Who resists algorithmic advice?

Cognitive style precludes algorithmic aversion

Heather Yang

Massachusetts Institute of Technology

Renée Richardson Gosline

Massachusetts Institute of Technology

PLEASE DOWNLOAD LATEST VERSION ON http://heatheryang.com/crt-paper

Author note:

Heather Yang, Work and Organization Studies, Sloan School of Management, Massachusetts Institute of Technology. Renée Richardson Gosline, Management Science, Sloan School of Management, Massachusetts Institute of Technology. The authors gratefully acknowledge the comments and suggestions from John Carroll, Sinan Aral, Nick Stagnaro, Dave Rand, JoAnne Yates, Alan Zhang, MIT Initiative on The Digital Economy, MIT Behavioral Lab, the Hehman lab at McGill University; and Hannah Schiller, Leonardo Trigo, Anna-Louise Todsen for research assistance. Correspondence about this manuscript should be addressed to Heather Yang at heather@sloan.mit.edu.

Abstract

As technology and artificially-intelligent (AI) algorithms become increasingly prevalent in all aspects of life, individuals have more opportunity to rely on these sources of advice when making consequential decisions. Recent research has documented both *algorithmic aversion* and *appreciation*, but little is known about who prefers algorithmic over human advice. In this paper, we present the first evidence that peoples' cognitive style corresponds to the kind of advice they prefer to seek out. We present results from 4 studies (combined $N = 2{,}450$) showing that cognitive style consistently predicts the degree to which users seek input on their decisions from AI versus human advisors. Those scoring higher on the Cognitive Reflection Test are more likely to embrace input from artificially-intelligent algorithms into their decisions making, demonstrating *algorithmic appreciation*. We find that this effect is partially mediated by users' perceptions of the expected accuracy of the advisor: people who are more cognitively reflective expect the algorithmic advisor to be more accurate than the human advisor.

Keywords: artificial intelligence; algorithms; advice-seeking; judgment and decision-making; machine assisted decision aids

Cognitive Style Predisposes Preferences for Algorithmic or Human Advice

People frequently interact with artificially intelligent (AI) agents in their personal and professional lives, including chatbots,[1] robo-advisors,[2] and virtual assistants (such as Amazon's Alexa or Apple's Siri). The dawn of digitally-mediated customer experiences portends the rise of greater user empowerment, as individuals are afforded the power to seek advice from multiple sources and choose the degree to which algorithms play a role in choices as wide-ranging as health, finance, and dating. Yet, research about attitudes toward artificially-intelligent algorithmic input demonstrates mixed reactions. According to a recent survey by Forrester, 54% of U.S. consumers think that interactions with chatbots will negatively impact the quality of their lives (2019). A report by the Center for the Governance on AI at Oxford University found mixed support for AI, with 41% supporting the development of AI agents compared to a smaller 22% who opposed it (Zhang and Dafoe, 2019). At the other end of the spectrum, evidence points to a preference among many users for technologically-mediated experiences: 70% of users surveyed said that they sought the power to solve customer service issues on their own and would prefer to do so via text, chat, or messaging if "it were done right" (Aspect Customer Experience Survey, 2016). Moreover, 69% of these users say that they interact with an intelligent assistant or chatbot at least once a month.

[1] A chatbot is a piece of software that conducts a conversation via auditory or textual methods. Such programs are often designed to convincingly simulate how a human would behave as a conversational partner.
[2] Robo-advisors are a class of financial advisor that provide financial advice or investment management online with moderate to minimal human intervention. They provide digital financial advice based on mathematical rules or algorithms.

Despite the potential benefits of AI, the potential drawbacks have caught the eye of keen observers, from academia to Hollywood. Politicians and researchers are raising concerns about the potential harm of the prevalence of AI for jobs, judgment, and the amplification of societal ills due to algorithmic bias (Kochan, 2019; Banker and Khetani 2019; Bo and Tucker 2019; Buolamwini and Gebru, 2018; Noble 2018; Brynjolfsson and McAfee 2011). The dystopian narratives prevalent in popular culture about societal takeover by malicious robots (e.g., "The Terminator", "The Matrix", "I, Robot") are reflected in the implicit biases people hold against new technologies (Elsbach and Stigliani, 2019).

Given the potential for both helpful and harmful roles of AI in our daily lives, it is imperative that we better understand how humans decide to incorporate algorithmic input in their decision. In this paper, we seek to shed light on the relationship between "man" and "machine" by examining how a human's cognitive style – whether they rely on heuristics or on deliberation (Kahneman 2011) -- shapes their preference for algorithmic advice over human advice, a phenomenon referred to as algorithmic appreciation (Logg, Minson and Moore 2019). Using data from 11 online studies, we test the association between an individual's cognitive style and their propensity to seek advice from an algorithmic source and present two main findings. First, building upon research on the Cognitive Reflection Test (Frederick 2005), we suggest that a person's tendency to over-rely on their incorrect, but automatic, responses is related to their preference for human input when making decisions; conversely, those who have a more analytic cognitive style, and engage in greater reflection to find correct answers, are more likely to embrace input from algorithms. Secondly, we identify the expectations of accuracy of the advisor (human or algorithm), as a partial mediator of the relationship between cognitive style and algorithmic appreciation: deliberate thinkers expect algorithmic advisers to be more accurate than do intuitive thinkers. We suggest that cognitive style is a useful lens through which we can understand divergent attitudes toward AI and examine the possibility of algorithmic appreciation in the context of decision domains where people commonly seek to improve their choices. We discuss the implications of this research for underscoring the importance of combatting algorithmic bias and identifying a population that may be susceptible to this problem due to their predilection for AI advisors.

Theoretical Background

Research suggests that algorithms influence choices in both beneficial and harmful ways. Artificial intelligence has been found to be more accurate than the human eye when analyzing some medical data (like ultrasounds for the detection of cancerous lesions; see Golden, 2017; Rodriguez-Ruiz et al., 2019; Liu et al., 2018). Medical chatbots have provided mental health support that has been shown helpful to people less inclined or able to seek human advice, with one study experimentally showing that participants engaged more with a counseling chatbot when described as being algorithmically operated than when described as being staffed by a person (Lucas et al., 2017).

Perhaps it is no wonder, then, that recent research has detailed instances in which individuals prefer advice from an algorithm to that from another human. Recent research has shown that individuals are more likely to take advice for forecasting and estimation tasks when it is framed as being from an algorithm, rather than from other people, controlling for quality (Logg, Minson, and Moore, 2019). Moreover, when individuals can tweak such algorithmically generated estimates, they are even more likely to adopt them, even when the tweaks make them less accurate (Dietvorst, Simmons, and Massey, 2016). Taken together, this stream of research has established the phenomenon of *algorithmic appreciation*: the preference for advice from artificially-intelligent sources.

Indeed, the benefits of technology are widespread. On the bright side, for example, learning software can improve performance outcomes and bridge digital divides in groups where resources are scarce or locations not convenient for traditional classrooms (Vander Ark, 2018; Marr, 2018). However, perceived improvements are not always the same as actual improvements. For example, recent research has shown that, despite feeling better prepared and having a higher willingness to pay, learning software can result in impaired performance when users experience anxiety due to status-based cognitive associations (Banker, Gosline and Lee 2020). Therefore, in addition to the necessary conversations about algorithmic bias, we must also recognize that humans may hold biases with respect to the accuracy and efficacy of AI tools, affecting their impact. This is an important feature of the interaction between man and machine: despite the widespread adoption of artificially intelligent technologies in our homes, schools, and workplaces, and even a preference for algorithmic sources, the effects of reliance on technology are not always as helpful as expected.

Human Bias For Or Against Algorithmic Input

Skepticism among researchers about algorithmic superiority dates back to Paul Meehl's actuarial models in predicting human behavior (Dawes, Faust, and Meehl, 1989). Although simple statistical models outperformed experts in the field, people were reluctant to accept the superiority of outputs from a non-expert, non-human. In more recent research, we see preference for human sources in subjective tasks, even to the detriment of their accuracy (Yeomans, Shah, Mullainathan, and Kleinberg, 2018). For example, participants, predicting how funny someone would consider a joke and whether someone would find someone else attractive for a date, preferred taking the advice of another human to that of a high-quality algorithm. Other research has shown that, when asked to choose between a healthcare recommendation from a human physician versus one from an artificially intelligent program, participants preferred the human, ostensibly due to the ability to shift responsibility for a poor decision to another person as opposed to a blameless tool (Promberger and Baron, 2006). Similarly, within the healthcare domain, Longoni, Bonezzi, and Morewedge (2018) identified "uniqueness neglect" in medical care patients who felt that, although algorithmic medical recommendations may be superior on the whole, their unique situation would be an outlier and thus less accurate than a recommendation from a human (2018).

Although most of the research described has compared information adoption choices from either an algorithmic or human choice, research directly comparing the relative amount of advice taking by Onkal and colleagues (2009) found that participants took more advice from a human expert as opposed to an algorithmic source when divergent pieces of advice were presented simultaneously. These studies establish the phenomenon of *algorithmic aversion*, where humans avoid artificially-intelligent sources of advice in favor of human input.

What could account for the evidence for both *algorithmic appreciation* and *algorithmic aversion*? Most of the research to date has focused on the features of the artificially-intelligent advisor, or the decision context. In a series of experiments, Riva, Sacchi, and Brambilla (2015) examined the role of anthropomorphizing algorithms on gambling behavior and found that gamblers who read anthropomorphized descriptions of a slot machine gambled more than their counterparts in the non-humanized conditions. In the judgment and decision-making literature, research has focused on features of algorithms that appeal to users, such as adjustability

(Dietvorst et al., 2016), transparency of the rationale (Yeomans et al., 2019), accuracy (Dietvorst et al., 2015), task difficulty (Kaufman and Budescu, 2019), and epistemic probability of the item being predicted (Dietvorst and Bharti, 2019).

While scholars have recognized a variety of source-based factors as influencing the preference for human versus AI input, they understand much less about the features of the user that predict whether she will demonstrate algorithmic appreciation or algorithmic aversion. In one of the only papers we found examining individual differences in support for algorithms, Zhang and Dafoe (2019) demonstrated that greater understanding of algorithms and computer science education is predictive of support for algorithmic development. In this paper, we control for this and explore a novel individual difference, cognitive style, as a key predictor of algorithmic appreciation or aversion in advice acceptance behaviors.

Cognitive Style, Advice Acceptance, and Behavior

Cognitive style, as measured by the Cognitive Reflection Test (Frederick 2005), is an individual's tendency to suppress intuitive thoughts and subsequently engage in deliberate, intentional, System 2 thinking. It is considered representative of the dual systems theory advanced by Kahneman (2011), where, in contrast to System 2, System 1 is characterized by automatic, heuristic-based, "fast" cognitive processing. CRT as a formal assessment tool builds on decades of work that discussed cognitive style and advice taking, but under different names like "need for cognition", "intuitive vs rational thinking", and "experiential vs rational" modes of thinking (see Cacioppo, J., Petty, R., Feinstein, J., and Jarvis, W. 1996 for review). For instance, research from the 1950's on advice taking behaviors among mothers showed that those with a high need for cognition are more likely to seek advice from doctors about their child's behavioral problems (Adams 1959). Other research on cognitive style and advice acceptance showed that people processing information rationally are willing to pay substantially more for advice than those who are processing information experientially, and that this effect is moderated by the individual's decision-specific knowledge (Godek and Murray 2008).

The later, widespread adoption of the Cognitive Reflection Test followed, as it proved to be a parsimonious measure of which system is dominant in a person that can be as predictive as the equivalent of 3.5 hours of multiple cognitive tests such as the Need for Cognition scale (self-reported measure of tendency to engage in and enjoy thinking), SAT, ACT, and the Wonderlic Personnel inventory (a measure of general cognitive ability). The CRT has a strong predictive power particularly for biases that may arise in problems for which there exists a correct solution and where analytical skills are helpful to derive this solution (Hoppe and Kusterer 2011). CRT is also a more potent predictor of performance on a wide sample of tasks than measures of cognitive ability, thinking disposition and executive functioning (Toplak, West, and Stanovich 2011). This informed our embrace of CRT for our research design. We initially undertook the examination of cognitive style via CRT as a key individual level difference not only due to it being a highly foundational individual difference that is related to many judgments and behavioral decisions (see next section for a brief review), but also because of the societally widespread mistrust of novel technology (see Glikson and Woolley, 2020 for a meta-analysis) and the relationship between cognitive style and critical belief processing. Cognitive style, we suspected, could help us explore users' relationships with technology by measuring how individuals think. A robust stream of recent research has shown that intuitive cognitive style (measured by lower scores on the Cognitive Reflection Test) is linked with a belief in misinformation (Bago, Rand and Pennycook 2020; Pennycook and Rand 2020, 2018), conspiracy theories (Pennycook and Rand 2019), and pseudo-profound 'bullshit' (Pennycook et al., 2015)[3] Other research has identified "cognitive miserliness" as a predictor of CRT performance (Toplak, West and Stanovich 2014; Stupple, et al., 2017). Given the multi-screen, hyper-stimulated pace of information, humans have more multi-media data to process than ever, affecting everything from purchasing behaviors, to opinions, to social network patterns (Aral 2020). We are just beginning to understand how cognitive style intersects with human reliance on technology, both as a predictor of use, and as an outcome (Vujic 2017). This issue is of great importance to use, especially as we seek to understand how human cognitive bias may be amplified or attenuated by artificial intelligence, and which populations may be more vulnerable to any deleterious effects (Noble 2018; O'Neil 2016).

In this paper, we propose that CRT may hold great potential in helping us understand the pressing issue of pro- or anti-algorithmic bias among users, and related behavioral outcomes. Thus, we explore: do individual level differences in cognitive style affect humans' advice acceptance or avoidance of input from algorithms? And, if this relationship is discovered, what might explain it? The extant literature suggests two possible explanatory variables: familiarity with technology and perceived accuracy. With respect to the former, research has shown that high a CRT score reflects a thinking disposition that interacts with knowledge, domain-specific heuristics and characteristics of the environment Campitelli and Labollita 2010) – including, we hypothesize, familiarity with technology. High CRT scores have been linked with less reliance on smartphones, instead of thinking (Barr et al., 2015); this suggests a potential pattern of

[3] Since artificial intelligence posing an existential threat to humankind is a common trope in popular culture, we initially wanted to see whether individuals prone to believing conspiracy theories would also show a preference for human (vs. algorithmic) advisors. However, neither belief in conspiracy theories nor any of a variety of potential mediators that we examined in our surveys affected the relationship between Cognitive Reflection Test (CRT) score and preference for type of advisor.

technology usage that could result in algorithmic aversion or appreciation according to cognitive style. Moreover, the mere exposure effect, or familiarity principle (Bornstein and D'Agostino 1997), suggests that familiarity with a technological product of system may increase preference. If people receive a low score on the CRT, they may lean on more intuitive cognitive processing and prefer an advisory source that feels familiar. In such a case, familiarity with an algorithmic (human) advisor would lead to higher (lower) displays of algorithmic appreciation. We would expect then, that lower CRT people would be more likely to prefer intuitive experiences, leading to algorithmic aversion relative to higher CRT people, while higher CRT people can control this instinct. We seek to address this issue by measuring how familiarity with algorithms (as well as the decision domain) affects algorithmic appreciation (aversion) according to cognitive style.

Unfamiliar tasks may also prompt perceptions of advisor accuracy. One is more likely to accept advice if the source is presumed more accurate, as it will improve the decision task. This may be even more so the case when a task requires more cognitive skill, as in the case of financial decisions or personal healthcare management. Past research has shown that confidence in the accuracy of one's intuition (fast thinking) depends on the difficulty of the cognitive task (Gill, Swann and Silvera 1998). Perhaps, then, people's judgments about the accuracy of their judgments, and that of the advisor, will affect advice acceptance. However, being offered the choice between a human and an algorithmic advisor may feel like an easy choice in that people display a clear intuition about this, leading System 1-dominant people to rely on their automatic processing, and casting algorithmic advisors in a relatively less appealing light. System 2 thinkers, however, demonstrate the ability to interrupt this automatic process and engage in more deliberative cognitive processing. This leads us to expect that higher perception of accuracy of the advisor would predict advice-seeking behavior, with low CRT anticipating higher accuracy among human advisors, and high CRT people expecting the inverse.

Overview of Studies

The purpose of this paper is to establish whether or not a person’s cognitive style is related to that person’s relative preference for input from an algorithmic advisor versus a human advisor. We sought to present a scenario that participants may plausibly confront in real life, and thus asked participants to interact with either a human or algorithmic advisor through a chat textbox. We measured the participant’s preference for an algorithmic and a human advisor and then examined whether it varied along with the participant’s cognitive style. Given the documented implicit biases held by individuals against novel technologies (Elsbach and Stigliani, 2019), and the research showing analytical cognitive style relating to greater ability to suppress instinctive biases (Toplak, Stanovich, and West, 2011), we expected that those scoring higher in cognitive reflection would be able to suppress their biases against novel technologies and would prefer the algorithmic advisor over the human advisor in comparison to their less reflective counterparts.

The choice between human and AI sources of information is not binary; people can and do refer to multiple sources for advice. Therefore, we operationalize algorithmic aversion and algorithmic appreciation as the stated preference for the relative input of a human and an algorithmic advisor. That is, participants may choose one or the other, and relative proportion in between. This allows us to not only measure this variable continuously, but more faithfully represents to real choices that people make as they navigate technology in their choice tasks.

To better understand potential mediators and alternative explanations, we also measured (Study 2) participants' general comfort with technology, accuracy expectations, the efficacy of their prior experiences using AI (both a binary measure of whether a participant had any experience or not, and additionally the qualitative experience using AI in the past). Initial pilot testing revealed that a commonly cited reason for preferring the non-human advisor was participant social anxiety, and thus we measured participant social anxiety as a potential factor driving the advisor decision. Finally, we measured participant's self-reported confidence in their financial literacy, in case those who were less confident in their financial literacy (who presumably may be less cognitively reflective) may have felt shame discussing this topic area with another human.

Taken together, our studies suggest that CRT predicts algorithmic aversion or appreciation, such that cognitive style can help us understand who is more likely to incorporate input from artificially-intelligent sources. When people engage in fast, intuitive thinking (low-CRT-scoring individuals), we hypothesize that they may be more likely to prefer interaction with other humans as more accurate and familiar to their own intuitive processing style. Conversely, high-CRT scoring people are likely to be algorithmically appreciative, given their ability to suppress their initial, intuitive concerns against novel technology and deliberately recognize the benefits. We test the mediating roles of familiarity with technology and perceived accuracy as potential explanations for the effect. Finally, we discuss the implications of this for behavioral outcomes and the societal implications for widespread algorithmic influence.

Study 1

The data in this study are aggregated from a series of 11 surveys conducted on Amazon's Mechanical Turk from May 2019 to January 2020. The relationship remains robust at about the same effect size across all studies, except for the smallest study (Study 6) that was consistent in effect size and direction but not statistically significant (see Table 1 for a summary of the studies and the correlation sizes). In order to make sure that the relationship between cognitive style was not limited to the original 3-item CRT, we also used an alternative measure of cognitive style for studies 3, 5, 6, and 7.

Table 1.

Study #	*N*	Effect size (Pearson's *r)*	# CRT item	Domain	Comfort w Tech	Prior exp w AI (1-5)	Prior exp w AI (Y/N)	Social Anxiety (SIAS)	BIG 5	Warmth / Competence	Seen CRT before?	Text response	Self-rated intelligence	Pre-registered?
1	183	-0.22*	3	Finance	Yes	Yes	-	Yes	Yes	-	-	-	-	No
2	270	-0.22***	3	Finance	Yes	Yes	-	Yes	Yes	-	-	-	-	No
3	190	-0.20**	7	Mixed	Yes	Yes	-	Yes	Yes	-	-	-	Yes	No
4	323	-0.14*	3	Mixed	Yes	Yes	-	Yes	Yes	-	-	-	Yes	No
5	207	-0.23**	7	Finance	-	-	-	-	-	-	-	-	Yes	No
6	90	-0.11	7	Finance	Yes	Yes	Yes	Yes	-	-	Yes	Yes	-	No
7	99	-0.20*	7	Finance	Yes	Yes	Yes	-	-	-	Yes	Yes	-	No
8	197	-0.24***	3	Finance	-	Yes	Yes	-	-	Yes	Yes	Yes	-	No
9	217	-0.19**	3	Finance	-	Yes	Yes	-	-	Yes	Yes	Yes	-	Yes
10	203	-0.16*	3	Finance	-	Yes	Yes	-	-	Yes	Yes	Yes	-	Yes
11	435	-0.16***	3	Finance	-	Yes	Yes	-	-	Yes	Yes	-	-	Yes

Total: 2414

* $p < 0.05$. ** $p < 0.01$. *** $p < 0.001$.

Method

Participants. Across 11 surveys, 2414 US residents were recruited from Amazon Mechanical Turk in return for market-rate compensation (M_{age} = 40.1, 57.1% women). We excluded any participants for not giving consent or for failing basic attention checks such as "Please choose option 5 for this question".

Measures.

Advisor preference. The focal outcome of advisor choice was self-reported on a sliding scale ranging from 0-100, with all advice coming from the algorithmic advisor being zero, all advice coming from the human advisor being 100, and an even split of advice from both being 50.

Cognitive style.

i. Original Analytical Style (Three Item CRT). We measured cognitive style through the original Cognitive Reflection Test (CRT) (Frederick, 2005). This three-item measure is comprised of questions with an intuitive (but incorrect) answer that must be overcome with deliberation in order to reach to the correct result. Participants were able to give any numerical response; we coded them as correct (1) or incorrect (0) and summed (0-3) for the CRT score (Cronbach's $\alpha = 0.74$).

ii. Alternative Analytical Style (Seven Item CRT). In order to ensure that our results in Study 1 were not limited to one measure of cognitive style or in any way tainted by possible repeated exposure, we measured cognitive style through a modified 7-item version of the original Cognitive Reflection Test (CRT) that has been found to be suitably reliable by other researchers (Pennycook and Rand, 2019). This combined measured consisted of three reworded items from the original CRT that were computationally equivalent but semantically different, e.g., estimating the size of a patch of mold on a loaf of bread (modified) instead of estimating the lily-pad coverage of a pond (original) (Shenhav, Rand, and Greene, 2012) and four non-numeric items that had a false lure answer (e.g., "Emily's father has three daughters. The first two are named May and June. What is the name of the third daughter?") that needed to be suppressed in order to arrive at the correct answer (e.g., Emily) from Thomson and Oppenheimer (2016). The 7-item measure had acceptable reliability, Cronbach's $\alpha = 0.78$. Scores could range from zero (indicating all questions were answered incorrectly and thus the least amount of cognitive reflection) to seven (indicating the highest amount of cognitive reflection).

Intuitive CRT score. For both forms of the Cognitive Reflection Test, we also calculated a separate score that measured participants' intuitive answering (Shenhav, Rand, Greene, 2012). We gave participants one point for each intuitive (but incorrect) lure answer for each question of the CRT. For the three-item CRT, these answers were 10 cents, 100 minutes, and 24 days. For the seven-item CRT, the intuitive answers were 1st place, 7 sheep, June, 27 cubic feet (Thomson and Oppenheimer, 2016), 8 years old, 50 minutes, and 100 days.

To ease interpretation of the all of CRT scores across the studies, we used a standardized percentage of items answered correct (0-100%) instead of the raw scores from the test (either 0-3 or 0-7).

Prior experience with AI.

i. Rated Efficacy. In order to control for participants' previous experience with artificially intelligent technology, we asked them to rate the effectiveness of their previous experience accomplishing a task, getting information, and getting tailored advice using artificial intelligence based on two specific examples for each experience (e.g., making a reservation, figuring out how to make a return, creating a weight management plan). Participants answered these items on a 5-point Likert scale (1 = *Not at all effective*, 5 = *Very effective*). The summed total was averaged across the number of questions answered, leaving a score ranging from one (worst experiences) to five (best experience).

ii. Amount of Experience. Later studies also included the option to answer "No experience" if the participant had not undertaken the task. If the participant reported having some experience with AI, their answer was coded "Yes (some experience)". If the participant reported having no experience at all with AI, their answer was coded "No (no experience with AI)".

Comfort with Technology. We adapted a comfort-with-technology scale used to relate student motivation to learn in online-education settings for a broader general audience (Rodriguez, Ooms, and Montañez, 2008). Respondents used this 8-item scale (Cronbach's $\alpha = 0.88$) to rate their comfort on a 5-point Likert scale (1 = *Very comfortable*, 5 = *Very uncomfortable*) accomplishing a variety of tasks involving technology of varying levels of novelty (e.g., "Downloading and reading e-books," "Use social media to connect with a stranger," "Save and retrieve files in the cloud"). To ease interpretation of the score, we reverse coded the summed score, so that higher numbers signify greater comfort with technology.

BFI-10 Personality index. To control for possible concerns about the individual difference of "Openness to Experience" explaining any variation in preference for algorithmic advisor, we gave participants the short-form (10 item) of the Big Five Inventory (Rammstedt and John, 2007). This version has been shown to have similar reliability and predictive power as the longer version of the Big Five inventory (44 items) that measures openness to experience (Cronbach's $\alpha = 0.61$), neuroticism ($\alpha = 0.78$), agreeableness ($\alpha = 0.41$), conscientiousness ($\alpha = 0.62$), and extroversion ($\alpha = 0.67$). As the preeminent inventory in personality research, the Big Five traits have decades of research demonstrating their external validity and factorial uniqueness (John and Srivastava, 1999; McCrae and Costa, 1987)

Social Anxiety. To measure social anxiety as a possible alternative explanation for desire to interact with an algorithmic advisor, as opposed to another human, we gave participants the short form of the Social Interaction Anxiety Scale (SIAS) that is widely used by clinicians (Fergus et al., 2014). This six-item scale (Cronbach's $\alpha = 0.86$) has been validated by both clinical and non-clinical samples to be predictive of interpersonal functioning, and has shown convergence with other measures of social anxiety (Fergus et al., 2012).

Mindset. To measure participant's lay theory of cognitive malleability, we gave them a three-item measure of intelligence mindset (Dweck, Chiu, and Hong, 1995). This consisted of items assessing the participant's acceptance of statements concerning the malleability (vs. innateness) of learning, such as "You have a certain amount of intelligence and you can't really do much to change it". The mean score on the measure of Mindset was 6.8 out of a potential 15 (*SD* = 4.2), suggesting on average that participants in this sample generally "mostly agreed" with statements implying a fixed mindset.

Confidence in Financial Literacy. Participants self-reported their confidence in their financial literacy on the one-item question through a 5-point Likert scale (1 = *Very confident*, 7 = *Not at all confident*).

Time Taken. We also collected the response time taken to make the focal advisor decision, measured in seconds.

Design and procedure. We advised participants that the survey comprised a series of unrelated tasks. We introduced the Cognitive Reflection Test as an "intelligence test" and we informed the participants that, even though the test was composed of only three questions, they should take their time answering it.

We then asked them to take part in a separate task involving a financial decision-making scenario. We asked participants to imagine themselves in a scenario in which they were looking for advice on managing an investment portfolio of financial assets and then told them that we would give them the opportunity to interact with a financial advisor over a text chat box.

Participants had two options for their advisor —an algorithmic advisor and a human advisor stated to be of equivalent ability– to answer questions that they might have. The advisor, they were told, would ask them questions about their lifestyle, goals, and background and then tailor their advice on the investment portfolio to their answers. They were then provided a continuous sliding scale whereby they could choose the relative amount of advice from both the algorithmic advisor and the human advisor. The bipolar scale ranged from the left end meaning that all of the advice would come from the algorithmic advisor, to the right end meaning that all of the advice would come from a human advisor. We allowed participants to choose along this continuum, so that a participant could choose any combination of advice from either of the advisors. For example, a participant who put the slider at 60% would have received more advice from the human advisor than from the algorithmic advisor.

After choosing the advisor, participants saw an error message that stated that the rest of the financial advice scenario could not be loaded, but that their work would still be compensated. As we were interested in advice seeking behavior that we collected through their choice on the sliding scale, we did not actually let the participants interact with either of the (hypothetical) advisors, nor receive any advice. The participants went through the remaining scales and answered demographic questions before being thanked and paid for their time.

Decision-Making Domain. The main decision paradigm was exactly the same throughout the eleven studies, with Studies 3 and 4 only differing on the inclusion of three additional decision domains: College Admissions, Employee Hiring, and Healthcare Management (The full survey materials are available to download at https://osf.io/sz354/). In these studies, we randomly assigned participants to one of the four decision domain paradigms, where they were asked to imagine themselves as making a decision in that domain area and needing additional advice in order to make the decision (Medical condition: *You have been diagnosed with a health condition and are wanting advice on how to manage it;* College admissions: *You are helping make college admissions decisions and are looking for advice on how to evaluate applicants*; Employee hiring: *You are in the HR department and are looking for advice on how to evaluate applicants for hiring*). The survey only varied in the task introduction, and used exactly the same information to introduce the algorithmic and human advisors across the decision domains.

Results

Our analyses looked at the relationship between cognitive style (as measured by a standardized score on the CRT) and advisor preference (human vs. algorithmic). We conducted *available-case* analyses for the control variable data collected across the 11 studies. There was no difference between the studies aggregated in their relationship of cognitive style and advisor preference, once the decision domain was controlled for (See Table 1, also discussed in more detail below).

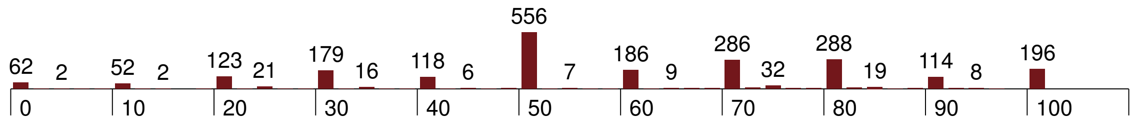
Descriptive Statistics

The average score on the CRT for the participants in this sample was 44% (95% CI: 42.9%, 45.9%), indicating that most participants scored just over one out of the three on the original scale, or about three out of seven questions correctly. As a point of reference, the average CRT score collected across 3,000+ individuals in a diverse range of locations was 1.24 (Frederick, 2005), suggesting that our online sample answered in line with what would be expected and was not unduly swayed by any professional mTurkers who may have been exposed to the CRT before.[4] In our sample, we did find that the number of items that participants had self-reported as having seen before was significantly positively correlated with the number of items that they answered correctly, $r(1141) = 0.39$, $p < 0.001$. We note, however, that this results in a more conservative test of the relationship between CRT and advisor preference, as 'true' deliberation would be reduced by the inflated scores of those who had seen the test before.

Across all CRT scorers, participants slightly preferred more advice coming from a human advisor over an algorithmic advisor (M = 58.0%, SD = 1.0%, Min = 0%, Max = 100%; where 0% = all advice from algorithmic advisor, and 100 = all advice from human advisor. See Figure 1 for a complete distribution).

Figure 1. Distribution of choice frequency for preferred advice proportion

[4] Research has consistently shown that CRT is robust to repeated exposures, where less reflective individuals are less likely to realize the true difficulty of the problems even after multiple exposures (Bialek and Pennycook, 2018). A study by Meyer, Zhou, and Frederick (2018) exposed participants to the test up to 25 times, and found that, on average, each additional exposure only increased participants' scores by 0.024 items. It also follows that those who are less reflective will not doubt the lure answer that they intuitively arrive at enough to double check it with an answer they find online.



Cognitive Style and Algorithmic Appreciation

Our findings revealed that participants who had higher scores on the CRT were more likely to prefer greater artificial-intelligence-based input, Pearson's $r(2413) = -0.187$, $p < 0.001$.

When we subjected the data to an ordinary least squares regression (Table 1) without controls, we see that preference for advice from an algorithmic advisor increased as cognitive reflection increased ($b_{cognitive\ reflection} = -12.396$, $p < 0.001$, 95% CI: [-15.506, -1.536]; see Model 1 in Table 1).

Demographic differences

Adding the demographic controls of participant age, gender, highest education attained, race, annual household income (before taxes), and employment status, we see that the pattern remains the same ($b_{cognitive\ reflection} = -10.690$, $p < 0.001$, 95% CI: [-13.4, -7.982]; see Table 2).

Notably, of the demographic factors, only being female and either Democrat or Independent politically were related to preference for algorithmic advisors. Female participants, compared to male and a small number ($N = 11$) of 'other' gender participants, were more likely to choose a greater proportion of advice from a human advisor ($b_{Female} = 3.849$, $p < 0.001$, 95% CI: [1.787, 5.911]). This result is in line with previous work, such as Zhang and Dafoe's (2019) finding that women were less supportive of the development of AI technologies as compared to men. Democrats and Independents were more likely to prefer a greater proportion of advice from an algorithmic advisor ($b_{Democrat}$ = -5.533, $p < 0.001$, 95% CI: [1.787, 5.911]; $b_{Independent}$ = -5.018, $p < 0.001$, 95% CI: [1.787, 5.911], as compared to Republican participants. This finding is also consistent with Zhang and Dafoe's (2019) work, which showed that being a Democrat (as opposed to a Republican) was a significant predictor of support for AI development.

Robustness checks

Prior experience with AI and Comfort with Technology

Unsurprisingly, prior experience with AI was significantly related to advisor preference, where the more efficacious the prior experiences the participants had with AI, the greater the preference for advice from the algorithmic advisor (M = 3.82 out of a potential 5, SD = 0.039; $b_{Prior\ exp.\ w/\ AI}$ = -2.676, $p < 0.001$, 95% CI: [-4.34, -1.01]; see Model 3a in Table 3). Comfort with technology was also significantly related to preference for algorithmic advice (M = 14.5 out of a potential 40, SD = 0.66; $b_{Comfort\ w/\ Tech}$ = -0.212, $p < 0.001$, 95% CI: [-0.36, -0.065]).

Surprisingly, even after controlling for these factors, individual cognitive style was still predictive of advisor choice ($b_{cognitive\ reflection}$ = -12.489, $p < 0.001$, 95% CI: [-16.27, -8.71]). All variables across all six models had Variance Inflation Factors smaller than 5, mitigating any concerns of multicollinearity between cognitive style and the other controls.

Individual differences

Social Anxiety

We had hypothesized that social anxiety may drive whether or not participants would want to interact with another human (vs. an algorithm), as socially anxious individuals may be more likely to avoid interacting with another human due to the potential judgment that might come with an interaction. However, this factor was not significant in predicting advisor choice (b_{SIAS} = 0.077, *p* = 0.57; See Model 3b in Table 3). Overall, participants scored a mean of 18.3 (*SD* = 0.3, *Min* = 6, *Max* = 30) out of a potential 30.

BIG-5 Personality Index

To rule out the possibility of personality-driven individual differences being predominant in driving desires to engage in a new technology, we also submitted the scores of all five factors into the model. We had hypothesized that extroverts may prefer the interaction with another human instead of what may be seen as a task-focused algorithmic advisor, and that individuals high on openness to experience would be more open to engaging with a novel algorithmic advisor. We had no predictions for conscientiousness, neuroticism, or agreeableness.

When we included the BIG-5 personality traits, the relationship between cognitive style and advisor preference was unchanged, ($b_{cognitive\ reflection}$ = -11.370, $p < 0.001$, 95% CI: [-15.39, -7.35]; see Model 3c). Although small, Extraversion was positively and significantly related to greater preference for human advice, as predicted ($b_{Extraversion}$ = 0.818, *p* = 0.04, 95% CI: [0.041, 1.60]; see Model 3c). Openness, however, was not significantly related to preference for advisor (b_{Open} = -0.128, *p* = 0.76; same model), suggesting that highly open individuals were not only open towards advice from the algorithmic advisor, but also open towards the human advisor. Although not hypothesized (and outside the scope of this paper), neuroticism, agreeableness, and conscientiousness were all significantly related to greater preference for advice from the human advisor (see Model 3c for coefficients). The more neurotic, agreeable, or conscientious individuals preferred their advice from a human advisor.

Confidence in financial literacy

Given the financial advice seeking scenario (for the majority of studies in this paper), we included a measure of confidence in financial literacy in case individuals who were highly confident in their financial ability might be more likely to dismiss the potential benefits of talking to another person and thus more receptive to a novel advice source of an algorithmic advisor. On the other hand, individuals confident in their financial literacy may feel less shame discussing their situation with another person and therefore show greater preference for a human advisor.

Although we were also initially concerned that less cognitively reflective participants would be less confident in their financial literacy, we found there was no significant relationship between scores on the CRT and self-reported confidence levels (Pearson's $r(451) = -0.09$, $p = 0.07$). Nor did we find a relationship between advice preference and self-reported financial literacy ($b_{Fin.\ Lit} = -0.29$, $p = 0.81$). This may be due to more reflective individuals being better calibrated in their confidence, whereas less reflective individuals may not account for what they might not know. Differently put, less reflective individuals may not reflect enough to recognize what they do not know, whereas highly reflective individuals might possess greater meta-cognition of the limits of their ability, and thus report lower financial literacy than an overconfident low-reflective peer who rates herself as more able than she is.

Regardless of potential reasons for why confidence in financial literacy was not related to preference for advisor, the inclusion of the control did not affect the focal relationship between cognitive style and advisor preference, ($b_{cognitive\ reflection} = -12.113$, $p < 0.001$, 95% CI: [-17.87, -6.35]; see Model 4a, Table 3).

Self-Perceived Intelligence

In a similar vein as our logic for including confidence in financial literacy, we included individuals' perceptions of their own intelligence in comparison to other people in case individuals who felt that they were much smarter than the average person would feel less inclined to talk to another person, and would therefore prefer more input from an algorithmic advisor.

In line with the 'Lake Wobegon'[5] effect of generalized overconfidence, individuals rated themselves smarter than 65.6% (SD = 1.16) of the population, on average. This was somewhat merited, however, as we found a significant, positive correlation between standardized CRT score and self-perceived intelligence, with individuals scoring higher on the CRT rating themselves as smarter than more of the population (Pearson's $r(718) = 0.13$, $p < 0.001$).

Similar to our null finding for the Confidence in Financial literacy control, we did not find a relationship between self-perceived intelligence and advisor preference ($b_{Perceived\ Intelligence} = -0.01$, $p = 0.82$; See Model 5a, Table 3), nor did its inclusion change the focal relationship between standardized CRT and advisor preference ($b_{cognitive\ reflection} = -12.673$, $p < 0.001$, 95% CI: [-17.7, -7.64]; see Model 5a).

Mindset

Despite a large body of work on the effect of a 'growth' versus 'fixed' mindset on motivation (Dweck, 1986), we found no work that linked cognitive style with mindset (see Rattan and Georgeac, 2017, for a review on implicit theories on social interactions). We investigated

[5] For readers unfamiliar with the reference, 'Lake Wobegon' is a fictional town, envisioned by popular radio personality Garrison Keillor, where "all the women are strong, all the men are good-looking, and all the children are above average".

whether belief in malleability of characteristics (that vary along a continuum from an 'entity theory' to an 'incremental theory') could vary between those who are cognitively reflective versus more intuitive. If cognitive intuition can be conceptualized as the tendency to rely on automatic and heuristic-based processing, then it could follow that individuals who are low on cognitive reflection may also tend toward a more fixed mindset. More cognitively reflective individuals may be able to come to the reasoned position that their abilities are able to improve given practice and effort, as opposed to being an innate trait. These mindsets may affect whether or not individuals decide to interact with a novel technology of an algorithmic advisor that may not have previously been a part of their skillset.

However, this was not the case; there was no significant relationship between mindset and CRT score (Pearson's $r(205) = -0.02$, $p = 0.75$), nor a relationship between mindset and advisor preference ($b_{Mindset} = 0.565$, $p = 0.18$; See Model 5b, Table 3). Interestingly, the inclusion of mindset as a control resulted in self-perceived intelligence becoming a significant predictor of advisor preference, with those higher in self-perceived intelligence being slightly more likely to prefer advice from a human ($b_{Perceived\ Intelligence} = 0.237$, $p < 0.001$; See Model 5b). Despite self-perceived intelligence becoming a significant factor, the focal effect of cognitive style and advisor preference remained, and benefited from a small boost in strength ($b_{cognitive\ reflection} = -19.43$, $p < 0.001$, 95% CI: [-30.7, -8.18]; see Model 5b).

Desire for Control

To address some concerns that individuals who desire more control would opt for more advice from a human advisor, we measured desire for control as a control variable. It did not have any relationship to advisor preference ($b_{Desire\ for\ Control} = 0.045$, $p = 0.68$; See Model 6c, Table 4), nor substantively change the focal relationship between cognitive style and advisor preference ($b_{cognitive\ reflection}$ = -7.656, $p < 0.001$, 95% CI: [-15.0, -0.29]).

Locus of Control

Whereas the Desire for Control variable measured an individual's preference for being in control, the Locus of Control variable measured an individual's beliefs about how strongly they perceive control over the situations they experience in their lives (cite). Locus of Control did not have any relationship to advisor preference (b_{LOC} = 0.162, p = 0.52; See Model 6c, Table 4), nor substantively change the focal relationship between cognitive style an advisor preference (same model).

Scope Conditions

Although the data suggest that the decision domain has a significant effect on how much advice individuals prefer coming from human vs. algorithmic sources, analytical cognitive style still has a predictive capability of advisor preference.

In evaluating the differences between the studies aggregated in this dataset, we found that studies 3 and 4 differed significantly in their mean advisor preference, with participants in those two studies preferring more advice coming from the human advisor than those in the other nine studies ($b_{Study\ 3}$ = 5.506, p = 0.01, 95% CI: [0.502, 10.51]; $b_{Study\ 4}$ = 7.901, $p < 0.001$, 95% CI: [-3.431, 12.37; see Model 2 in Table 1). Study 3 and 4 differed from the other studies in the inclusion of multiple decision-making domains, namely healthcare management, college admissions, and employee hiring. Once decision domain was included as a control variable, the effect of Study 3 and 4 disappeared completely (see Model 2 and 3 in Table 1). Across the four decision-making domains, only healthcare management was significantly different in predicting preference for algorithmic vs. human advice ($b_{Healthcare\ Domain}$ = 11.950, $p < 0.001$, 95% CI: [5.925, 17.98]), with individuals preferring more advice coming from a human (vs. algorithmic) advisor than individuals making decisions in the other domain areas. This is in line with the findings of Promberger and Baron (2006) and Longoni, Bonezzi, and Morewedge (2018).

Despite this difference, individuals in the healthcare management decision domain were nonetheless subject to the predictive relationship of CRT and advisor preference, with higher CRT scores associated with greater advice sought from the algorithmic advisor, (Pearson's $r(128) = -0.17$, $p = 0.048$).

Across all three models (without study controls, with study controls, and with study and decision domain controls), the relationship between cognitive analytical style and advisor preference remain substantively unchanged, ($b_{cognitive\ reflection,\ Model\ 3}$ = -12.239, $p < 0.001$, 95% CI: [-14.85, -9.63]; See Table 1 for all coefficients).

General Discussion

In this paper, we find a robust relationship between cognitive style and a preference for advice from an algorithmic vs. human advisor. The focal relationship between cognitive style and preference for advice from algorithmic advisor was consistent, whether using the original 3-item CRT or the revised 7-item CRT, with more reflective individuals preferring greater advice from an algorithmic (vs. human) source. This was even after controlling for the BIG-5 personality traits, social anxiety, comfort with technology, prior experiences with AI, cognitive mindset, self-perceived intelligence, self-reported financial literacy (for the financial decision-making scenarios), desire for control, locus of control, decision domains, and demographic variables.

Our work builds on previous research that has demonstrated the relationship between cognitive style and behavioral outcomes such as voting patterns (Pennycook and Rand 2019), and information search during food choice (Mawad et al. 2015; Ares, et al. 2014). It also relates to previous research that examines how technological innovations interact with cognitive styles to augment cognitive biases, as in the case of misinformation in social media and reliance on technology (Pennycook and Rand 2019). In this study, we use CRT to focus on how cognitive style can lead to the phenomenon of algorithmic appreciation or algorithmic aversion in decision tasks that may seem prosaic, but can aggregate to important patterns.

We believe our results raise an important question about how human bias may interact with algorithmic bias to reproduce inequality. If, as our research suggests, high CRT people are more prone to preferring algorithmic sources of input over human ones, then this would underscore the importance of rooting out algorithmic bias. Research has shown that high CRT scores are predictive of the ability to interrupt flawed heuristics, resulting in behavioral outcomes that are linked to traditionally-held markers of "success:" understanding science, financial impulse control, and success in second language acquisition, for example (Sheremeta 2018; Pennycook, Fuselgang, and Kohler 2015; Shtulman and McCallum 2014; Jamieson 1992). Within the business sphere, high CRT has been shown to correlate with tendencies to invest on the long term and delay gratification (Białek and Sawicki, P., 2018) and with greater risk-taking (Thoma V. et al, 2015). So, though analytical thinking is correlated with types of successful outcomes as well as lower prejudice and religiosity (Franks and Scherr 2017; Karadöller, Yılmaz and Sofuoglu 2015), algorithms that have baked-in bias may undermine these deliberative processes as humans increasingly "outsource" cognitive tasks to artificial intelligence. We hope to further explore this dynamic directly and hope other researchers will continue to examine how cognitive styles among those most powerful interact with algorithmic bias to amplify social stratification, despite loftier intentions.

This work, of course, has its limitations. As online sourcing of participants has become established as a widespread practice, researchers across psychology and political science have investigated and found that samples from mTurk respond similarly to in-lab and field samples, and have the additional benefit of being more demographically representative than convenience samples such as undergraduate students (for a review, see Horton et al., 2011; Berinsky, Huber, and Lenz, 2012; Rand, 2012; Buhrmester, Kwang, and Gosling, 2016). Indeed, we did replicate results from other studies, such as that more educated respondents had higher CRT scores (Pearson's $r(2412) = 0.15$, $p < 0.001$). The implications of the current findings should also caution researchers studying AI adoption to account for the higher than average cognitive reflection of student samples (Frederick, 2009) that may skew results to be more pro-AI than the general population. For practitioners rolling out novel AI agent-based features, these results suggest an easier adoption trajectory for target audiences naturally higher in cognitive reflection (e.g., non-religious, socially liberal, low testosterone individuals, Shenhav, Rand, and Green, 2012; Pennycook, Ross, Koehler, and Fugelsang, 2016; Bahcekapili and Yilmaz, 2017; Deppe et al., 2015; Iyer, Koleva, Graham, Ditto, and Haidt, 2012; Nadler, Jiao, Peiran, and Johnson, 2017). Future work can build on these findings to identify a manipulatable mechanism that could be applied across a range of work contexts.

It does not seem that low CRT scorers were answering at random. 71.8% of those with incorrect CRT responses supplied the 'intuitively' correct answer. This suggests that they were responding using their cognitive heuristics, and not scoring poorly due to inattention.

Additionally, the variance of preference choice between higher cognitive reflective individuals and lower cognitively reflective individuals was similar ($SD_{highCRT}$ = 25.3, SD_{lowCRT} = 23.1). This suggests that participants who scored poorly on the CRT were not paying less attention to the focal question than those who scored well. Overall, we found little evidence that responses in this data set are tainted by gross inattention by mTurk participants.

Our findings also show that perceptions of advisor accuracy, which are biased in and of themselves, partially mediates the relationship between cognitive style and algorithmic acceptance (aversion). Yet, despite using measures for social anxiety and comfort with technology, we did not find any evidence for these factors influencing preference for interacting with an algorithmic vs. a human advisor. One possible explanation for the lack of relationship could be the purported equivalence of the human and algorithmic advisors presented in the task. Since both advisors were presented as ready to ask the same questions and then present their advice, and there was no personal interaction or visual cues, it may not trigger any additional social anxiety. Similarly, comfort with technology may not have entered in the decision-making of the participants in this paradigm as the actions required of each advisor were held constant. Thus, the findings from this study suggest the existence of reasons outside of comfort with technology and social anxiety that drive the preference for an algorithmic advisor.

Despite limitations, these studies contribute to the judgment and decision-making literatures surrounding advice-seeker attributes, as well as to the algorithmic decision-making literature in identifying a novel individual level difference that that influences algorithmic appreciation consistently from those who are averse. It builds on the cognitive style literature in showing another behavioral difference between those who rely predominantly on their intuitive, System 1 thinking and those who are more deliberative and rely on their System 2 thinking.

References

Adams, J. (1959). Advice seeking of mothers as a function of need for cognition. Child Development, 30(1), 171-176.

Aral, Sinan. The Hype Machine: how social media disrupts our elections, our economy, and our health–and how we must adapt. Penguin Random House, 2020.

Ares, G., Mawad, F., Giménez, A., and Maiche, A. (2014). Influence of rational and intuitive thinking styles on food choice: Preliminary evidence from an eye-tracking study with yogurt labels. Food Quality and Preference, 31, 28-37.

Aspect Customer Experience Survey (2016) https://www.aspect.com/globalassets/2016-aspect-consumer-experience-index-survey_index-results-final.pdf

Bago, Bence, David G. Rand, and Gordon Pennycook. "Fake news, fast and slow: Deliberation reduces belief in false (but not true) news headlines." *Journal of experimental psychology: general* (2020).

Banker, Sachin, Renée Richardson Gosline, and Jeffrey K. Lee. "Reversing the Placebo: Performance-Branded Experiences Can Undermine Consumer Performance." *Journal of Consumer Psychology* 30.1 (2020): 140-148.

Banker, Sachin, and Salil Khetani. "Algorithm Overdependence: How the Use of Algorithmic Recommendation Systems Can Increase Risks to Consumer Well-Being." *Journal of Public Policy and Marketing* 38.4 (2019): 500-515.

Baron, R. M., and Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of personality and social psychology*, *51*(6), 1173.

Barr N., Pennycook G., Stolz J.A., and Fugelsang J.A. (2015). The brain in your pocket: Evidence that Smartphones are used to supplant thinking. Computers in Human Behavior, 48, 473–480.

Berinsky, A. J., Huber, G. A., and Lenz, G. S. (2012). Evaluating online labor markets for experimental research: Amazon.com's Mechanical Turk. *Political analysis*, *20*(3), 351-368.

Bialek, M., and Pennycook, G. (2018). The cognitive reflection test is robust to multiple exposures. *Behavior research methods*, *50*(5), 1953-1959.

Bornstein, Robert F., and Paul R. D'Agostino. "Stimulus recognition and the mere exposure effect." Journal of personality and social psychology 63.4 (1992): 545.

Browne, M., Pennycook, G., Goodwin, B., and McHenry, M. (2014). Reflective minds and open hearts: Cognitive style and personality predict religiosity and spiritual thinking in a community sample. *European Journal of Social Psychology*, *44*(7), 736-742.

Brynjolfsson, Erik, and Andrew McAfee. Race against the machine: How the digital revolution is accelerating innovation, driving productivity, and irreversibly transforming employment and the economy. Brynjolfsson and McAfee, 2011.

Buhrmester, M., Kwang, T., and Gosling, S. D. (2016). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality data? In A. E. Kazdin (Ed.), *Methodological issues and strategies in clinical research* (p. 133–139). American Psychological Association.

Cacioppo, J., Petty, R., Feinstein, J., and Jarvis, W. (1996). Dispositional Differences in Cognitive Motivation: The Life and Times of Individuals Varying in Need for Cognition. Psychological Bulletin, 119(2), 197-253.

Campitelli, Guillermo and Labollita, Martin. (2010). Correlations of cognitive reflection with judgments and choices. Judgment and Decision-making. 5. 182-191.

Chen, B. X., and Metz, C. (2019, May 22). *Google's Duplex Uses A.I. to Mimic Humans (Sometimes)*. New York Times.
https://www.nytimes.com/2019/05/22/technology/personaltech/ai-google-duplex.html

Cowgill, Bo, and Catherine E. Tucker. "Economics, fairness and algorithmic bias." preparation for: Journal of Economic Perspectives (2019).

Dawes, R. M., Faust, D., and Meehl, P. E. (1989). Clinical versus actuarial judgment. *Science,* 243(4899), 1668-1674.

Dietvorst, Berkeley J., Joseph P. Simmons, and Cade Massey. "Algorithm aversion: People erroneously avoid algorithms after seeing them err." Journal of Experimental Psychology: General 144.1 (2015): 114.

Dweck, C. S. (1986). Motivational processes affecting learning. *American psychologist*, *41*(10), 1040.

Dweck, C. S., Chiu, C.-Y., and Hong, Y.-Y. (1995). Implicit Theories and Their Role in Judgments and Reactions: A Word From Two Perspectives. Psychological Inquiry, 6(4), 267–285. http://doi.org/10.1207/s15327965pli0604_1

Elsbach, K. D., and Stigliani, I. (2019). New information technology and implicit bias. *Academy of Management Perspectives*, *33*(2), 185-206.

Fergus, T. A., Valentiner, D. P., McGrath, P. B., Gier-Lonsway, S. L., and Kim, H. S. (2012). Short-forms of the Social Interaction Anxiety Scale and the Social Phobia Scale. *Journal of Personality Assessment, 94*, 310-320. doi:10.1080/00223891.2012.660291

Fergus, T. A., Valentiner, D. P., Kim, H.-S., and McGrath, P. B. (2014). The Social Interaction Anxiety Scale (SIAS) and the Social Phobia Scale (SPS): A comparison of two short-form versions. *Psychological Assessment, 26*(4), 1281–1291. https://doi.org/10.1037/a0037313

Franks, Andrew and Scherr, Kyle. (2017). Analytic Thinking Reduces Anti-Atheist Bias in Voting Intentions. The International Journal for the Psychology of Religion. 10.1080/10508619.2017.1313013.

Frederick, Shane. "Cognitive reflection and decision-making." Journal of Economic perspectives 19.4 (2005): 25-42.

Gill, M. J., Swann, W. B., Jr., and Silvera, D. H. (1998). On the genesis of confidence. Journal of Personality and Social Psychology, 75, 1101– 1114.

Godek, J., and Murray, K. (2008). Willingness to pay for advice: The role of rational and experiential processing. Organizational Behavior and Human Decision Processes, 106(1), 77-87.

Golden, J. A. (2017). Deep learning algorithms for detection of lymph node metastases from breast cancer: helping artificial intelligence be seen. *Jama*, *318*(22), 2184-2186.

Heimberg, R. G., Horner, K. J., Juster, H. R., Safren, S. A., Brown, E. J., Schneier, F. R., and Liebowitz, M. R. (1999). Psychometric properties of the Liebowitz social anxiety scale. *Psychological medicine*, *29*(1), 199-212.

Hoppe, E. I., and Kusterer, D. J. (2011). Behavioral biases and cognitive reflection. Econ. Lett. 110, 97–100. doi: 10.1016/j.econlet.2010.11.015

Horton, J. J., Rand, D. G., and Zeckhauser, R. J. (2011). The online laboratory: Conducting experiments in a real labor market. *Experimental economics*, *14*(3), 399-425.

John, O. P., and Srivastava, S. (1999). The Big 5 trait taxonomy: History, measurement, and theoretical perspectives. In L. A. Pervin and O. P. John (Eds.), Handbook of personality: Theory and research (2nd ed., pp. 102-138). New York: Guilford.

Kahneman, Daniel. Thinking, fast and slow. Macmillan, 2011.

Karadöller, Dilay and Yılmaz, Onurcan and Sofuoglu, Gamze. (2015). Analytic Thinking, Religion and Prejudice: An Experimental Testing of the Dual-Process Model of Mind. 10.13140/RG.2.2.24522.77764.

Knight, Will. "How to tell if you're talking to a bot." (2018).

Liebowitz, Michael R (1987). "Social Phobia". Anxiety. Modern Problems of Pharmacopsychiatry. *Modern Trends in Pharmacopsychiatry. 22.*

Liu, Y., Kohlberger, T., Norouzi, M., Dahl, G. E., Smith, J. L., Mohtashamian, A., ... and Stumpe, M. C. (2018). Artificial Intelligence–Based Breast Cancer Nodal Metastasis Detection: Insights Into the Black Box for Pathologists. *Archives of pathology and laboratory medicine*.

Logg, Jennifer M., Julia A. Minson, and Don A. Moore. "Algorithm appreciation: People prefer algorithmic to human judgment." Organizational Behavior and Human Decision Processes 151 (2019): 90-103.

Longoni, C., Bonezzi, A., and K Morewedge, C. (2018). Consumer Reluctance Toward Medical Artificial Intelligence: the Underlying Role of Uniqueness Neglect. *ACR North American Advances*.

Lucas, G. M., Rizzo, A., Gratch, J., Scherer, S., Stratou, G., Boberg, J., and Morency, L. P. (2017). Reporting mental health symptoms: breaking down barriers to care with virtual human interviewers. *Frontiers in Robotics and AI*, *4*, 51.

Marr, 2018. https://www.forbes.com/sites/bernardmarr/2018/07/25/how-is-ai-used-in-education-real-world-examples-of-today-and-a-peek-into-the-future/#146ee5a8586e

Meyer, A., Zhou, E., and Shane, F. (2018). The non-effects of repeated exposure to the Cognitive Reflection Test. *Judgment and Decision-making*, *13*(3), 246.

McCrae, R. R., and Costa, P. T. (1987). Validation of the five-factor model of personality across instruments and observers. *Journal of personality and social psychology*, *52*(1), 81.

Noble, Safiya. Algorithms of Oppression: How search engines reinforce racism. NYU Press, 2018.

O'Neil, Cathy. Weapons of math destruction: How big data increases inequality and threatens democracy. Broadway Books, 2016.

Pennycook, G., Cheyne, J. A., Barr, N., Koehler, D. J., and Fugelsang, J. A. (2015). On the reception and detection of pseudo-profound bullshit. *Judgment and Decision-making*, *10*(6), 549-563.

Pennycook, Gordon, and David G. Rand. "Who falls for fake news? The roles of bullshit receptivity, overclaiming, familiarity, and analytic thinking." *Journal of personality* 88.2 (2020): 185-200.

Pennycook, Gordon, and David G. Rand. "Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning." Cognition 188 (2019): 39-50.

Rand, D. G. (2012). The promise of Mechanical Turk: How online labor markets can help theorists run behavioral experiments. *Journal of theoretical biology*, *299*, 172-179.

Rammstedt, B., and John, O. P. (2007). Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. *Journal of research in Personality*, *41*(1), 203-212.

Riva, P., Sacchi, S., and Brambilla, M. (2015). Humanizing machines: Anthropomorphization of slot machines increases gambling. *Journal of Experimental Psychology: Applied*, *21*(4), 313.

Rodriguez, M. C., Ooms, A., and Montañez, M. (2008). Students' perceptions of online-learning quality given comfort, motivation, satisfaction, and experience. *Journal of interactive online learning*, *7*(2), 105-125.

Rodriguez-Ruiz, A., Lång, K., Gubern-Merida, A., Broeders, M., Gennaro, G., Clauser, P., ... and Wallis, M. G. (2019). Stand-alone artificial intelligence for breast cancer detection in mammography: comparison with 101 radiologists. *JNCI: Journal of the National Cancer Institute*, *111*(9), 916-922.

Sheremeta, Roman M., Impulsive Behavior in Competition: Testing Theories of Overbidding in Rent-Seeking Contests (April 18, 2018). Available at SSRN: https://ssrn.com/abstract=2676419 or http://dx.doi.org/10.2139/ssrn.2676419

Shtulman, A., & McCallum, K. (2014). Cognitive reflection predicts science understanding. Proceedings of the 36th Annual Conference of the Cognitive Science Society, 2937–2942.

Sinkovics, R. R., Stöttinger, B., Schlegelmilch, B. B., and Ram, S. (2002). Reluctance to use technology-related products: Development of a technophobia scale. *Thunderbird International Business Review*, *44*(4), 477-494.

Swami, Viren, et al. "Analytic thinking reduces belief in conspiracy theories." Cognition 133.3 (2014): 572-585.

Toplak, Maggie E., Richard F. West, and Keith E. Stanovich. "Assessing miserly information processing: An expansion of the Cognitive Reflection Test." *Thinking and Reasoning* 20.2 (2014): 147-168.

Toplak, M. E., West, R. F., and Stanovich, K. E. (2011). The Cognitive Reflection Test as a predictor of performance on heuristics-and-biases tasks. *Memory and cognition*, *39*(7), 1275.

Vander Ark, 2018 https://www.gettingsmart.com/2018/08/32-ways-ai-is-improving-education/

Vujic, A. (2017). Switching on or switching off? Everyday computer use as a predictor of sustained attention and cognitive reflection. Computers in Human Behavior, 72, 152-162.

Yeomans, M., Shah, A., Mullainathan, S., and Kleinberg, J. (2019). Making sense of recommendations. *Journal of Behavioral Decision-making.*

Zhang, B., and Dafoe, A. (2019). Artificial intelligence: American attitudes and trends. *Available at SSRN 3312874*.

Table 1. OLS Regressions Predicting Relative Preference for Advice from Human (vs. Algorithmic) Advisor, controlling for data source and decision domain

	Preference for human advisor (vs. algorithmic) (0-100)		
	Model (1)	Model (2)	Model (3)
Standardized CRT score	**-12.396*****	**-12.341*****	**-12.239*****
	1.362	1.335	1.332
Study (Reference Group: Study 1)			
Study 2		1.524 (2.361)	1.517 (2.355)
Study 3 (Mixed decision domain)		5.506* (2.552)	0.166 (3.167)
Study 4 (Mixed decision domain)		7.901*** (2.280)	2.565 (2.971)
Study 5		1.172 (2.501)	1.168 (2.494)
Study 6		3.581 (3.162)	3.574 (3.165)
Study 7		3.362 (3.074)	3.358 (3.066)
Study 8		-0.745 (2.530)	-0.749 (2.523)
Study 9		3.929 (2.474)	3.934 (2.467)
Study 10		2.417 (2.512)	2.421 (2.505)
Study 11		1.064 (2.173)	1.071 (2.167)
Decision Domain (Reference group: Financial)			
College Admissions			4.433 (3.109)
Healthcare			11.950*** (3.073)
Employee Hiring			4.774 (3.055)
Constant	63.515***	60.785***	60.740***
	0.775	1.916	1.91
Observations	2414	2414	2414
Adjusted R^2	0.035	0.041	0.046
Residual Std. Error	24.7	24.6	24.6
	(df = 2412)	(df = 2402)	(df = 2399)
F Statistic	86.6 ***	10.4***	9.31***
	(df = 1; 2412)	(df = 11; 2402)	(df = 14; 2399)

Note. This table shows ordinary least squares (OLS) regressions predicting relative proportion of advice from a human (vs. algorithmic) advisor using an indicator of individual cognitive reflection. Negative coefficients denote preference for greater advice from algorithmic advisor. Standard errors are in parentheses. †, *, **, and *** denote significance levels of <0.10, <0.05, <0.01 and <0.001% levels, respectively.

Table 2. OLS Regressions Predicting Relative Preference for Advice from Human (vs. Algorithmic) Advisor, controlling for demographic variables

	Preference for human advisor (vs. algorithmic) (0-100)
Standardized CRT score	**-10.690*** (1.381)**
Age	-0.062 (0.05)
Gender (Reference Group: Male)	
Female	3.849*** (1.052)
Other Gender	-11.081 (7.533)
Race *(Reference group: White)*	
Black	3.04† (1.75)
American Indian or Alaska Native	4.905 (5.539)
Asian	-0.565 (1.968)
Native Hawaiian or Pacific Islander	-2.127 (11.045)
Other Race	-2.399 (2.194)
Employment status *(Reference group: Employed)*	
Unemployed and searching	2.852 (1.891)
Unemployed and not searching	-9.837 (0.636)
Retired	3.206 (2.759)
Student	-1.092 (2.474)
Political Affiliation *(Reference group: Republican)*	
Democrat	-5.533*** (1.314)
Independent	-5.018*** (1.436)
Other	-2.936 (4.282)
None	0.556 (2.889)
Income	0.000 (0.000)
Education (continuous)	-0.283 (0.395)
Constant	59.625*** (1.222)
Observations	2414
Adjusted R^2	0.047
Residual Std. Error	24.6 (df = 2394)
F Statistic	7.29*** (df = 19; 2394)

Note. This table shows ordinary least squares (OLS) regressions predicting relative proportion of advice from a human (vs. algorithmic) advisor using an indicator of individual cognitive reflection. Negative coefficients denote preference for greater advice from algorithmic advisor. Standard errors are in parentheses. All continuous predictors are mean-centered and scaled by 1 standard deviation. †, *, **, and *** denote significance levels of <0.10, <0.05, <0.01 and <0.001% levels, respectively.

Table 3. OLS Regressions Predicting Relative Preference for Advice from Human (vs. Algorithmic) Advisor, controlling for individual differences

	Model 3a	**Model 3b**	**Model 3c**	**Model 4a**	**Model 5a**	**Model 5b**
CRT Score	**-12.489*****	**-12.093*****	**-11.370*****	**-12.113*****	**-12.673*****	**-19.430*****
	(1.929)	(1.978)	(2.047)	(2.932)	(2.562)	(5.704)
Prior Experience with AI	-2.676***	-2.473***	-3.110***	-6.077***		
	(0.850)	(0.894)	(0.960)	(1.630)		
Comfort with Technology	-0.212***	-0.247***	-0.281***	-0.440**		
	(0.075)	(0.078)	(0.080)	(0.214)		
Social Anxiety		0.077	0.092			
		(0.138)	(0.144)			
BIG-5						
Agreeableness			1.230***			
			(0.442)			
Conscientiousness			1.196**			
			(0.488)			
Extraversion			0.818**			
			(0.396)			
Neuroticism			0.931**			
			(0.404)			
Openness			-0.128			
			(0.417)			
Confidence in Financial Literacy				-0.29		
				(1.224)		
Self-perceived Intelligence					-0.014	0.237**
					(0.059)	(0.119)
Mindset						0.565
						(0.423)
Constant	77.797***	76.132***	51.977***	95.907***	67.168***	45.868***
	(3.461)	(4.658)	(8.236)	(8.666)	(3.988)	(8.455)
Num.Obs.	1155	1056	966	453	720	207
R2	0.051	0.053	0.078	0.092	0.034	0.083
Adj.R2	0.048	0.05	0.069	0.084	0.031	0.069

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Note. This table shows ordinary least squares (OLS) regressions predicting relative proportion of advice from a human (vs. algorithmic) advisor using an indicator of individual cognitive reflection. Negative coefficients denote preference for greater advice from algorithmic advisor. Standard errors are in parentheses. †, *, **, and *** denote significance levels of <0.10, <0.05, <0.01 and <0.001% levels, respectively.

Table 4. OLS Regressions Predicting Relative Preference for Advice from Human (vs. Algorithmic) Advisor, controlling for advisor beliefs and individual differences

	Preference for human advisor (vs. algorithmic) (0-100)		
	Model 6a	Model 6b	Model 6c
CRT Score	**-7.167***** (1.510)	**-5.826***** (1.689)	**-7.656**** (3.734)
Advisor beliefs			
Algorithm Accuracy	-11.861*** (0.710)	-11.185*** (0.792)	-11.514*** (1.848)
Human Accuracy	12.153*** (0.872)	12.217*** (0.982)	12.788*** (1.848)
Algorithm Impartiality	2.120*** (0.558)	2.483*** (0.624)	2.630* (1.436)
Human Impartiality	-4.129*** (0.636)	-3.655*** (0.710)	-1.842 (1.620)
Algorithm Objectivity	1.233** (0.592)	1.477** (0.644)	2.182 (1.383)
Human Objectivity	-0.185 (0.568)	0.043 (0.629)	2.102 (1.507)
Advisor Preferences			
Warmth		0.965*** (0.240)	1.189** (0.493)
Competence		-0.964*** (0.270)	-1.768*** (0.646)
Desire for Control			0.045 (0.107)
Locus of Control			0.162 (0.250)
Constant	94.574*** (20.125)	72.479*** (22.856)	10.159 (54.495)
N	1052	849	197
R^2	0.433	0.456	0.482
Adjusted R^2	0.428	0.449	0.448