, one asking them to explain the probability judgement, and one to explain their reliability judgement. The coding frequencies for both preferences and reasons can be seen in Table 1. Further detail on each code is given in the next section.

Table 1. Participants' preference determined through qualitative analysis of their open text responses. Within each preference, number of reasons stated. Different reasons include 1) prediction: a) positive attributes or b) negative attributes; 2) accurate prediction; 3) known evidence; 4) confirmation bias; 5) identical evidence; and other. In measurement column, P signifies the explanation of probability judgments, and R the explanation of reliability judgments.

	Preference		Pred. > Accom.		Pred.< Accom.		Indifferent	
	Measurement		P	R	P	R	P	R
	Total number		29	33	13	11	59	57
Reason	1) Prediction	a) +	25	23				
		b) –	14	14				
	2) Accurate pred.		14	11				
	3) Known ev.				9	8		
	4) Conf. bias				6	4		
	5) Identical ev.						39	26
	Other		3	5	3	2	20	31

We conducted a sign test to determine if the number of participants coded as preferring prediction was significantly greater than those preferring accommodation. For the probability judgement open text data (29 vs. 13) the sign test showed there was a significantly higher proportion of participants preferring prediction over accommodation ($p = .02$). The same was true for the reliability judgement (33 vs. 11; $p = .001$). Note that codes of the reliability judgment were broadly in line with those of the probability judgment. For the sake of brevity, the qualitative analysis of the reliability judgments is solely shown in Table 1 and not further outlined in the section below.

Preference: Prediction > Accommodation

Reason 1: Presence and/or lack of prediction

a) Positive attributes for prediction (+): Increased reliability

For more than one half of the participants who preferred the predicted diagnosis the mere fact of having made a prediction seemed to reveal positive attributes about the predicting doctor. Specifically, participants generated positive expectations about the doctor's reliability. It was assumed that the predicting doctor outclassed the accommodating

doctor in terms of level of reliability making the diagnosis of the former more likely. Some participants inferred that the predicting doctor must have more expertise (i.e., the capacity to formulate an accurate diagnosis), e.g., P16 "The first doctor made predictions which suggests he may be more knowledgeable about the condition.", while others mentioned greater trustworthiness (i.e., the willingness to formulate an accurate diagnosis) e.g., P12 "I have more faith in the [predicting] doctor. Without seeing the results and by assessing the symptoms, and hypothesising a diagnosis BEFORE seeing the results, I automatically feel more confident in them."

Some participants ascribed further attributes related to reliability such as increased familiarity, knowledge, and experience (i.e., frequency of exposure) to the doctor given the prediction e.g., P94 "[…] It probable shows [predicting doctor] has other patients with similar symptoms and he analysed based on past data […]".

b) Negative attributes for lack of prediction (–): Overreliance on test results In the absence of prediction, participants ascribed negative attributes to the accommodating approach. In particular, participants argued that the accommodating doctor based his inference exclusively on the outcome of the test results (which actually was not stated as such in the experimental scenario), e.g., P94 "[…] [accommodating doctor] on the other hand predicted based on only the test result" and neglected additional evidence (e.g., symptoms) in contrast to the predicting doctor e.g., P52 "[predicting] Doctor B made a diagnosis and then predicted which test results would support this hypothesis (which they did), whereas [accommodating] Doctor A saw the results first meaning that he was more influenced by these than by the overall patient symptoms and situation."

For some participants the overreliance on test results indicated that the accommodating doctor must lack expertise relative to his predicting colleague, e.g., P8 "[Predicting] Doctor B must have a lot of experience to have been able to diagnose without any tests, whereas [accommodating] Doctor A seems to have solely relied on testing rather than his own experience". Likewise, according to some participants by overly relying on the test results, the accommodating doctor must be less confident in his diagnostic decision, e.g., P28 "[accommodating] Doctor A seems to rely on test results alone and perhaps is unsure about giving a diagnosis".

Reason 2: Accurate prediction Within the preference for prediction over accommodation the accuracy of the prediction was a strongly prevalent reason. The fact that the prediction turns out to be correct (above and beyond the prediction itself), seemed to constitute a highly convincing factor why participants regarded the diagnosis of the predicting doctor as being more likely, e.g., P69 "[Predicting] Doctor A thought that certain tests will come out positive because it is related to the Ebrosis. He was right as those tests came out positive". The test results further validate the mere prediction and highlight the success of the doctor's a priori assumption.

> P50 "I believe [predicting] Doctor B had a higher chance of being correct as he had an idea and prediction of what Mr. Smith could be suffering with prior to any evidence. His prediction is then further solidified when the results come back, solidifying his prior notions."

Note that in order to elucidate why participants judged the likelihood of the disease in the prediction condition as more likely, two possible inferences have to be distinguished. It could be that (i) participants were impressed by the correct prediction per se, which constitutes direct evidence for the disease and makes that *specific* diagnosis more likely. Alternatively (but nonexclusively), (ii) participants could have regarded the success of the prediction as evidence for the doctors' reliability (e.g., experience) *in general* (indirect evidence), which raises the likelihood of a true diagnosis by that doctor. The data of the current reason fail to provide a clear distinction since participants refer to 'the doctor is predicting correctly' leaving this nuance unseparated.

Preference: Prediction < Accommodation

Interestingly, some participants made inferences that point in the opposite direction in relation to the preference discussed so far and therefore favoured accommodation. Nonetheless, in a similar vein as seen in prediction favouring statements their interpretations targeted identical concepts such as the doctors' reliability (but drawing the opposite conclusion).

Reason 3: Inference from known (> predicted) evidence The majority of participants who leant towards the accommodated diagnosis expressed a preference for the diagnosis inferred from known evidence rather than from predicted (unknown) evidence. From that some participants concluded that the accommodated diagnosis must be more reliable e.g., P32 "[accommodating] Dr A's prognosis is more reliable as they came to their conclusion based on the test results. [Predicting] Dr B made their prognosis based on an assumption/prediction of the test results." while others mentioned heightened trustworthiness of the accommodating doctor relative to the predicting equivalent.

> P54 "It does not give me confidence that Doctor A seems to be "predicting" (to a certain degree) what Mr. Smith is suffering from. [Predicting] Doctor A is expecting to see certain results before those results are available to him/her. I suppose that I would have more trust in a Doctor that reaches conclusions after being provided with all test results, and having time to analyse them."

Reason 4: Confirmation bias Confirmation bias emerged as the second most important reason within the Prediction < Accommodation preference. Participants stated that the prediction leads to a premature narrowing down of potential diseases and the predicting doctor seeks for confirmatory

evidence. The test results simply reinforce the prior prediction. This could be partially confounded by the participants' misunderstanding that the doctor has any control over the procedure of testing, which in fact, as stated in the scenario, is a standard procedure that takes place entirely independent from both doctors.

P68 "[Predicting] Doctor A was seeking to confirm a theory they had already had, and which the test results reinforced. [Accommodating] Doctor B viewed the results somewhat more objectively as they were not seeking to confirm a conclusion they had already drawn."

Preference: Indifferent

Reason 5: Identical (and lack of discriminative) evidence

Participants in the indifferent category regarded the presence of the two diseases as equiprobable and mostly failed to appreciate the prediction/accommodation cue. Common answers referred to the identical evidence (i.e., test results and/or symptoms) within the scenario e.g., P60 "Both doctors have access to the same information about symptoms and the same test results" as well as to the lack (and therefore the need) of discriminative evidence that allows a probabilistic differentiation between the two diseases e.g., P4 "There is no certainty as with same results 2 diagnosis have been given. For the right diagnosis further tests or symptoms need to be taken into account for a differential diagnosis."

It is not clear from the majority of those statements whether participants failed to recognise the prediction versus accommodation cue, or if they regarded it as irrelevant or unintelligible. Only very few participants explicitly mentioned the prediction cue but considered it as non-discriminating evidence. Interestingly, some participants arrived at an equiprobability since the advantages and disadvantages of both approaches cancelled each other out.

P33 "Doctor A waited for all the test results before coming to a conclusion. He, therefore, seems to have no preconceived notions and can be viewed as impartial. Doctor B clearly has serious medical knowledge if he can predict test results before they happen. However, he may be picking and choosing evidence from the results that match his expectations. Therefore, he can't be regarded as fully impartial. Either, way, it can't be certain who's right and who's wrong and, therefore, I cannot have more than 50% certainty in either doctor's opinion."

Discussion

The present study tackled the question whether people intuitively differentiate between prediction and accommodation within a naturalistic reasoning task. Up to now, the existence of a 'special psychological effect' to prefer prediction over accommodation has been simply presupposed and justified based on the philosophical discourse (e.g., McIntyre, 2001) despite the fact that it has not been empirically tested. We found tentative evidence from both quantitative and qualitative data for participants' inclination to judge the theory (about the disease) that successfully predicted the evidence (test results) as being more likely to be true than the one that solely explained the known evidence. Individuals also ascribed a slightly higher reliability to the predicting agent (doctor) relative to his accommodating colleague. Interestingly, individuals' open text explanations also revealed a larger number preferring prediction over accommodation. Nevertheless, our open text findings are preliminary as they do not reveal the exact relationship between people's reliability and probability judgments. Hence, a more concrete testing whereby individuals indicate how much they agree with closed opinion statements (e.g., "Doctor A's correct prediction is direct evidence that he must be more reliable compared to Doctor B") would be insightful.

Prediction preference Participants' explanations for preferring prediction sit well with preexisting philosophical arguments. As participants themselves stated, the persuasive element of making a correct prediction may touch on the doctor's wisdom and skills behind the prediction. Some general background knowledge is usually entailed in theories (Worrall, 2002). To predict an outcome, one has to possess the conceptual framework that includes the characteristics of the prediction (Macintyre, 1979) and the credibility of the predictor is strengthened by her predictive success (Barnes, 2008). Doctors as well as scientists hold preconceptions about phenomena in the world. A clinician's belief stems from prior experiences with similar patients or knowledge of disease mechanisms (Wulff, 1981) and the ability to predict successfully may indeed indicate superior reliability, as participants' averaged judgments imply. Besides, for several competing theories the most severely tested should win the battle (van Dongen et al., 2022) and since the accommodated diagnosis cannot be risky, the predicted diagnosis is supposed to count more, which is in harmony with our provisional findings for a small perceived epistemic advantage.

Divergent preferences Still, we showed that participants diverge in their preferences. The aversion towards the predicting and/or accommodating doctor may be conflated by the moral question of what constitutes a good or bad approach by the doctor? Perhaps more so than what is most likely to lead to true or false diagnoses. There are two (sometimes conflicting) goals: Being accurate and favouring certain beliefs (Kunda, 1990). For instance, the opacity of how the doctor arrives at the anticipatory prediction may be a feature that impresses some, but is dubious to others, especially in the context of these dissonant theories between doctors. This in turn shows how individuals reading the identical story come up with conflicting interpretations that may be rooted in their ideological belief system preferences (see Jost et al., 2008), such as judging moral duty, ethical responsibility, or the integrity of the character. Moreover, people's responses may be shaped by features such as individuals' risk preferences (Schildberg-Hörisch, 2018) when regarding the high error-proneness among risky predictions. A further layer

of complexity is added when considering people's theory of mind in that respect. Participants are naïve to the doctor's private knowledge. When speculating about the uncertainties in the hypothetical scenario, it could be that the accommodating doctor had a prediction in mind but did not communicate it before the test results came out (see Jellema, 2021), hence the mere absence of (evidence about) the prediction may not be evidence of the prediction's absence. For instance, even two doctors with identical knowledge could simply have different inner confidence thresholds. Whilst some participants consider the correct prediction as indication of competence and the lack thereof as incompetence, others interpret a communicated prediction as overconfidence, conforming with the slogan 'pride goes before a fall' (cf. Kahn et al., 1996).

Fudging The arguments offered by all three preference camps appear sophisticated and plausible. According to Hodson (1996), predictions and explanations are equally influenced by one's prior conceptual understanding which often leads to the denial of conflicting observational evidence. Using evidence at hand to preserve one's belief is a feature of confirmation bias (Klayman, 1995), which is often seen to be a shortcoming of accommodation. Conversely, a biased selection of data can still occur even if a prediction has been made (Mayo, 1991). Lipton's (2005) "fudging argument" can go both ways and is thus a possibility for each account: a *theory is fudged* when it is constructed to confirm and align with known evidence and *evidence is fudged* when it is selected and explained to fit with one's preconceived theory (Jellema, 2021). Arguably both doctors could be guilty of what Klayman and Ha (1987) termed as 'positive search strategy', whereby the human proclivity is to investigate those properties that are *expected* to occur or that are already *known* - more so than testing the lack thereof. In other words, they are looking for positive tests (present features) favouring the theory more than examining disconfirming lacking evidence (absent features) that provides a better profile about the alternatives. The antidote to a fudging mindset would be a falsificationist mindset (Jellema, 2021). In essence, the prediction camp seems to believe that the predicting doctor has a falsificationist mindset and the accommodating doctor has a fudging mindset, while the accommodation camp ascribed these mindsets conversely. The indifferent camp may, if recognizing this nuance (and regarding it as relevant), be suspicious of the mutually non-exclusive fudging possibilities that could cancel each other out.

Limitations The confirmed novel evidence is not a neat counterfactual comparison to the explained known evidence in our scenario. Although the doctors are mutually blind about each other's work, from the participants' third-person perspective the evidence (test results) may have generally gained in probative value simply due to the doctor's confirmed risky prediction (see Mackor, 2017) and therefore the accommodating doctor may automatically gain an upgrade in his theory since he '(re)uses' the identical evidence. Even though the test results' evidential value was not central to our manipulation, we acknowledge the dependence between conditions via this evidence link. Arguably, the influence would constitute an increase of the estimated probability and reliability in the accommodation condition. In consequence, a diminished difference between the conditions is possible. Nonetheless, for a storyline that approximates at least some complexity as encountered in the real world by combining and contrasting the two, we acknowledge the danger of an order-independent carryover effect. We therefore encourage future endeavors to include scenarios as (1) between-participants design and as (2) modified within-participant design including a predictor and accommodator who independently agree on the same theory (see thought experiment; Lipton, 2005). This may allow us to discover boundary conditions and generalisations above and beyond the particularities within the current scenario. Another shortcoming in the present study is whereas in science the approach of repeated successful predictions and multiple testing can be impressive, within the present case only a single prediction has been made and tested. Against this backdrop, individuals may have been less impressed by such a single predictive success (Harker, 2006; Leplin, 1984). Further, our isolated predicted/accommodated diagnosis may appear contrived since a real diagnosis rather resembles a perpetual information-gathering exercise. Calibrating the experimental scenario in being neither too complex (to reduce noise) nor too impoverished (for ecological validity) is a tightrope walk and thus debatable (see Schwarz, 1996).

Converging findings Taking together, we can certainly consolidate people's partially opposed intuitions on prediction versus accommodation. Often, one accommodates some data to then predict novel ones – thus, well supported theories make both (Lipton, 2005). Humans as sensemaking creatures may operate in a dance between accommodations and predictions, both which can be contaminated by confirmation bias and its sibling 'fudging'. As Popper (1963) pointed out, even theories that satisfy the praised criterion of falsifiability often have their origin in mythological 'accommodating' grounds and hence it does not predicate that they are insignificant, meaningless, or untrue, but rather that predictions are a means to stand up to scientific standards. The state-of-the art in healthcare practices is far from being optimal and has therefore been called upon to go back to the roots of philosophy of science (Worrall, 2010). Although our work provides no final solution to this, empirical continuations of similar, but modified nature would be fruitful for various contexts also besides medicine such as law, science, history, and artificial intelligence – practically every domain where predictions and accommodations are purportedly germane. In sum, our preliminary findings suggest that people are generally sensitive to the predicting versus accommodating approach - not in a uniform, but heterogenous sense that mirrors the distinct philosophical standpoints astonishingly well.

References

Achinstein, P. (1994). Explanation v. Prediction: Which Carries More Weight?. *Proceedings of the biennial meeting of the philosophy of science association* (pp. 156-164).

Barnes, E. C. (2008). The Paradox of Predictivism. Cambridge University Press.

Collins, R. (1994). Against the epistemic value of prediction over accommodation. *Nous*, *28*, 210-224.

Fry, M. (2018). Predictive hypotheses are ineffectual in resolving complex biochemical systems. *History and Philosophy of the Life Sciences*, *40*, 1-46.

Graber, M. L. (2013). The incidence of diagnostic error in medicine. *BMJ Quality & Safety*, *22*, 21-27.

Harker, D. (2006). Accommodation and prediction: The case of the persistent head. *The British Journal for the Philosophy of Science*.

Hitchcock, C., & Sober, E. (2004). Prediction versus accommodation and the risk of overfitting. *British Journal for the Philosophy of Science*, *55*.

Hodson, D. (1996). Laboratory work as scientific method: Three decades of confusion and distortion. *Journal of Curriculum Studies*, *28*, 115-135.

Hogarth, R. M., & Einhorn, H. J. (1992). Order effects in belief updating: The belief-adjustment model. *Cognitive Psychology*, *24*, 1-55.

Horwich, P. (1982). *Probability and Evidence*. Cambridge University Press.

Jellema, H. (2021). The values of prediction in criminal cases. *The International Journal of Evidence & Proof*, *25*, 163-179.

Jost, J. T., Nosek, B. A., & Gosling, S. D. (2008). Ideology: Its resurgence in social, personality, and political psychology. *Perspectives on Psychological Science*, *3*, 126-136.

Kahn, J. A., Landsburg, S. E., & Stockman, A. C. (1996). The positive economics of methodology. *Journal of Economic Theory*, *68*, 64-76.

Klayman, J. (1995). Varieties of confirmation bias. In *Psychology of Learning and Motivation* (pp. 385-418). Academic Press.

Klayman, J., & Ha, Y. W. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review*, *94*, 211.

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, *108*, 480.

Lagnado, D. A. (2021). *Explaining the evidence: How the mind investigates the world*. Cambridge University Press.

Leplin, J. (Eds.). (1984). *Scientific realism*. Scientific Realism, Berkeley, CA: University of California Press, 1984.

Lipton, P. (1991). *Inference to the best explanation*. London: Routledge.

Lipton, P. (2005). Testing hypotheses: prediction and prejudice. *Science*, *307*, 219-221.

MacIntyre, A. (1979). Seven Traits for the Future. *The Hastings Center Report*, *9*, 5–7.

Mackor, A. R. (2017). Novel facts: The relevance of predictions in criminal law. Strafblad 15: 145–156.

Maher, P. (1988). Prediction, accommodation, and the logic of discovery. *Proceedings of the biennial meeting of the philosophy of science association* (pp. 273–285). Philosophy of Science Association.

Makary, M. A., & Daniel, M. (2016). Medical error—the third leading cause of death in the US. *BMJ*, *353*.

Mayo, D. G. (1991). Novel Evidence and Severe Tests. *Philosophy of Science*, *58*, 523–552.

Mayo, D. G. (2018). *Statistical inference as severe testing: How to get beyond the statistics wars*. Cambridge University Press.

McIntyre, L. (2001). Accommodation, prediction, and confirmation. *Perspectives on Science*, *9*, 308-323.

Meehl, P. E. (1978). Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald, and the slow progress of soft psychology. *Journal of Consulting and Clinical Psychology*, *46*, 806.

Nendaz, M., & Perrier, A. (2012). Diagnostic errors and flaws in clinical reasoning: mechanisms and prevention in practice. *Swiss Medical Weekly, 42*, 13706.

Oaksford, M., & Chater, N. (2020). New paradigms in the psychology of reasoning. *Annual Review of Psychology*, *71*.

Persson, U. (2016). Is falsification falsifiable?. *Foundations of Science*, *21*, 461-475.

Popper, K. R. (1963). Science as falsification. *Conjectures and Refutations*, *1*, 33-39.

Popper, K. R. (2002). *The Logic of Scientific Discovery* (2nd ed.). Routledge.

Schildberg-Hörisch, H. (2018). Are risk preferences stable?. *Journal of Economic Perspectives*, *32*, 135-54.

Schlesinger, G. N. (1987). Accommodation and prediction. *Australasian Journal of Philosophy*, *65*, 33-42.

Schwarz, N. (1996). *Cognition and communication: Judgmental biases, research methods and the logic of conversation*. Hillsdale, NJ: Erlbaum.

Thammasitboon, S., & Cutrer, W. B. (2013). Diagnostic decision-making and strategies to improve diagnosis. *Current Problems in Pediatric and Adolescent Health Care*, *43*, 232-241.

van Dongen, N., Sprenger, J., & Wagenmakers, E. J. (2022). A Bayesian perspective on severity: Risky predictions and specific hypotheses. *Psychonomic Bulletin & Review*, 1-18.

Whewell, W. (1860). *Philosophy of the inductive sciences*. London: John W. Parker.

White, R. (2003). The epistemic advantage of prediction over accommodation. *Mind*, *112*, 653-683.

Worrall, J. (1985). Scientific discovery and theory-confirmation. In *Change and Progress in Modern Science* (pp. 301-331). Springer, Dordrecht.

Worrall, J. (2002). New Evidence for Old, in P. Gardenfors, J. Wloenski, and K. Kijania-Placek (Eds.), In the Scope of Logic, Methodology and Philosophy of Science (pp. 191–209). Springer, Dordrecht.

Worrall, J. (2010). Evidence: philosophy of science meets medicine. *Journal of Evaluation in Clinical Practice*, *16*, 356-362.

Wulff, H. R. (1981). How to make the best decision: Philosophical aspects of clinical decision theory. *Medical Decision Making*, *1*, 277-283.