*Journal of Behavioral Data Science*, 2022, *2* (1), 106–123.
DOI: https://doi.org/10.35566/jbds/v2n1/p4

# The Role of Personality in Trust in Public Policy Automation

Philip D. Waggoner*[1] and Ryan Kennedy[2]

[1] Columbia University ,YouGov America
pdw2119@columbia.edu
[2] University of Houston

**Abstract.** Algorithms play an increasingly important role in public policy decision-making. Despite this consequential role, little effort has been made to evaluate the extent to which people trust algorithms in decision-making, much less the personality characteristics associated with higher levels of trust. Such evaluations inform the widespread adoption and efficacy of algorithms in public policy decision-making. We explore the role of major personality inventories – need for cognition, need to evaluate, the "Big 5" – in shaping an individual's trust in public policy algorithms, specifically dealing with criminal justice sentencing. To explore personality in this context, we fielded an original survey experiment aimed at assessing the impact of varying advice sources on forecasting criminal recidivism, conditioned by personality traits. We found strong correlations between all personality types and general levels of trust in automation, as expected. Further, we uncovered evidence that need for cognition increases the weight given to advice from an algorithm relative to humans, and "agreeableness" decreases the distance between respondents' expectations and advice from a judge, relative to advice from a crowd.

*Keywords:* Personality · Trust in automation · Public policy · Decision-making

## 1 Introduction

Algorithms are increasingly important in public policy implementation (Kennedy, Waggoner, & Ward, 2022). Algorithms assist officials in major US cities to allocate resources (O'Brien, 2015), judges in detecting gerrymandering (Bernstein & Duchin, 2017), and the military to control weapons (Scharre, 2018). Recently, algorithms have also begun to play a role in criminal sentencing, where algorithms are used by judges to inform expectations on a defendant's probability of recidivating (Waggoner & Macmillen, 2021). Such a hybrid-decision making process between humans and algorithms influences the parameters, duration, and severity of sentencing (Dressel & Farid, 2018).

Despite the rise in interest about automation and algorithms, little attention has been paid in public policy to algorithms or the psychological factors that influence trust in them. Horowitz (2016) explored situations under which people approve development of autonomous weapons systems, but this reveals little about the underlying trust people have in algorithms in practice. Further, some have debated whether individuals place low levels of trust in algorithms, "algorithm aversion" (Dietvorst, Simmons, & Massey, 2015), or high levels of trust, "algorithm bias" (Logg, 2016). Still, very little attention has been paid to how individuals' psychological characteristics might influence attitudes towards algorithms. Instead, the literature tends to focus on demographic or cultural factors (Hoff & Bashir, 2015).

To address this gap, we recently fielded a criminal sentencing survey experiment and leveraged three major inventories of psychological measures of personality to explore who is more or less trusting of algorithms: "need for cognition" (NC) (Cacioppo & Petty, 1982), "need to evaluate" (NE) (Bizer et al., 2004), and the "Big 5" (Norman, 1963).

The survey experiment was primarily interested in assessing the impact of varying advice sources (judge, algorithm, a "crowd" of peers) on respondents' forecasts of criminal recidivism. Of primary interest was the conditioning role of personality in this forecasting effort. The details of and findings from the experiment are detailed throughout the remainder of the paper.

### 1.1 Personality Inventories

The first inventory, need for cognition (NC), is associated with individuals who have a strong desire to learn and grow (Cacioppo & Petty, 1982). Some previous studies on NC in similar contexts have suggested that when high NC individuals are asked to undertake a task in which they are given little information and then provided expert advice, they are more likely to assign greater weight to that advice, rather than relying on heuristics (Leippe, Eisenstadt, Rauch, & Seib, 2004). This suggests that high NC individuals will be more likely to *take* advice insofar as they view that advice as "expert," given their more elaborate processing of information (Sicilia, Ruiz, & Munuera, 2005).

Our second, need to evaluate (NE), is associated with individuals who tend to generate and retain their own attitudes (Bizer et al., 2004). This "self-monitoring" personality is also associated with the need to control (Snyder, 1974) and constantly evaluate social surroundings (Jarvis & Petty, 1996). Such attributes induce greater reliance on intuition over outside sources. Past work has demonstrated that high NE individuals tend to make spontaneous judgements in *response* to stimuli (Tormala & Petty, 2001). This "on-line" form of information processing suggests that when people come into contact with outside information, their personality plays a key role in determining their levels of acceptance of the information. Resultant attitudes are much stronger than those in the alternative, "memory-based" processing (Bizer, Tormala, Rucker, & Petty, 2006). Other studies have also leveraged NE to explain information processing (Druckman & Nelson, 2003). For our context, we expect high NE individuals exhibit

greater reliance on their own intuition compared to other sources, which should lead to distrust. When a high NE individual is confronted by advice from an outside source, we expect these individuals to be less trusting of advice, *regardless* of its origin. This is in line with recent work suggesting that errors experienced from an algorithm provoke a stronger distrust of that advice than do errors experienced from other sources (Dietvorst et al., 2015).

For measurement, we used the two-item battery for each personality type (NC and NE), totaling four questions for both personality types. Question wording is in the Appendix. Of note, though there is a tradeoff between "internal reliability and brevity" (Gerber, Huber, Doherty, & Dowling, 2011), we opted for the smaller battery for two reasons. First, it is the exact same approach as using the common, reliable TIPI inventory to measure each of the Big 5 traits (Gosling, Rentfrow, & Swann Jr, 2003). Both approaches uses two items per personality trait to generate a measure. Second, we wanted to ensure high response rates, given the inclusion of the personality batteries in addition to our main experiment. To minimize the burden on the respondent and with the "the benefit of being short enough to be included in large political surveys," (Gerber et al., 2011, 268), we opted for the smaller battery. Ultimately, we selected these measures of personality given their widespread use in a variety of fields including political science (Jost, Glaser, Kruglanski, & Sulloway, 2003), public policy (Sargent, 2004), psychology (Cacioppo, Petty, & Feng Kao, 1984), and others (Luttrell, Petty, & Xu, 2017).

Turning now to the Big 5, we use only the "agreeableness" and "openness to experience" traits in our study as they can be most clearly linked to trust in automation. We selected only two instead of all five traits, because, as Gerber et al. (2011) note, "in most cases only some of the Big Five traits significantly predict outcomes of interest" (268). Our approach is similar to other studies on the role of the Big 5 in behavior that select only the specific personality traits that can be clearly linked to substantive phenomena (Quintelier, 2014).

For agreeableness, Gerber et al. (2011) and John and Srivastava (1999) note that "agreeableness contrasts a prosocial and communal orientation toward others with antagonism and includes traits such as altruism, tender-mindedness, trust, and modesty." Agreeableness is also associated with social conformity (Fiske, 1949) and compliance (Digman & Takemoto-Chock, 1981). In our context, being given advice from an "expert," and then asked whether they wish to update their expectation, we expect agreeable individuals should positively respond to the advice-giver, regardless of the source of advice. In an effort to conform to the reigning wisdom via the advice treatment, individuals who are high on agreeableness should trust automation, positively weight expert advice, and also align with the advice-giver.

Second, openness is associated with originality (Gerber et al., 2011; John & Srivastava, 1999), intellectual curiosity (Peabody & Goldberg, 1989), and an eagerness to learn (Barrick & Mount, 1991). As individuals who are open to experiences come into contact with outside advice in an unfamiliar realm, they

should positively respond to the advice treatment across all three measures of trust discussed below.

We leveraged the Ten-Item Personality Inventory (TIPI) (Gosling et al., 2003) to measure these traits. Two items containing personality adjectives are associated with each trait, with one phrase coded normally and the other reverse coded (e.g., for "agreeableness": *item 7* = sympathetic, warm and *item 2* (reverse coded) = critical, quarrelsome).

## 2 Method

### 2.1 Participants

We utilized Amazon's Mechanical Turk (*MTurk*) to recruit 395 subjects, each of whom were paid $2.00 for participation. MTurk is a valid, widely used platform to field similar political, psychological, and social experiments such as ours (Clifford, Jewell, & Waggoner, 2015). Additional details of the study design are included in the Appendix.

### 2.2 Procedure

Our study contains observational (general trust) and experimental (behavioral impact) components. For the observational component, respondents were given an eight-item battery of questions related to degrees of trust in automation (Kim, Ferrin, & Rao, 2008). Respondents were asked their level of agreement on a scale from 1 (strongly disagree) to 7 (strongly agree) for statements like, "Using algorithms improves the output quality for organizations." These were aggregated into a 7 point scale where 7 indicates high trust in algorithms, while 1 indicates low trust. The wording for all of the items is available in the supporting information. This scale is the dependent variable for the first stage of the analysis, which is analyzed using OLS regression and presented in Table 1.

For the experimental component, respondents were asked to forecast the probability of a defendant committing *another* crime within two years for one of eight real, randomly selected criminal profiles based on criminal history and defendant demographic characteristics. Then, the respondent was given "advice" from a source (listed below), and asked whether they wanted to update predictions or leave them the same (manual entry required both times). The shifts in respondents' predictions (or lack thereof) is the quantity of interest in our study. We included two attention checks throughout to minimize satisficing (Hauser & Schwarz, 2016). Specifically, respondents were warned if they missed one attention check, and then were removed and not paid if they failed both. About 80% of respondents who attempted the survey passed the checks and completed the survey.

The presentation of our criminal profiles mimics the formatting of Dressel and Farid (2018), which was shown to be a sufficient amount of detail for an average MTurk participant to make an informed judgment, with expected accuracy

similar to the popular "COMPAS" algorithm. The full wording is available in the supporting information. We randomly selected 20 pre-trial defendants from the 2013-14 from Broward County, FL database, who all had a risk scores between 2 and 8 (derived from the COMPAS algorithm, which ranked defendants from 1 to 10, with 10 being the most likely to recidivate). This pool of defendants was winnowed when the crime involved was obscure, and then reduced again randomly to reduce the task burden on respondents, which left us with eight profiles.

For each profile, respondents were randomly assigned to one of the three advice conditions: judge with 10 years of experience in criminal sentencing; computer algorithm designed by computer scientists and criminal justice officials; average of a previous survey of 300 Turkers. The treatment conditions are coded as separate dummy variables for whether the individual saw advice from an algorithm or a judge in the scenario, with the previous MTurk survey as the baseline condition. And, in addition to the main personality predictors, we control for several common factors in public policy experiments, including age, education, gender, and partisanship.

We evaluate two measures of trust. The first measure is "weight of advice" (Gino & Moore, 2007; Logg, 2016). This variable is calculated as $\mid u_{2i} - u_{1i} \mid / \mid a_i - u_{1i} \mid$, where $u_{2i}$ is respondent $i$'s final assigned probability for recidivism, $u_{1i}$ is their initial prediction, and $a_i$ is the advice they were given from one of the sources. A score of 1 suggests the respondent only used the advice from the source, where as 0.5 suggests they weighted the source and their prediction equally, and 0 means the respondent ignored the advice. Our second measure is the average distance to advice, measured as $\mid a_i - u_{ri} \mid$. Lower values indicate that there was less distance between the respondent's final forecast and the advice they were given.

We modeled the weight and distance measures by fitting multilevel regressions to the data after pooling across all criminal profiles and specifying varying intercepts for defendant descriptions and respondent. Multilevel models were chosen to account for unobserved heterogeneity on both the individual respondent and scenario level. This provides an efficient and accurate estimates for experiments where respondents evaluate multiple, different scenarios (Gelman & Hill, 2006). The model was specified as

$$y_{ijk} = a_{ijk} + \zeta_j + \phi_k + \beta * X + e_{ijk} \tag{1}$$

where $a_{ij}$ is the overall intercept, $\zeta_j \sim N(0,1)$ is the random intercept based on the defendant description, $\phi_k \sim N(0,1)$ is the random intercept based on the individual respondent, $\beta$ is an array of coefficients for the treatments $X$, and $e_{ij}$ is the error term. Results are presented in Table 2.

## 3 Results

For the observational analysis, note the significance and large magnitudes of effects for all personality indicators in the top four rows of Table 1.[1] High NE respondents are less likely to trust algorithms ($\beta = -0.14$), compared to those higher on NC ($\beta = 0.25$), agreeableness ($\beta = 0.04$), and openness ($\beta = 0.07$), all of whom are eager to learn. The latter group of respondents is more trusting of automation in line with expectations.

**Table 1.** The Impact of Personality on Trust in Automation

| | *Dependent variable:* | |
|---|---|---|
| | Trust in Automation | |
| | (1) | (2) |
| Need for Cognition | $0.245^{***}$ (0.023) | |
| Need to Evaluate | $-0.136^{***}$ (0.021) | |
| Agreeableness | | $0.040^{**}$ (0.016) |
| Openness to Experience | | $0.066^{***}$ (0.015) |
| Age | $0.005^{***}$ (0.002) | $0.007^{***}$ (0.002) |
| Education | $0.131^{***}$ (0.029) | $-0.020$ (0.040) |
| Female | $-0.212^{***}$ (0.035) | $-0.336^{***}$ (0.040) |
| Partisanship | $-0.047^{***}$ (0.009) | $-0.028^{***}$ (0.010) |
| Algorithm Condition | $-0.000$ (0.00000) | $-0.000$ (0.00000) |
| Judge Condition | $-0.000$ (0.00000) | 0.000 (0.00000) |
| Constant | $3.783^{***}$ (0.124) | $4.047^{***}$ (0.173) |
| N | 3,022 | 2,233 |
| Log Likelihood | 32,311.620 | 24,075.100 |
| Akaike Inf. Crit. | $-64{,}599.240$ | $-48{,}126.190$ |
| Bayesian Inf. Crit. | $-64{,}527.070$ | $-48{,}057.660$ |

*Note:* $^{*}$p<0.1; $^{**}$p<0.05; $^{***}$p<0.01

The experimental stage exploring the impact of personality on behavioral tasks seen in Table 2 and Figure 1. Here our N is higher because each respondent evaluated 8 defendant profiles. NC plays a strong conditioning role in the relative weight respondents' assign to advice across both "expert" conditions in comparison to the baseline category. The degree to which NC conditions trust in *algorithms* is nearly doubled that of the *judge* condition ($\beta = 0.09$ compared to $\beta = 0.05$). Further, the weight effect is opposite for NE individuals in the algorithm condition ($\beta = -0.04$), and indistinguishable from zero in the

[1] Of note, the trust in automation index is measured at the individual-level, not the scenario-level. Hence the larger N in the tables, relative to the number of individual recruited subjects.

judge condition. Results are similar for high openness personality types in their weighting of algorithmic advice relative to humans.

**Table 2.** The Impact of Personality on Behavior

| | *Dependent variable:* | | | |
|---|---|---|---|---|
| | Weight | Distance | Weight | Distance |
| | (1) | (2) | (3) | (4) |
| Need for Cognition | $-0.047^{***}$ (0.018) | $1.687^{*}$ (0.892) | | |
| Need to Evaluate | 0.012 (0.017) | −0.964 (0.839) | | |
| Agreeableness | | | 0.018 (0.013) | −0.501 (0.642) |
| Openness | | | −0.017 (0.012) | −0.126 (0.606) |
| Age | 0.0002 (0.001) | 0.008 (0.043) | 0.001 (0.001) | −0.012 (0.051) |
| Education | 0.010 (0.019) | $-1.879^{**}$ (0.857) | 0.010 (0.022) | −0.571 (1.008) |
| Female | 0.015 (0.020) | −0.678 (0.930) | −0.001 (0.025) | 0.758 (1.111) |
| Partisanship | 0.004 (0.005) | −0.187 (0.231) | 0.005 (0.006) | −0.128 (0.279) |
| Algorithm Cond. | −0.072 (0.079) | −1.701 (4.101) | 0.120 (0.097) | $-9.706^{*}$ (5.081) |
| Judge Cond. | 0.006 (0.083) | −1.969 (4.307) | −0.030 (0.097) | 1.395 (5.144) |
| Alg. x NC | $0.093^{***}$ (0.024) | $-3.007^{**}$ (1.251) | | |
| Alg. x NE | $-0.035^{*}$ (0.021) | $1.905^{*}$ (1.106) | | |
| Judge x NC | $0.054^{**}$ (0.022) | $-1.929^{*}$ (1.168) | | |
| Judge x NE | −0.033 (0.022) | 1.645 (1.140) | | |
| Alg. x Agreeable | | | −0.024 (0.016) | 0.610 (0.851) |
| Alg. x Openness | | | $0.028^{*}$ (0.015) | 0.228 (0.801) |
| Judge x Agreeable | | | $0.031^{*}$ (0.016) | $-2.216^{**}$ (0.876) |
| Judge x Openness | | | −0.005 (0.016) | 1.275 (0.854) |
| Constant | $0.225^{**}$ (0.093) | $30.664^{***}$ (6.182) | 0.076 (0.115) | $32.606^{***}$ (7.002) |
| N | 3,022 | 3,022 | 2,233 | 2,233 |
| Log Likelihood | −403.901 | −12,702.400 | −349.032 | −9,388.203 |
| Akaike Inf. Crit. | 839.802 | 25,436.800 | 730.063 | 18,808.410 |
| Bayesian Inf. Crit. | 936.021 | 25,533.020 | 821.441 | 18,899.780 |

$^{*}$p<0.1; $^{**}$p<0.05; $^{***}$p<0.01

Notably, NC strongly conditions trust in algorithms, but less so compared to advice from crowds or human experts. This is seen most clearly when comparing panels (a) and (b) in Figure 1. The gaps between the fit lines are more distinct in the algorithm condition (a) compared to the judge condition (b). There is only a modest distinction at the tails in the judge condition.

Across all treatment conditions, the advice given by all three sources was the same, and we found no differences when we presented values centered around those derived from the COMPAS algorithm or when the advice was randomly chosen.

Of note, for the weight and distance multilevel models, to test if there was any impact of varying the treatments by scenario or by respondent, we ran-

domly allocated half of our respondents to each type of assignment. We found no difference (i.e., no detection of study purpose) in the results.

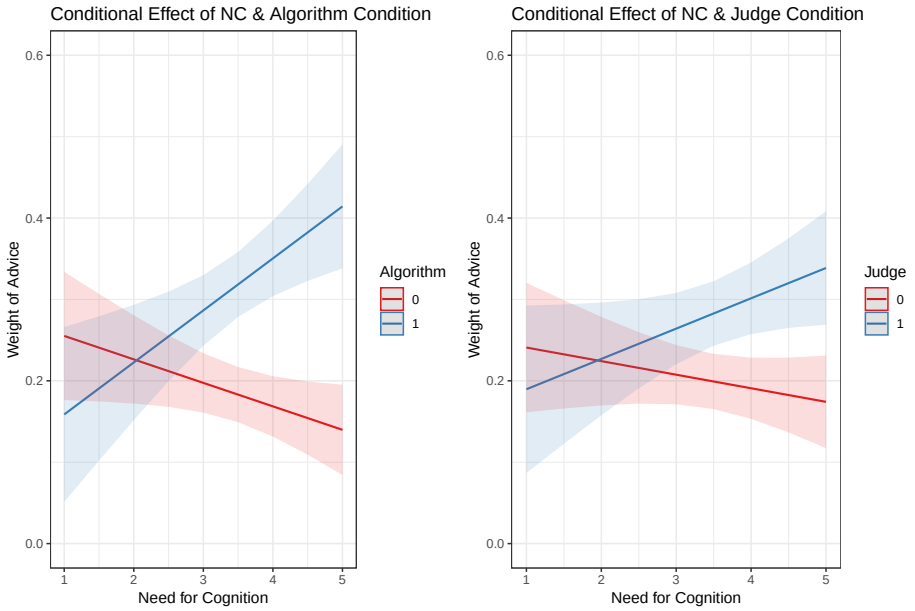


**Figure 1.** Conditional Impacts of NfC on Behavior

## 4 Discussion

Overall, we found that personality influences trust in automation, as well as behavioral tasks related to public policy decision-making. In the first stage, there was a pronounced impact of personality on general levels of trust. In line with research finding high levels of trust in algorithms (Goddard, Roudsari, & Wyatt, 2011), the significant conditioning role of these personality inventories suggests that personalities associated with intellectual curiosity, agreeableness, openness to advice-givers, as well as being highly aware of environments and more skeptical are strongly associated with levels of trust in automation. The former group comprised of individuals who are more accepting of new information and experiences is more trusting, while the latter group, who tends to be threatened by exogenous sources of information, is less trusting.

Regarding changes in respondents' *behavioral* indicators of trust, high NC individuals are much more trusting of algorithms than of the wisdom of the crowd or, to a lesser extent, a human expert. And the NE personality trait, which we expected to be threatened by exogenous advice, weighs advice from algorithms

less than human advice sources. Surprisingly, no effects were observed for agreeable individuals in the algorithmic advice condition for weighting, though there were weak effects for judges. Strikingly, high NE and NC individuals reacted to algorithmic advice over all other advice sources, though the effect size is nearly doubled for NC individuals compared to NE individuals and is more statistically stable. We also saw a significant effect of personality conditioning behavior in decision making tasks, especially related to their trust in advice from an algorithm.

Though a blend of significant and null results, we remain encouraged by our findings for two reasons. First, we uncovered strong evidence of personality influencing behavior and general levels of trust in automation, in line with our main goal. Given the newness of this topic, these results are useful for motivating future work on trust in automation and personality. Second, in line with Gerber et al. (2011), it would be unrealistic to expect *all* personality measures to explain *all* behavior. Of the Big 5 they note, "these traits have predictive power in an impressive variety of domains but are not universal predictors of all outcomes" (268). Our results corroborate this sentiment that personality plays a role in trust in automation, though it does not explain the breadth of general trust *and* behavior.

Regarding generalizability, while people generally trust algorithmic advice relative to other advice sources, levels of trust are influenced by personality traits. As not all people retain the same personalities, not all people equally trust algorithms to make consequential decisions.

## 5 Limitations and Future Directions

While we offer a starting place for future work on personality and trust in automation, a key limitation of our study is focusing only on criminal justice. Should we expect similar results in other subfields, such as automation in medicine, for example? Also, though Dietvorst et al. (2015) demonstrate trust in algorithms wanes when mistakes are introduced, this phenomenon may be more likely for high NE individuals relative to high NC individuals, given the starting place of skepticism for high NE individuals. Further, algorithm aversion may not be detectable for high NC individuals, while it may drive levels of trust for high NE individuals. Or, do the other three Big 5 personality traits (extroversion, conscientiousness, and emotional stability) impact trust in automation? In sum, we suggest researchers in this realm consider personalities to provide a fuller picture of trust in a variety of subfields.

An additional limitation that may be addressed in future work is the nature of MTurk respondents in general, in that they are typically higher educated and more liberal for example (Berinsky, Huber, & Lenz, 2012; Clifford et al., 2015), and thus may be more likely to trust automation. Such a possibility suggests the potential for future and different samples to yield potentially different results. More analysis and experiments in this vein would deepen the impact of our initial findings in this research.

## 6 Concluding Remarks

In this study, we have demonstrated that personality plays a strong role in impacting individuals' levels of trust in automation as they make public policy decisions. We bring psychology into the trust in automation discussion for several reasons. First, such an approach offers a baseline for understanding the role of innate, heritable characteristics and their influence on trust in automation.[2] Such an understanding makes it clearer where to look for greater or lesser trust in algorithms, and where the basis of trust lies. These psychological characteristics are also widely used in many fields to describe human behavior both inside (Sargent, 2004) and outside (Hill, Foster, Sofko, Elliott, & Shelton, 2016) of public policy. Given the rapid increase of algorithms and algorithmic advice in everyday life (Logg, 2016), the role of psychological characteristics conditioning virtually all human behavior (Eysenck, 1963), and also the recent surge in research on algorithmic transparency (Rudin & Ustun, 2018), our study offers a timely exploration of the intersection of trust in automation and personality.

## Acknowledgement

The authors thank Scott Clifford for his comments on a previous version of this paper. All errors are the authors'. This research is based upon work supported in part by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), via 2017-17061500008. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of ODNI, IARPA, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright annotation therein.

## References

Barrick, M. R., & Mount, M. K. (1991). The big five personality dimensions and job performance: a meta-analysis. *Personnel psychology*, *44*(1), 1–26. doi: https://doi.org/10.1111/j.1744-6570.1991.tb00688.x

Berinsky, A. J., Huber, G. A., & Lenz, G. S. (2012). Evaluating online labor markets for experimental research: Amazon. com's mechanical turk. *Political analysis*, *20*(3), 351–368. doi: https://doi.org/10.1093/pan/mpr057

Bernstein, M., & Duchin, M. (2017). A formula goes to court: Partisan gerrymandering and the efficiency gap. *Notices of the AMS*, *64*(9). doi: https://doi.org/10.1090/noti1573

[2] In using the terms "innate" and "heritable," we are referring to the vast work on the stability of personality throughout one's life (Caspi, Roberts, & Shiner, 2005), the heritable nature of personality (Van Gestel & Van Broeckhoven, 2003), the genetic aspects of personality (Canli, 2008), and even the biological link with personality (DeYoung et al., 2010).

Bizer, G. Y., Krosnick, J. A., Holbrook, A. L., Christian Wheeler, S., Rucker, D. D., & Petty, R. E. (2004). The impact of personality on cognitive, behavioral, and affective political processes: The effects of need to evaluate. *Journal of personality*, *72*(5), 995–1028. doi: https://doi.org/10.1111/j.0022-3506.2004.00288.x

Bizer, G. Y., Tormala, Z. L., Rucker, D. D., & Petty, R. E. (2006). Memory-based versus on-line processing: Implications for attitude strength. *Journal of Experimental Social Psychology*, *42*(5), 646–653. doi: https://doi.org/10.1016/j.jesp.2005.09.002

Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of personality and social psychology*, *42*(1), 116. doi: https://doi.org/10.1037/0022-3514.42.1.116

Cacioppo, J. T., Petty, R. E., & Feng Kao, C. (1984). The efficient assessment of need for cognition. *Journal of personality assessment*, *48*(3), 306–307. doi: https://doi.org/10.1207/s15327752jpa4803_13

Canli, T. (2008). Toward a molecular psychology of personality. *Handbook of personality: Theory and research*, 311–327.

Caspi, A., Roberts, B. W., & Shiner, R. L. (2005). Personality development: Stability and change. *Annu. Rev. Psychol.*, *56*, 453–484. doi: https://doi.org/10.1146/annurev.psych.55.090902.141913

Clifford, S., Jewell, R. M., & Waggoner, P. D. (2015). Are samples drawn from mechanical turk valid for research on political ideology? *Research & Politics*, *2*(4), 2053168015622072. doi: https://doi.org/10.1177/2053168015622072

DeYoung, C. G., Hirsh, J. B., Shane, M. S., Papademetris, X., Rajeevan, N., & Gray, J. R. (2010). Testing predictions from personality neuroscience: Brain structure and the big five. *Psychological science*, *21*(6), 820–828. doi: https://doi.org/10.1177/0956797610370159

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General*, *144*(1), 114. doi: https://doi.org/10.1037/xge0000033

Digman, J. M., & Takemoto-Chock, N. K. (1981). Factors in the natural language of personality: Re-analysis, comparison, and interpretation of six major studies. *Multivariate behavioral research*, *16*(2), 149–170. doi: https://doi.org/10.1207/s15327906mbr1602_2

Dressel, J., & Farid, H. (2018). The accuracy, fairness, and limits of predicting recidivism. *Science Advances*, *4*(1), 55–80. doi: https://doi.org/10.1126/sciadv.aao5580

Druckman, J. N., & Nelson, K. R. (2003). Framing and deliberation: How citizens' conversations limit elite influence. *American Journal of Political Science*, *47*(4), 729–745. doi: https://doi.org/10.1111/1540-5907.00051

Eysenck, H. (1963). *The biological basis of personality*. Routledge. doi: https://doi.org/10.4324/9781351305280

Fiske, D. W. (1949). Consistency of the factorial structures of personality ratings

from different sources. *The Journal of Abnormal and Social Psychology*, *44*(3), 329. doi: https://doi.org/10.1037/h0057198

Gelman, A., & Hill, J. (2006). *Data analysis using regression and multilevel/hierarchical models*. Cambridge university press. doi: https://doi.org/10.1017/cbo9780511790942

Gerber, A. S., Huber, G. A., Doherty, D., & Dowling, C. M. (2011). The big five personality traits in the political arena. *Annual Review of Political Science*, *14*. doi: https://doi.org/10.1146/annurev-polisci-051010-111659

Gino, F., & Moore, D. A. (2007). Effects of task difficulty on use of advice. *Journal of Behavioral Decision Making*, *20*(1), 21–35.

Goddard, K., Roudsari, A., & Wyatt, J. C. (2011). Automation bias: a systematic review of frequency, effect mediators, and mitigators. *Journal of the American Medical Informatics Association*, *19*(1), 121–127. doi: https://doi.org/10.1136/amiajnl-2011-000089

Gosling, S. D., Rentfrow, P. J., & Swann Jr, W. B. (2003). A very brief measure of the big-five personality domains. *Journal of Research in personality*, *37*(6), 504–528. doi: https://doi.org/10.1016/s0092-6566(03)00046-1

Hauser, D. J., & Schwarz, N. (2016). Attentive turkers: Mturk participants perform better on online attention checks than do subject pool participants. *Behavior research methods*, *48*(1), 400–407.

Hill, B. D., Foster, J. D., Sofko, C., Elliott, E. M., & Shelton, J. T. (2016). The interaction of ability and motivation: Average working memory is required for need for cognition to positively benefit intelligence and the effect increases with ability. *Personality and Individual Differences*, *98*, 225–228. doi: https://doi.org/10.1016/j.paid.2016.04.043

Hoff, K. A., & Bashir, M. (2015). Trust in automation: Integrating empirical evidence on factors that influence trust. *Human Factors*, *57*(3), 407–434.

Horowitz, M. C. (2016). Public opinion and the politics of the killer robots debate. *Research & Politics*, *3*(1), 2053168015627183.

Jarvis, W. B. G., & Petty, R. E. (1996). The need to evaluate. *Journal of personality and social psychology*, *70*(1), 172. doi: https://doi.org/10.1037/0022-3514.70.1.172

John, O. P., & Srivastava, S. (1999). The big five trait taxonomy: History, measurement, and theoretical perspectives. *Handbook of personality: Theory and research*, *2*(1999), 102–138.

Jost, J. T., Glaser, J., Kruglanski, A. W., & Sulloway, F. J. (2003). Political conservatism as motivated social cognition. *Psychological bulletin*, *129*(3), 339. doi: https://doi.org/10.4324/9781315175867-5

Kennedy, R. P., Waggoner, P. D., & Ward, M. M. (2022). Trust in public policy algorithms. *The Journal of Politics*, *84*(2). doi: https://doi.org/10.1086/716283

Kim, D. J., Ferrin, D. L., & Rao, H. R. (2008). A trust-based consumer decision-making model in electronic commerce: The role of trust, perceived risk, and their antecedents. *Decision support systems*, *44*(2), 544–564. doi: https://doi.org/10.1016/j.dss.2007.07.001

Leippe, M. R., Eisenstadt, D., Rauch, S. M., & Seib, H. M. (2004). Timing of eyewitness expert testimony, jurors' need for cognition, and case strength as determinants of trial verdicts. *Journal of Applied Psychology*, *89*(3), 524. doi: https://doi.org/10.1037/0021-9010.89.3.524

Logg, J. M. (2016). *When do people rely on algorithms?* (Unpublished doctoral dissertation). University of California, Berkeley.

Luttrell, A., Petty, R. E., & Xu, M. (2017). Replicating and fixing failed replications: The case of need for cognition and argument quality. *Journal of Experimental Social Psychology*, *69*, 178–183. doi: https://doi.org/10.1016/j.jesp.2016.09.006

Norman, W. T. (1963). Toward an adequate taxonomy of personality attributes: Replicated factor structure in peer nomination personality ratings. *The Journal of Abnormal and Social Psychology*, *66*(6), 574. doi: https://doi.org/10.1037/h0040291

O'Brien, D. T. (2015). Custodians and custodianship in urban neighborhoods: A methodology using reports of public issues received by a city's 311 hotline. *Environment and Behavior*, *47*(3), 304–327.

Peabody, D., & Goldberg, L. R. (1989). Some determinants of factor structures from personality-trait descriptors. *Journal of personality and social psychology*, *57*(3), 552. doi: https://doi.org/10.1037/0022-3514.57.3.552

Quintelier, E. (2014). The influence of the big 5 personality traits on young people's political consumer behavior. *Young Consumers*, *15*(4), 342–352. doi: https://doi.org/10.1108/yc-09-2013-00395

Rudin, C., & Ustun, B. (2018). Optimized scoring systems: Toward trust in machine learning for healthcare and criminal justice. *Interfaces*, *48*(5), 449–466. doi: https://doi.org/10.1287/inte.2018.0957

Sargent, M. J. (2004). Less thought, more punishment: Need for cognition predicts support for punitive responses to crime. *Personality and Social Psychology Bulletin*, *30*(11), 1485–1493. doi: https://doi.org/10.1177/0146167204264481

Scharre, P. (2018). *Army of none: Autonomous weapons and the future of war*. New York: WW Norton and Company.

Sicilia, M., Ruiz, S., & Munuera, J. L. (2005). Effects of interactivity in a web site: The moderating effect of need for cognition. *Journal of advertising*, *34*(3), 31–44. doi: https://doi.org/10.1080/00913367.2005.10639202

Snyder, M. (1974). Self-monitoring of expressive behavior. *Journal of personality and social psychology*, *30*(4), 526. doi: https://doi.org/10.1037/h0037039

Tormala, Z. L., & Petty, R. E. (2001). On-line versus memory-based processing: The role of "need to evaluate" in person perception. *Personality and social psychology bulletin*, *27*(12), 1599–1612. doi: https://doi.org/10.1177/01461672012712004

Van Gestel, S., & Van Broeckhoven, C. (2003). Genetics of personality: are we making progress? *Molecular psychiatry*, *8*(10), 840. doi: https://doi.org/10.1038/sj.mp.4001367

Waggoner, P. D., & Macmillen, A. (2021). Pursuing open-source de-

velopment of predictive algorithms: the case of criminal sentencing algorithms. *Journal of Computational Social Science*, 1–21. doi: https://doi.org/10.1007/s42001-021-00122-y

# Appendix

## A Support Information

### A.1 Task Wording

It has generally been found that even untrained individuals can do very well, sometimes even better than trained people or computer algorithms, at determining the likelihood of a person committing another crime after their initial arrest.

We are interested in knowing whether this accuracy can be further improved by combining individual judgement with the advice of crowds, experts, or algorithms. In what follows, you will be given an actual arrest record for a person arrested in Broward County, Florida. We already know whether the person committed another crime within the next two years. You will be asked to give us a probability of the person re-offending along the following lines.

We have collected advice from several sources:

- Several Mechanical Turk surveys of people like yourself.
- A judge with over 10 years of experience.
- A machine learning algorithms, developed by computer scientists and criminal justice experts, that use historic recidivism data to predict probability of re-offending.

**Warning: There are attention checks in this survey. We reserve the right to deny payment if a participant fails these checks, as that indicates the participant is not actually doing the tasks.**

### A.2 Defendant Profiles

The defendant is a male aged 22. They have been charged with: Possession of Cocaine. This crime is classified as a felony. They have been convicted of 0 prior crimes. They have 0 juvenile felony charges and 0 juvenile misdemeanor charges on their record.

The defendant is a male aged 38. They have been charged with: Manufacturing Cannabis/Marijuana. This crime is classified as a felony. They have been convicted of 3 prior crimes. They have 0 juvenile felony charges and 0 juvenile misdemeanor charges on their record.

The defendant is a male aged 23. They have been charged with: Grand Theft. This crime is classified as a felony. They have been convicted of 3 prior crimes. They have 0 juvenile felony charges and 0 juvenile misdemeanor charges on their record.

The defendant is a male aged 27. They have been charged with: Possession of Meth. This crime is classified as a felony. They have been convicted of 5 prior crimes. They have 0 juvenile felony charges and 0 juvenile misdemeanor charges on their record.

The defendant is a male aged 24. They have been charged with: Driving with a Revoked License. This crime is classified as a felony. They have been convicted of 2 prior crimes. They have 0 juvenile felony charges and 0 juvenile misdemeanor charges on their record.

The defendant is a female aged 33. They have been charged with: Child Neglect. This crime is classified as a felony. They have been convicted of 1 prior crimes. They have 0 juvenile felony charges and 0 juvenile misdemeanor charges on their record.

The defendant is a male aged 22. They have been charged with: Disorderly Conduct. This crime is classified as a misdemeanor. They have been convicted of 0 prior crimes. They have 0 juvenile felony charges and 0 juvenile misdemeanor charges on their record.

The defendant is a male aged 24. They have been charged with: Resisting an Officer with Violence. This crime is classified as a felony. They have been convicted of 0 prior crimes. They have 0 juvenile felony charges and 0 juvenile misdemeanor charges on their record.

### A.3 Examples of Treatment

A group of 200 people recruited from Mechanical Turk, on average rated the defendant as 80% likely to commit another felony crime within the next two years.

Previously, you forecast that the defendant was [RESPONDENT'S PREVIOUS FORECAST] likely to commit another felony crime within the next two years.

**If you would like to update your forecast, you can do so now. If not, just enter the same numbers as you entered previously.**

A judge with more than 10 years of experience rated the defendant as 80% likely to commit another felony crime within the next two years.

Previously, you forecast that the defendant was [RESPONDENT'S PREVIOUS FORECAST] likely to commit another felony crime within the next two years.

**If you would like to update your forecast, you can do so now. If not, just enter the same numbers as you entered previously.**

An algorithm developed by computer scientists and criminal justice researchers, based on a statistical analysis of thousands of past defendant records, rated the defendant as 80% likely to commit another felony crime within the next two years.

Previously, you forecast that the defendant was [RESPONDENT'S PREVIOUS FORECAST] likely to commit another felony crime within the next two years.

**If you would like to update your forecast, you can do so now. If not, just enter the same numbers as you entered previously.**

### A.4 MTurk Study Specifics

In addition to the specifics of the design included in the manuscript, below are some additional specific items related to fielding the study on MTurk:

1. **Approval Rate**: HIT Approval Rate > 95%
2. **Location**: United States
3. **Study Description**: Respondents will be asked to evaluate a series of real criminal profiles and asked to predict the likelihood of recidivism with and without the help of advice.
4. **Keywords**: survey, criminal justice, forecasting, predication

### A.5 Personality Inventories

#### A.5.1 NC and NE

Please indicate the extent to which these statements are characteristic or uncharacteristic of you (On a scale from 1 to 5, with being extremely characteristic and 5 being extremely uncharacteristic).

1. I have opinions about almost everything.
2. I like having responsibility for handling situations that require a lot of thinking.
3. It is very important to me to hold strong opinions.
4. I often prefer to remain neutral about complex issues.

#### A.5.2 TIPI for Big 5

Here are a number of personality traits that may or may not apply to you. Please indicate the extent to which you agree or disagree that these characteristics apply to you. You should rate the extent to which the pair of traits applies to you, even if one characteristic applies more strongly than the other. (On a scale from 1 to 7, with 1 = strongly disagree and 7 = strongly agree).

1. Extroverted, enthusiastic
2. Critical, quarrelsome
3. Dependable, self-disciplined
4. Anxious, easily upset
5. Open to new experiences, complex
6. Reserved, quiet
7. Sympathetic, warm
8. Disorganized, careless
9. Calm, emotionally stable
10. Conventional, uncreative

### A.6 Trust in Automation Index

Many organizations now use algorithms to make forecasts. Some high profile examples include the use of statistics in baseball to choose players (Moneyball) or Nate Silvers use of statistics to predict elections. To what extent do you agree or disagree with the following statements about algorithms? (On a scale from 1 to 7, with 1 = strongly agree and 7 = strongly disagree). Given the variance in valence, items were coded so that the highest end of the response range (7) indicates *high* trust in automation and the lowest end of the range (1) indicates *low* trust.

1. Using algorithms increases the chances of organizations achieving their goals.
2. Using algorithms increases the effectiveness of organizations in making good decisions.
3. Using algorithms improves the output quality for organizations.
4. Using algorithms makes it more likely for organizations to make errors.
5. Modern organizations rely too much on algorithms to make decisions about the future.
6. Using algorithms is an effective way to overcome human biases.
7. When I am uncertain about something, I will trust the information from an algorithm in place of my own judgement.
8. When I am uncertain about something, I will tend to trust my own intuition and judgement over the information from an algorithm.

### A.7 Base Relationships: Empirical Motivation

As an empirical motivation for our full study, we offer a short discussion of our base findings of relative influence of the treatment conditions in the experiment. We show the impact of advice from an algorithm or a judge relative to the baseline category of average past MTurk respondents for our two behavioral measures of trust in Table 3: advice weight and distance to advice. The strong positive impacts from the first model (column 1) for each condition suggest respondents are reacting to the advice, with the magnitude of the effect in the algorithm condition nearly twice that of the judge condition. Second, the pronounced negative effects in the second model (column 2) demonstrate the impact of the algorithm and judge treatments on reducing the distance between respondents' predictions and the advice-giver relative to the baseline category. Similarly, the effects are nearly doubled in the algorithm condition.

These results demonstrate two things. First, respondents were significantly more likely to change their evaluations based on the advice of "experts," whether human or machine-derived than they were to trust the "wisdom of the crowd." And second the algorithm condition is where the strongest effects are observed, suggesting respondents trust algorithms to a greater degree than advice from humans. This is an important finding by itself, and one we explore in greater detail in a separate paper.

**Table 3.** Experimental Impacts on Dependent Variables of Interest

| | *Dependent variable:* | |
|---|---|---|
| | Advice Weight | Distance to Advice |
| | (1) | (2) |
| Algorithm Condition | $0.134^{***}$ (0.013) | $-6.563^{***}$ (0.864) |
| Judge Condition | $0.073^{***}$ (0.013) | $-3.812^{***}$ (0.857) |
| Constant | $0.156^{***}$ (0.009) | $27.353^{***}$ (0.608) |
| Observations | 3,274 | 3,274 |
| $R^2$ | 0.031 | 0.017 |
| Adjusted $R^2$ | 0.030 | 0.017 |
| Residual Std. Error (df = 3271) | 0.305 | 20.113 |
| F Statistic (df = 2; 3271) | $52.076^{***}$ | $29.117^{***}$ |

$^{*}$p<0.1; $^{**}$p<0.05; $^{***}$p<0.01