

# ALGORITHM AVERSION: SENSITIVITY TO INTERVENTIONS AND THE RELATIONSHIP WITH NUMERACY<sup>1</sup>

**Michał Dzieżyk\***

**SWPS University of Social Sciences and Humanities**

**Weronika Hetmańczuk\*\***

**SWPS University of Social Sciences and Humanities**

**Jakub Traczyk\*\*\***

**SWPS University of Social Sciences and Humanities**

***Abstract:** The main goal of this research was to investigate whether people exhibit algorithm aversion—a tendency to avoid using an imperfect algorithm even if it outperforms human judgments—in the case of estimating students' percentile scores on a standardized math test. We also explored the relationships between numeracy and algorithm aversion and tested two interventions aimed at reducing algorithm aversion. In two studies, we asked participants to estimate the percentiles of 46 real 15-year-old Polish students on a standardized math test. Participants were offered the opportunity to compare their estimates with the forecasts of an algorithm—a statistical model that predicted real percentile scores based on five explanatory variables (i.e., gender, repeating a class, the number of pages read before the exam, the frequency of playing online games, socioeconomic status). Across two studies, we demonstrated that even though the predictions of the statistical model were closer to students' percentile scores, participants were less likely to rely on the statistical model predictions in*

<sup>1</sup> This research was funded by Wrocławskie Centrum Akademickie (FAST program BWU-22/2020/F2) and SWPS University. We want to thank Jakub Figol for his support in this project.

\* Michał Dzieżyk, SWPS University of Social Sciences and Humanities, Faculty of Psychology in Wrocław, Center for Research on Improving Decision Making (CRIDM), [www.cridm.edu.pl](http://www.cridm.edu.pl), Ostrowskiego 30b, 53-238 Wrocław, Poland

\*\* Weronika Hetmańczuk, SWPS University of Social Sciences and Humanities, Faculty of Psychology in Wrocław, Center for Research on Improving Decision Making (CRIDM), [www.cridm.edu.pl](http://www.cridm.edu.pl), Ostrowskiego 30b, 53-238 Wrocław, Poland

\*\*\* Jakub Traczyk, corresponding author, SWPS University of Social Sciences and Humanities, Faculty of Psychology in Wrocław, Center for Research on Improving Decision Making (CRIDM), [www.cridm.edu.pl](http://www.cridm.edu.pl), Ostrowskiego 30b, 53-238 Wrocław, Poland, e-mail: [jtraczyk@swps.edu.pl](mailto:jtraczyk@swps.edu.pl)

making forecasts. We also found that higher statistical numeracy was related to a higher reluctance to use the algorithm. In Study 2, we introduced two interventions to reduce algorithm aversion. Depending on the experimental condition, participants either received feedback on statistical model predictions or were provided with a detailed description of the statistical model. We found that people, especially those with higher statistical numeracy, avoided using the imperfect algorithm even though it outperformed human judgments. Interestingly, a simple intervention that explained how the statistical model works led to better performance in an estimation task.

**Key words:** algorithm aversion, numeracy, estimation, augmented decisions, good decisions.

### **AWERSJA DO ALGORYTMÓW: WRAŻLIWOŚĆ NA INTERWENCJE ORAZ ZWIĄZEK Z POZIOMEM ZDOLNOŚCI NUMERYCZNYCH**

**Streszczenie:** Głównym celem tego projektu było zweryfikowanie hipotezy dotyczącej niechęci do korzystania z algorytmów (ang. algorithm aversion) – tendencji do unikania stosowania niedoskonałego algorytmu nawet wtedy, gdy w swoich przewidywaniach przewyższa on ludzkie sądy. W tym celu posłużyliśmy się przykładem wyników w teście z matematyki. Dodatkowo zbadaliśmy związki między zdolnościami numerycznymi, a niechęcią do algorytmów oraz zweryfikowaliśmy skuteczność dwóch interwencji, które miały na celu zmniejszenie awersji do algorytmów. W dwóch badaniach poprosiliśmy uczestników o oszacowanie wyników centylowych 46 piętnastoletnich uczniów w standaryzowanym teście z matematyki. Uczestnicy mogli oszacować wyniki samodzielnie lub na podstawie prognoz modelu statystycznego, który przewidywał rzeczywiste wyniki w oparciu o pięć predyktorów (płeć, powtarzanie zajęć, liczba stron przeczytanych przed egzaminem, częstość grania w gry online oraz status ekonomiczny). W obu badaniach wykazaliśmy, że osoby badane rzadziej polegały na przewidywaniach algorytmu, mimo że oszacowania modelu statystycznego były bliższe rzeczywistym wynikom uczniów niż oszacowania uczestników badań. Wykazaliśmy także, że osoby o większych zdolnościach numerycznych wykazywały silniejszą niechęć do korzystania z prognoz algorytmu. W drugim badaniu testowaliśmy skuteczność dwóch interwencji mających na celu zmniejszenie awersji do algorytmów. W zależności od warunku badawczego uczestnicy otrzymywali informacje zwrotne na temat przewidywań modelu lub mogli zapoznać się ze szczegółowym opisem modelu statystycznego. Wykazaliśmy, że proste objaśnienie działania modelu doprowadziło do lepszego szacowania wyników przez badanych.

**Słowa kluczowe:** *awersja do algorytmów, zdolności numeryczne, oszacowanie, wspomaganie decyzji, dobre decyzje.*

## **ALGORITHM AVERSION: SENSITIVITY TO INTERVENTIONS AND THE RELATIONSHIP WITH NUMERACY**

In a fast-changing world of new technologies, people are often confronted with an overwhelming amount of information. Various algorithms have been developed to improve judgments and decisions in such complex and dynamic environments (Fry, 2018). For example, laypersons who intend to invest money in stocks can consult with statistical models predicting future revenues of specific companies; music lovers are able to explore new albums suggested by an algorithm learning from their previous music preferences, and clinicians can use tailored medical algorithms to reach a diagnosis. Although algorithms are often better than humans in making accurate forecasts, it has been documented that people exhibit algorithm aversion — a tendency to avoid using an imperfect algorithm even if it outperforms human judgments (Dietvorst et al., 2018). In this study, we investigated algorithm aversion in the case of estimating students' percentile scores on a standardized math test. In contrast to other similar studies, we additionally explored the role of numeracy in avoiding algorithmic aids and experimentally tested two interventions aimed at reducing algorithm aversion.

### **Algorithm aversion**

Research comparing the effectiveness of human and algorithmic predictions has shown that algorithms regularly surpass humans (Dawes et al., 1989; Meehl, 1954) in a wide variety of domains (Grove et al., 2000): from forecasting product sales (Fildes et al., 2009) to medical diagnosing and management decision skills (Esmacilzadeh et al., 2015; Inthorn et al., 2015; Prahel et al., 2013). Despite this, people still decide to take advice from a human counselor than from the algorithmic one, thus frequently diminishing the value of the algorithmic superiority (Önköl et al., 2009). This irrational devaluation of algorithmic advice has its roots in the debate between clinical and actuarial psychology (Meehl, 1954). In the 1950's, algorithms were created as a solution for our cognitive limitations. Simple models were outperforming experts during psychological examination (Dawes, 1979), but many clinical psychologists and researchers were still against the statistical model, drawing their conclusions from traditional computing and diagnosing methods (Meehl, 1954). Today, the algorithm

devaluation effect raised in a seminal work by Meehl (1954) is known as algorithm aversion (Dietvorst et al., 2015).

The problem of algorithm aversion lies in defining the cause of the phenomenon and most of the explanations are based on anecdotes rather than empirical evidence (Dietvorst et al., 2015). The reasons for algorithm aversion incorporate the desire for perfect forecasts (Dawes, 1979; Einhorn, 1986; Highhouse, 2008), problems with learning (Dawes, 1979), the ability of human predictions to improve through experience (Highhouse, 2008), unwillingness to take into account the unique individual circumstances such as medical examination (Longoni et al., 2019), individual locus of control (Shaffer et al., 2013), and the widely understood need for human-like approach (Diab et al., 2011).

Dietvorst, Simmons, and Massey (2015) found that participants preferred to make their own predictions rather than rely on an algorithm that made mistakes. Even after the algorithms were seen as better in finding obvious errors, humans were seen as better in learning from mistakes and finding exceptions to the rule or detect challenging to understand “diamonds in the rough.” In another study, Dietvorst, Simmons, and Massey (2018) also found that if participants had a chance to slightly modify the outcome of algorithmic predictions, it made them more satisfied and tolerant of error. Additionally, their trust in superior algorithm predictions increased, and they were more likely to choose subsequent forecasts by the algorithm. These findings improve the potential to overcome algorithm aversion by introducing different interventions aimed at enhancing comprehension of an algorithmic mechanism and its predictions (e.g., explaining how an algorithm works or providing feedback on algorithmic predictions). Nevertheless, learning the algorithm and understanding it may be moderated by different individual factors (e.g., cognitive styles, Green & Hughes, 1986). In the present study, we decided to explore one of such factors related to cognitive abilities.

## Numeracy

Statistical numeracy — the ability to understand and use probabilistic and mathematical concepts — is regarded as one of the most robust predictors of superior decision making (Cokely et al., 2012, 2018; Garcia-Retamero et al., 2019; Ghazal et al., 2014; Sobków et al., 2020; Sobkow, Olszewska, et al., 2020; Weller et al., 2013). It has been repeatedly demonstrated that people with high numeracy, in comparison to people with low numeracy, make better decisions in both medical (Petrova et al., 2016, 2017) and financial (Estrada-Mejia et al., 2016; Estrada-Mejia

et al., 2020) domains. Interestingly, the positive relationship between numeracy and better decision making holds even when other cognitive abilities such as cognitive reflection and fluid intelligence are controlled (Peters & Bjälkebring, 2015; Sobkow, Garrido, et al., 2020; Sobkow, Olszewska, et al., 2020). Several studies have provided evidence for different psychological mechanisms that may explain more accurate judgments and better decisions made by people with high numeracy. For example, such individuals process probabilities (Millroth & Juslin, 2015; Petrova et al., 2014, 2019; Traczyk & Fulawka, 2016) and values (Schley & Peters, 2014) in a more linear and consistent (Traczyk, Fulawka, et al., 2020) way; have a better memory for outcomes and numerical information (Peters & Bjälkebring, 2015; Shoots-Reinhard et al., 2020; Sobkow, Olszewska, et al., 2020); deliberate more and explore a decision problem to a greater extent (Ashby, 2017; Cokely & Kelley, 2009; Jasper et al., 2017; Traczyk, Lenda, et al., 2018), employing their experience for judgments and choices (Traczyk, Lenda, et al., 2018; Traczyk, Sobkow, et al., 2020); use decision strategies adaptively (Jasper et al., 2013; Traczyk, Sobkow, et al., 2018); tend to draw different (generally stronger or more precise) affective meaning from numbers and numerical comparisons (Peters, 2012; Peters et al., 2006).

Interestingly, statistical numeracy is positively related to other facets of numerical abilities (Sobkow et al., 2019), such as subjective numeracy/numerical confidence (Fagerlin et al., 2007) and approximate numeracy (Peters & Bjälkebring, 2015), but it does not mean that it is a unitary construct always predicting decision-related variables in the same direction. For instance, Peters et al. (2019) demonstrated that more objectively numerate participants experienced a greater proportion of positive financial outcomes only when they were more numerically confident (i.e., scored higher in the subjective numeracy test). Furthermore, Sobkow et al. (2020) revealed that people with high approximate numeracy were more successful in avoiding negative decision outcomes in real-life while individuals with high subjective numeracy/numerical confidence experienced more negative outcomes (even if the relationship was controlled for statistical numeracy and other cognitive abilities).

Taking these findings into account, we can expect the two possible effects of numeracy on algorithm aversion. On the one hand, if people with high statistical numeracy have low subjective numeracy (i.e., they are less confident about their numerical abilities), such individuals might rely on algorithmic predictions more often (Logg et al., 2019). On the other hand, high statistical numeracy accompanied by high subjective numeracy might be associated with a greater algorithm aversion because of higher trust in individual numerical abilities and lower trust in algorithmic predictions.

## Overview

In two studies reported in this paper, we investigated algorithm aversion using an original statistical model predicting real students' percentile scores on a standardized math test. For this purpose, we asked participants to estimate percentiles of real students' scores. Participants were informed that they could provide estimates by themselves or use predictions of an 'imperfect' algorithm. We hypothesized that people would exhibit algorithm aversion, that is, they would tend to reject a statistical model's predictions. We also explored how people with high or low numeracy would rely on statistical model predictions to make forecasts about students' percentile scores. Finally, we tested two interventions aimed at reducing algorithm aversion. Specifically, we expected that interventions involving feedback on the algorithm's performance and the explanation of the statistical model would decrease reluctance to use algorithms.

## STUDY 1

### THE STATISTICAL MODEL PREDICTING PERCENTILE SCORES ON A STANDARDIZED MATH TEST

We developed a statistical model using data collected in the OECD's Program for International Student Assessment (PISA). PISA measures 15-year-olds' skills in reading, mathematics, and science (particularly, the role of these skills in solving real-life problems). We used data from PISA 2018 (the seventh round of the international assessment). Firstly, we selected data from Polish students ( $N = 5,625$ ) who completed PISA 2018. Secondly, based on exploratory correlation analyses, we selected five variables that were significant predictors of math scores (i.e., gender, repeating a class, the number of pages read before the exam, the frequency of playing online games, socioeconomic status). Additionally, we filtered out all observations with missing data in these predictors, resulting in a total of  $N = 4,639$  observations. Thirdly, we randomly divided dataset into three groups consisting of  $N = 3,526$ ,  $N = 1,067$ , and  $N = 46$  observations. These data sets were used to train the model, estimate prediction accuracy, and test participants' predictions in empirical studies, respectively.

The algorithm we employed in the current research was based on a linear regression model predicting percentiles of 3,526 real 15-year-old students on a standardized PISA math test (Table 1; see Table S1 in the Appendix for correlation coefficients

among predictors). The model explained  $R^2 = .30$  variance of the outcome variable; all five variables were significant predictors of the outcome variable (all  $ps < .001$ ).

**Table 1**  
*Regression model predicting the performance (i.e., the percentile) in a standardized math test being a part of PISA 2018 (narrowed down to a sample of Polish students)*

Estimate	<i>b</i>	<i>SE</i>	<i>t</i>	<i>p</i>
Intercept	2.19	2.10	1.04	.297
Gender (female)	6.75	0.83	8.11	< .001
Frequency of playing online games	-4.29	0.29	-14.69	<.001
Number of pages read before the exam	5.50	0.30	18.52	<.001
Repeating a class	-22.91	3.22	-7.13	<.001
Socioeconomic status	8.87	0.41	21.80	<.001

Prediction accuracy of the model was tested using an independent sample of  $N = 1,067$  students from the same PISA data set. We found that the model predictions were off by 20 percentiles on average from the real students' percentile scores. Additionally, in a pilot study, we found that participants ( $N = 24$ ) who were asked to estimate the percentiles of 30 real students performed worse than the model with a mean absolute deviation from the real students' percentile scores of  $M = 22.79$  ( $SD = 16.37$ ).

### Participants

One hundred native Polish speakers ( $M_{age} = 25.3$ ,  $SD_{age} = 7.9$ , 44 females) recruited from the Prolific participant's pool took part in this study. All participants gave informed consent. They were informed that the study is voluntary and anonymous.

### Materials and procedure

We informed participants that their task is to estimate the percentiles scores of a sample of 46 real 15-year-old Polish students on a standardized math test. We briefly explained how the percentiles should be interpreted using an illustrative example. We instructed participants that their estimates should be based on five features describing each of 46 students, explicitly presented and explained. The features were: 1) gender (male, female), 2) whether a student repeated a class (yes, no), 3) the number of pages read before the exam (five levels from 1 or fewer pages to 500 or more), 4) the frequency of playing online games (five levels from never to every day), 5) socioeconomic status (five levels from very low to very high).

Additionally, participants were informed that, based on these features and data collected from a large sample of real students, we developed a statistical model to support predicting the percentiles on a standardized math test. Similar to previous research by Dietvorst et al. (2018), we highlighted that the model is imperfect<sup>2</sup>, and its prediction error is +/- 20 percentiles. Having confirmed that all information about the task and the model are clear, participants started a training session. In this stage of the study, participants learned the experimental task and predictions of the model. In particular, 16 vignettes (Figure 1), including information about features describing 16 students, were displayed on separate screens along with the model predictions. Participants were asked to familiarize themselves with each vignette and type the model predictions (i.e., the percentile of a given student predicted by the model and presented along with other information about a student).

Id	355
Płeć	m
Powtarzanie klasy	nie
Ilość przeczytanych stron do egzaminu	między 101 a 500
Granie w gry online	prawie codziennie
Status socjoekonomiczny	wysoki
Wynik	91

**Figure 1.** Exemplary vignette presented to participants in a training session. In the test session, information about model prediction (i.e., Wynik) was unavailable. Płeć (gender), Powtarzanie klasy (repeating a class), Ilość przeczytanych stron do egzaminu (the number of pages read before the exam), Granie w gry online (the frequency of playing online games), Status socjoekonomiczny (socioeconomic status).

In the next stage of the study, participants were instructed to estimate the percentiles of the other group of 30 students using the same features as in the training session. However, there was no information available about the predictions of the model. Instead, participants had to decide whether they wanted to estimate each student's percentile by themselves or to use model prediction. If they decided individually, they were asked to type their estimate on the next screen. In contrast,

<sup>2</sup> We introduced information about the *imperfect* algorithm and its *prediction error* to clearly assure participants that their estimates can be more accurate than those provided by the statistical model. If the algorithm provided only correct estimates, then there would be no point for participants to make individual predictions.



if participants chose to use the model predictions, they saw the model's percentile estimate on the next screen.

There was no time limit for this task. Vignettes were presented in random order for each participant. The participants were not provided with feedback on their performance (i.e., how their estimate or the model estimate deviated from the real percentiles).

To measure numeracy, we used the Berlin Numeracy Test (BNT, Cokely et al., 2012). This test consists of four mathematical tasks that capture risk literacy, statistical numeracy, and comprehension of probability. BNT was always administered after the estimation task at the end of the study.

### **Data preprocessing and statistical approach**

For each participant estimating the percentile for each of 30 students, we extracted the number of decisions to choose the model predictions or make an individual prediction. Next, we computed absolute deviance from real students' percentile scores in the math test. Specifically, we subtracted each estimate (irrespective of whether it was made by the model or by the individual) from the real student's percentile score. A higher absolute value of this outcome indicated that participants' estimates deviated more from the percentile score.

The first variable served as an indicator of the algorithm aversion (i.e., whether people tend to use forecasts of an algorithm or decide on their own), whereas the latter indicated an accuracy of estimates (irrespective of the previous decision).

We tested our predictions using Bayesian multilevel regression analysis (specifying unique intercepts for participants) performed in the brms package (Bürkner, 2018) that uses the STAN programming language. For each model, we run four chains with 10,000 iterations. We burned the first 2,000 samples of the iterations. To regularize estimation of the models and rule out unreasonable parameter values, we used weakly informative priors for coefficients [Normal(0,10)] and a Half-Cauchy(10) prior for residual variation (Nalborczyk et al., 2019). We diagnosed chains for convergence using Gelman–Rubin statistics, visual inspection, and autocorrelation plots. We found that the sampling procedure was efficient (values of the Gelman–Rubin statistic were  $R < 1.01$ ).

## Results

Descriptive statistics for the measures used in Study 1 are presented in Table 2.

**Table 2**

*Descriptive statistics for the measures used in Study 1. The decision to use an algorithm is a dichotomous variable (1 – yes, 0 – no)*

	<i>M</i>	<i>SD</i>	<i>Min</i>	<i>Max</i>
BNT	1.81	1.24	0	4
Decision to use algorithm	0.33	0.27	0	1
Mean deviance from real percentile scores	21.16	2.82	14.18	31.67

In the first analysis, we fitted a varying intercept model (Bernoulli model with logit link function) with a participants' decision to use the statistical model predictions as an outcome variable and mean-centered BNT as a predictor (Table 3). The model explained  $R^2 = .31$  of variance. We found that participants tended to avoid using the statistical model predictions,  $a = -0.90$ , 95% CrI = [-1.25, -0.54]. Participants chose the statistical model instead of making an individual decision in 33% of cases on average. BNT was negatively related to these decisions,  $b = -0.29$ , 95% CrI = [-0.58, 0.00]. Additional analysis of posterior probability distribution showed that 97% of this distribution is below 0, suggesting that there is a 97% probability that people with higher numeracy are more likely to forecast students' percentile scores on their own.

**Table 3**

*Posterior mean, standard error, 95% credible interval, and Rhat statistic for each parameter of the model (with varying intercept by subject) predicting decision to use the algorithm (1 – yes, 0 – no) in Study 1*

Parameter	Mean	SE	CrI lower	CrI higher	Rhat
Intercept	-0.9	0.18	-1.25	-0.54	1.00
BNT	-0.29	0.15	-0.58	0	1.00
$\sigma_{\text{subject}}$	1.71	0.16	1.43	2.05	1.00

In the second analysis (Gaussian model), we predicted performance in the estimation task, which was operationalized as a mean absolute deviation from real students' percentile scores irrespective of whether participants provided estimations by themselves or used predictions of the statistical model. Lower numbers of this variable indicated better performance (i.e., participants provided percentile estimated closer to real students' percentile scores). We found that BNT was not a credible predictor of performance (Table 4), but the decision to use the statistical model to predict real students' percentile scores was related to better estimates of

the percentiles,  $b = -6.50$ , 95% CrI =  $[-7.66, -5.34]$ . The model explained  $R^2 = .04$  of variance.

**Table 4**  
*Posterior mean, standard error, 95% credible interval, and Rhat statistic for each parameter of the model (with varying intercept by subject) predicting performance in the estimation task in Study 1*

Parameter	Mean	SE	CrI lower	CrI higher	Rhat
Intercept	23.29	0.34	22.62	23.97	1.00
BNT	0.00	0.23	-0.45	0.46	1.00
Decision (to use the statistical model predictions)	-6.50	0.60	-7.66	-5.34	1.00
$\sigma_{\text{subject}}$	0.43	0.31	0.02	1.17	1.00
$\sigma_e$	15.33	0.20	14.95	15.73	1.00

Because features of the students were visible to subjects prior to their choices whether to use the algorithm or not (and consequently, participants' decisions/estimations might be influenced depending on the features of the student), we also fitted models including all features of each student as control variables. The general pattern of results and main conclusions remained the same (see Tables S2 and S3 in the Appendix), suggesting that specific characteristics of students presented on vignettes did not impact our findings.

## STUDY 2

In Study 1, we demonstrated that people avoided using predictions of the statistical model in making forecasts about the percentile of real students' percentile scores on a standardized math test. People with higher statistical numeracy were less likely to use the statistical model in their predictions. However, numeracy was not related to better performance in the estimation task. It was the decision to rely on the statistical model that led to better estimates. In Study 2, we further explored the factors associated with algorithm aversion. Specifically, we introduced another measure of numeracy (i.e., subjective numeracy; Fagerlin et al., 2007), which can be understood as a measure of numeric confidence (Peters et al., 2019). Additionally, we introduced two experimental manipulations that are likely to increase trust in statistical models. In the first experimental condition, we explained how the statistical model works, while in the second experimental condition, we provided direct feedback on the algorithm's predictions.

## Participants

One hundred and eighty-one native Polish speakers ( $M_{\text{age}} = 26.3$ ,  $SD = 8.6$ , 38 females) recruited from the Prolific participant's pool took part in this study. All participants gave informed consent. They were informed that the study is voluntary and anonymous. Additionally, participants were informed that three people who will provide the most accurate estimates would receive an additional bonus payment.

## Materials and procedure

We employed an experimental design similar to Study 1. Participants were asked to estimate the percentiles of 46 real 15-years-old Polish students on a standardized math test. The task was explained in the same way as in Study 1.

In contrast to Study 1, participants were randomly assigned to one of the three conditions. The control condition was identical to the training session of Study 1. That is, participants learned the model predictions by processing vignettes describing 16 students. In each case, participants were asked to type the percentile of each student that was explicitly provided in a vignette. In the *feedback* condition, participants were instructed to provide percentile estimates, but they did not have access to estimates recommended by the model (as in the control condition). Instead, they provided percentile estimates by themselves, followed by feedback information indicating predictions of the statistical model on the next screen. In the *description* condition, participants were provided with a detailed explanation of the statistical model. They were asked to read a brief description of all predictors/features attentively and focus on their contribution to math test scores. This information was presented in infographics. Additionally, a visual reminder of a model was displayed throughout the training session (Figure 2).



**Figure 2.** A graphical illustration of the statistical model demonstrating the character of relationships between predictors and the outcome variable (see Figure 1 for the translation)

The test session was identical to the session run in Study 1. That is, participants decided whether they wanted to choose the model predictions or determine the percentile by themselves.

Finally, we measured statistical numeracy with BNT. Additionally, we asked participants to complete the 8-item subjective numeracy test (SNS, Fagerlin et al., 2007), a measure of people’s beliefs about their skills performing mathematical operations, preferences for using numbers, and numerical confidence. These measures were administered at the end of the procedure.

**Results**

Descriptive statistics for the measures used in Study 2 are presented in Table 5.

**Table 5**  
*Descriptive statistics for the measures used in Study 2. The decision to use an algorithm is a dichotomous variable (1 – yes, 2 – no)*

Measure	Condition	N	M	SD	Minimum	Maximum
BNT	control	59	1.70	1.07	0	4
	description	55	1.71	1.29	0	4
	feedback	67	1.69	1.21	0	4
SNS	control	59	32.76	6.73	15	43
	description	55	33.62	5.41	17	43
	feedback	67	32.55	6.67	17	47
Decision to use algorithm	control	59	0.30	0.32	0	1
	description	55	0.32	0.28	0	1
	feedback	67	0.30	0.31	0	1
Mean deviance from real percentile scores	control	59	21.38	3.45	15.89	30.98
	description	55	20.16	3.21	15.10	34.30
	feedback	67	21.42	3.04	16.39	31.56

Similarly to Study 1, we fitted a varying intercept model with a participants’ decision to use the statistical model predictions as an outcome variable and mean-centered BNT and SNS as predictors, and dummy coded experimental conditions (with a control condition as a reference category; Table 6). The model explained  $R^2 = .41$  of variance. We found that participants tended to avoid using the statistical model predictions,  $\alpha = -1.55$ , 95% CrI = [-2.23, -0.90]. Participants chose the statistical model instead of making an individual decision in 31% of cases on average. BNT was negatively related to these decisions,  $b = -0.35$ , 95% CrI = [-0.70, 0.01]. There was a 97% probability that people with higher statistical numeracy would avoid basing their estimates on the statistical model predictions. In contrast, SNS was positively associated with the tendency to using the statistical model, but this effect was not credible,  $b = 0.04$ , 95% CrI = [-0.03, 0.11]. Neither feedback,  $b = -0.01$ , 95% CrI = [-0.91, 0.89], nor the description/explanation of the statistical model,  $b = 0.29$ , 95% CrI = [-0.64, 1.24], influenced decisions.

**Table 6**

*Posterior mean, standard error, 95% credible interval, and Rhat statistic for each parameter of the model (with varying intercept by subject) predicting decision to use the algorithm (1 – yes, 0 – no) in Study 2*

Parameter	Mean	SE	CrI lower	CrI higher	Rhat
Intercept	-1.55	0.34	-2.23	-0.9	-1.55
BNT	-0.35	0.18	-0.7	0.01	-0.35
SNS	0.04	0.04	-0.03	0.11	0.04
Description condition	0.29	0.48	-0.64	1.24	0.29
Feedback condition	-0.01	0.46	-0.91	0.89	-0.01
$\sigma_{\text{subject}}$	2.43	0.18	2.10	2.81	1.00

In the second analysis, we again predicted performance in the estimation task (Table 7). The model explained  $R^2 = .04$  of variance. We found that neither BNT,  $b = -0.13$ , 95% CrI = [-0.54, 0.29], nor SNS,  $b = 0.05$ , 95% CrI = [-0.03, 0.13], were credible predictors of performance. Again, the decision to use the statistical model to predict real students' percentile scores was related to better estimates of the percentiles,  $b = -6.47$ , 95% CrI = [-7.39, -5.54]. In comparison to the control condition, people in the description condition were substantially more accurate in estimating percentiles,  $b = -1.14$ , 95% CrI = [-2.22, -0.05]. We did not find a similar effect in the case of feedback,  $b = 0.05$ , 95% CrI = [-0.98, 1.09].

**Table 7**

*Posterior mean, standard error, 95% credible interval, and Rhat statistic for each parameter of the model (with varying intercept by subject) predicting performance in the estimation task in Study 2*

Parameter	Mean	SE	CrI lower	CrI higher	Rhat
Intercept	23.33	0.41	22.53	24.14	1.00
BNT	-0.13	0.21	-0.54	0.29	1.00
Decision (to use the statistical model predictions)	-6.47	0.47	-7.39	-5.54	1.00
SNS	0.05	0.04	-0.03	0.13	1.00
Description condition	-1.14	0.55	-2.22	-0.05	1.00
Feedback condition	0.05	0.53	-0.98	1.09	1.00
$\sigma_{\text{subject}}$	0.48	0.34	0.02	1.25	1.00
$\sigma_e$	15.90	0.15	15.61	16.20	1.00

Similarly to Study 1, we also fitted additional two models, including features of students as control variables, showing that these covariates did not influence the general pattern of results (Table S4 and Table S5).

## DISCUSSION

We demonstrated that participants (especially those with higher statistical numeracy) tended to avoid the statistical model predictions in making forecasts about students' percentile scores on a standardized math test even though the predictions of the statistical model were closer to real percentile. The level of algorithm aversion stayed strong across two studies and neither feedback nor descriptive intervention diminished it. Nevertheless, we demonstrated that a simple intervention explaining how the statistical model works led to better performance in an estimation task irrespective of algorithm aversion, which could be a promising direction of future research.

In a recent systematic review of algorithm aversion, Burton, Stain, and Jensen (Burton et al., 2020) identified and categorized possible causes of why people are unwilling to use superior but imperfect algorithms. These categories cover five themes, such as expectations and expertise, decision autonomy, incentivization, cognitive compatibility, and divergent rationalities. We believe that at least three themes can address the causes of algorithm aversion demonstrated in our research. In reference to the first theme (expectations and expertise), participants might utilize their past experience to make predictions of students' percentile scores. That is, years of experience at different education stages might induce an erroneous impression of being an expert in this area. Consequently, participants asked to make predictions of percentile scores on a math test were more likely to consult their subjective experience from elementary/high school rather than rely on the algorithmic aid. Additionally, the decision autonomy of our participants was restricted. In other words, they did not have control over algorithmic aid, resulting in a lack of decision control, opportunity to interact with the statistical model or calibrate it, which is a crucial factor in overcoming algorithm aversion (Dietvorst et al., 2018). Finally, the statistical model we used in our research could not be compatible with the cognitive processes underlying human statistical reasoning. While the model followed consecutive steps of a linear combination of features describing real students, participants could use different decision strategies. Such an explanation is exceptionally plausible among people with high numeracy, who are more likely to employ a greater number of various decision strategies from fast and frugal heuristics to elaborative and sophisticated decision rules, depending on task requirements and personal goals (Cokely et al., 2018; Cokely & Kelley, 2009; Traczyk, Sobkow, et al., 2018). Future studies should directly address these possible explanations and emphasize theory integration.

Our study results can also be interpreted in light of recent findings by Dietvorst and Bharti (2020). These authors provided ample evidence that people in uncertain

domains preferred human judgment to better-performing algorithmic decision aids. This effect can be explained by a diminishing sensitivity to forecasting error. People who exhibit diminishing sensitivity to forecasting error face a smaller subjective penalty for increasing error produced by the forecast. In consequence, such people would take risk, endeavoring to provide near-perfect forecasts. To do so, they would prefer forecast methods that have high variability, which may, on the one hand, result in producing near-perfect answers, and on the other hand, very inaccurate answers. In this sense, algorithm aversion is not about deciding between human judgment vs. algorithmic aid but rather between forecast methods of different accuracy variability.

In our study, participants with high numeracy avoided using the statistical model in their forecasts. It could be due to their diminishing sensitivity to forecasting errors (Dietvorst & Bharti, 2020). Highly numerate individuals who understand and use probabilistic and mathematical concepts may perceive themselves as better in numerical estimation than a statistical model and they want to provide the best possible estimation. Even though the statistical model outperforms human judgments on average, people with high numeracy may decide to take greater risk and make forecasts on their own to provide a near-perfect answer. As a result, they avoid a lower-variance statistical model.

We did not find evidence suggesting that our experimental manipulations decreased algorithm aversion. Neither feedback on the algorithm's estimates nor the explanation of how the algorithm influenced reluctance to use the algorithm. Interestingly, the latter intervention was related to better performance in the estimation task. We speculate that this manipulation might have influenced comprehension, algorithm literacy, or trust in the statistical model. Furthermore, participants might have used predictions of the statistical model adaptively. That is, during the learning process, they might identify cases in which the algorithm led to better estimations than human judgment. Based on implicit learning or intuitive abilities (Sobkow et al., 2018), some people could take advantage of additional information about the statistical mechanism underlying the algorithm. Due to the limitations of our experimental design, we are not able to address this hypothesis explicitly.

Despite the fact that our general findings and conclusions remain stable across the two studies, some open questions emerged as a consequence of the methods we used in our study. Firstly, it is essential to note that we are not able to assess the level of instruction comprehension in our study precisely. We controlled whether the participants understood the instruction using only one 'yes-no' question. Although all participants in Study 1 and Study 2 confirmed that the description of the algorithm and the instruction of the task were clear and understandable to them, it does not



imply that all participants understood the instructions to the same extent. For example, we explained that the statistical model is probabilistic and is off by 20 percentiles on average from students' percentile scores. However, some participants, especially those with low numeracy, could interpret this information as a deviation in percentage points or percent deviations. As a consequence, they could employ different strategies to make decisions. Furthermore, following previous studies (Dietvorst et al., 2018) we informed participants that the algorithm is imperfect. Such information, yet essential (see footnote on p. 74), could evoke negative attitudes towards the algorithm that might have an impact on the decisions to use it.

Secondly, in Study 2, in comparison to Study 1, participants had a chance to receive an additional bonus payment for making top predictions. This additional reward could result in a higher likelihood of making individual predictions. Additionally, highly numerate participants who wanted to get additional bonus payment could make individual decisions more often. In this sense, performance measures could be distorted by monetary reward. Nevertheless, in light of the data from Study 1 (with no reward) and Study 2 (with reward), we did not find substantial differences in algorithm avoidance tendency, performance in estimation task and the effect of numeracy on these outcome variables. Regardless of bonus payment, the results of Study 1 and Study 2 are very similar. Therefore, we cannot state that the remuneration influenced the observed effects. Nevertheless, these considerations (e.g., instruction comprehension and bonus reward) are important and should be taken into account in future research.

Thirdly, a selection of predictors for building the statistical model might raise some concerns. Since they are correlated, the algorithm's predictive performance may not be uniform across the predictor space. In other words, a specific configuration of some predictors can be more informative (e.g., it is less likely that girls play online games more often than boys) to make a decision whether to use the algorithm or not. We tried to rule out an alternative explanation of the algorithm aversion by performing additional statistical analyses that included students' features as control variables in the model. We confirmed that these additional predictors did not change the main conclusions of this research. Even though the algorithm was imperfect in its predictions and could be biased by specific configurations of students' features, it was still better than humans in predicting students' percentile scores.

To summarize, we demonstrated that algorithm aversion is a robust phenomenon. People underused algorithmic aid to make predictions about real students' percentile scores on a standardized math test even if the statistical model outperformed human judgment. Such a tendency was stronger among people with high statistical numeracy. Although brief interventions had no direct effect on overcoming algorithm aversion,

explaining how the algorithm works might influence the overall performance in estimation tasks.

## REFERENCES

- Ashby, N.J.S. (2017). Numeracy predicts preference consistency : Deliberative search heuristics increase choice consistency for choices from description and experience. *Judgement and Decision Making*, 12(2), 128–139.
- Bürkner, P.-C. (2018). Advanced Bayesian Multilevel Modeling with the R Package brms. *The R Journal*, 10(1), 395–411. <https://doi.org/10.32614/RJ-2018-017>
- Burton, J.W., Stein, M., & Jensen, T.B. (2020). A systematic review of algorithm aversion in augmented decision making. *Journal of Behavioral Decision Making*, 33(2), 220–239. <https://doi.org/10.1002/bdm.2155>
- Cokely, E.T., Feltz, A., Ghazal, S., Allan, J.N., Petrova, D., & Garcia-Retamero, R. (2018). Decision making skill: From intelligence to numeracy and expertise. In K.A. Ericsson, R.R. Hoffman, A. Kozbelt, & A.M. Williams (Eds.), *The Cambridge Handbook of Expertise and Expert Performance* (2nd ed., pp. 476–505). Cambridge University Press.
- Cokely, E.T., Galesic, M., Schulz, E., Ghazal, S., & Garcia-Retamero, R. (2012). Measuring risk literacy: The Berlin numeracy test. *Judgment and Decision Making*, 7(1), 25–47. <http://journal.sjdm.org/11/11808/jdm11808.html>
- Cokely, E.T., & Kelley, C.M. (2009). Cognitive abilities and superior decision making under risk : A protocol analysis and process model evaluation. *Judgement and Decision Making*, 4(1), 20–33. <http://journal.sjdm.org/81125/jdm81125.pdf>
- Dawes, R.M. (1979). The robust beauty of improper linear models in decision making. *American Psychologist*, 34(7), 571–582. <https://doi.org/10.1037/0003-066X.34.7.571>
- Dawes, R.M., Faust, D., & Meehl, P. (1989). Clinical versus actuarial judgment. *Science*, 243(4899), 1668–1674. <https://doi.org/10.1126/science.2648573>
- Diab, D.L., Pui, S.-Y., Yankelevich, M., & Highhouse, S. (2011). Lay Perceptions of Selection Decision Aids in US and Non-US Samples. *International Journal of Selection and Assessment*, 19(2), 209–216. <https://doi.org/10.1111/j.1468-2389.2011.00548.x>
- Dietvorst, B.J., & Bharti, S. (2020). People Reject Algorithms in Uncertain Decision Domains Because They Have Diminishing Sensitivity to Forecasting Error. *Psychological Science*, 31(10), 1302–1314. <https://doi.org/10.1177/0956797620948841>
- 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), 114–126. <https://doi.org/10.1037/xge0000033>
- Dietvorst, B.J., Simmons, J.P., & Massey, C. (2018). Overcoming Algorithm Aversion: People Will Use Imperfect Algorithms If They Can (Even Slightly) Modify Them. *Management Science*, 64(3), 1155–1170. <https://doi.org/10.1287/mnsc.2016.2643>
- Einhorn, H.J. (1986). Accepting Error to Make Less Error. *Journal of Personality Assessment*, 50(3), 387–395. [https://doi.org/10.1207/s15327752jpa5003\\_8](https://doi.org/10.1207/s15327752jpa5003_8)

- Esmaeilzadeh, P., Sambasivan, M., Kumar, N., & Nezakati, H. (2015). Adoption of clinical decision support systems in a developing country: Antecedents and outcomes of physician's threat to perceived professional autonomy. *International Journal of Medical Informatics*, 84(8), 548–560. <https://doi.org/10.1016/j.ijmedinf.2015.03.007>
- Estrada-Mejia, C., de Vries, M., & Zeelenberg, M. (2016). Numeracy and wealth. *Journal of Economic Psychology*, 54, 53–63. <https://doi.org/10.1016/j.joep.2016.02.011>
- Estrada-Mejia, C., Peters, E., Dieckmann, N.F., Zeelenberg, M., De Vries, M., & Baker, D.P. (2020). Schooling, numeracy, and wealth accumulation: A study involving an agrarian population. *Journal of Consumer Affairs*, 54(2), 648–674.
- Fagerlin, A., Zikmund-Fisher, B.J., Ubel, P.A., Jankovic, A., Derry, H.A., & Smith, D.M. (2007). Measuring numeracy without a math test: development of the Subjective Numeracy Scale. *Medical Decision Making*, 27(5), 672–680. <https://doi.org/10.1177/0272989X07304449>
- Fildes, R., Goodwin, P., Lawrence, M., & Nikolopoulos, K. (2009). Effective forecasting and judgmental adjustments: an empirical evaluation and strategies for improvement in supply-chain planning. *International Journal of Forecasting*, 25(1), 3–23. <https://doi.org/10.1016/j.ijforecast.2008.11.010>
- Fry, H. (2018). *Hello World: How to be Human in the Age of the Machine*. Random House.
- Garcia-Retamero, R., Sobkow, A., Petrova, D., Garrido, D., & Traczyk, J. (2019). Numeracy and Risk Literacy: What Have We Learned so Far? *The Spanish Journal of Psychology*, 22, E10. <https://doi.org/10.1017/sjp.2019.16>
- Ghazal, S., Cokely, E.T., & Garcia-Retamero, R. (2014). Predicting biases in very highly educated samples: Numeracy and metacognition. *Judgment and Decision Making*, 9(1), 15–34.
- Green, G.I., & Hughes, C.T. (1986). Effects of Decision Support Systems Training and Cognitive Style on Decision Process Attributes. *Journal of Management Information Systems*, 3(2), 83–93. <https://doi.org/10.1080/07421222.1986.11517764>
- Grove, W.M., Zald, D.H., Lebow, B.S., Snitz, B.E., & Nelson, C. (2000). Clinical versus mechanical prediction: A meta-analysis. *Psychological Assessment*, 12(1), 19–30. <https://doi.org/10.1037/1040-3590.12.1.19>
- Highhouse, S. (2008). Stubborn Reliance on Intuition and Subjectivity in Employee Selection. *Industrial and Organizational Psychology*, 1(3), 333–342. <https://doi.org/10.1111/j.1754-9434.2008.00058.x>
- Inthorn, J., Tabacchi, M. E., & Seising, R. (2015). Having the Final Say: Machine Support of Ethical Decisions of Doctors. In S.P. van Rysewyk & M. Pontier (Eds.), *Machine Medical Ethics* (pp. 181–206). Springer International Publishing. [https://doi.org/10.1007/978-3-319-08108-3\\_12](https://doi.org/10.1007/978-3-319-08108-3_12)
- Jasper, J.D., Bhattacharya, C., & Corser, R. (2017). Numeracy Predicts More Effortful and Elaborative Search Strategies in a Complex Risky Choice Context: A Process-Tracing Approach. *Journal of Behavioral Decision Making*, 30(2), 224–235. <https://doi.org/10.1002/bdm.1934>
- Jasper, J.D., Bhattacharya, C., Levin, I.P., Jones, L., & Bossard, E. (2013). Numeracy as a Predictor of Adaptive Risky Decision Making. *Journal of Behavioral Decision Making*, 26(2), 164–173. <https://doi.org/10.1002/bdm.1748>
- Logg, J.M., Minson, J.A., & Moore, D.A. (2019). Algorithm appreciation : People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes*, 151, 90–103. <https://doi.org/10.1016/j.obhdp.2018.12.005>

- Longoni, C., Bonezzi, A., & Morewedge, C.K. (2019). Resistance to Medical Artificial Intelligence. *Journal of Consumer Research*, 46(4), 629–650. <https://doi.org/10.1093/jcr/ucz013>
- Meehl, P. (1954). *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence*. University of Minnesota Press. <https://doi.org/10.1037/11281-000>
- Millroth, P., & Juslin, P. (2015). Prospect evaluation as a function of numeracy and probability denominator. *Cognition*, 138, 1–9. <https://doi.org/10.1016/j.cognition.2015.01.014>
- Nalborczyk, L., Batailler, C., Løevenbruck, H., Vilain, A., & Bürkner, P.-C. (2019). An Introduction to Bayesian Multilevel Models Using brms: A Case Study of Gender Effects on Vowel Variability in Standard Indonesian. *Journal of Speech, Language, and Hearing Research*, 62(5), 1225–1242. [https://doi.org/10.1044/2018\\_JSLHR-S-18-0006](https://doi.org/10.1044/2018_JSLHR-S-18-0006)
- Önkal, D., Goodwin, P., Thomson, M., Gönül, S., & Pollock, A. (2009). The relative influence of advice from human experts and statistical methods on forecast adjustments. *Journal of Behavioral Decision Making*, 22(4), 390–409. <https://doi.org/10.1002/bdm.637>
- Peters, E. (2012). Beyond Comprehension: The Role of Numeracy in Judgments and Decisions. *Current Directions in Psychological Science*, 21(1), 31–35. <https://doi.org/10.1177/0963721411429960>
- Peters, E., & Bjalkbebring, P. (2015). Multiple numeric competencies: When a number is not just a number. *Journal of Personality and Social Psychology*, 108(5), 802–822. <https://doi.org/10.1037/pspp0000019>
- Peters, E., Kate, M., Simon, A., Tompkins, M.K., Knoll, M.A.Z., Ardoin, S.P., Shoots-Reinhard, B., & Meara, A.S. (2019). Despite high objective numeracy, lower numeric confidence relates to worse financial and medical outcomes. *Proceedings of the National Academy of Sciences*, 116(39), 19386–19391. <https://doi.org/10.1073/pnas.1903126116>
- Peters, E., Västfjäll, D., Slovic, P., Mertz, C.K., Mazzocco, K., & Dickert, S. (2006). Numeracy and Decision Making. *Psychological Science*, 17(5), 407–413. <https://doi.org/10.1111/j.1467-9280.2006.01720.x>
- Petrova, D., Garcia-Retamero, R., Catena, A., Cokely, E.T., Heredia Carrasco, A., Arrebola Moreno, A., & Ramírez Hernández, J.A. (2017). Numeracy Predicts Risk of Pre-Hospital Decision Delay: a Retrospective Study of Acute Coronary Syndrome Survival. *Annals of Behavioral Medicine*, 51(2), 292–306. <https://doi.org/10.1007/s12160-016-9853-1>
- Petrova, D., Garcia-Retamero, R., Catena, A., & van der Pligt, J. (2016). To screen or not to screen: What factors influence complex screening decisions? *Journal of Experimental Psychology: Applied*, 22(2), 247–260. <https://doi.org/10.1037/xap0000086>
- Petrova, D., Traczyk, J., & Garcia-Retamero, R. (2019). What shapes the probability weighting function? Influence of affect, numeric competencies, and information formats. *Journal of Behavioral Decision Making*, 32(2), 124–139. <https://doi.org/10.1002/bdm.2100>
- Petrova, D., van der Pligt, J., & Garcia-Retamero, R. (2014). Feeling the Numbers: On the Interplay Between Risk, Affect, and Numeracy. *Journal of Behavioral Decision Making*, 27(3), 191–199. <https://doi.org/10.1002/bdm.1803>
- Prahl, A., Dexter, F., Braun, M.T., & Van Swol, L. (2013). Review of Experimental Studies in Social Psychology of Small Groups When an Optimal Choice Exists and Application to Operating Room Management Decision-Making. *Anesthesia & Analgesia*, 117(5), 1221–1229. <https://doi.org/10.1213/ANE.0b013e3182a0eed1>

- Schley, D.R., & Peters, E. (2014). Assessing “Economic Value” Symbolic-Number Mappings Predict Risky and Riskless Valuations. *Psychological Science*, 25(3), 753–761. <https://doi.org/10.1177/0956797613515485>
- Shaffer, V.A., Probst, C.A., Merkle, E.C., Arkes, H.R., & Medow, M.A. (2013). Why Do Patients Derogate Physicians Who Use a Computer-Based Diagnostic Support System? *Medical Decision Making*, 33(1), 108–118. <https://doi.org/10.1177/0272989X12453501>
- Shoots-Reinhard, B., Erford, B., Romer, D., Evans, A.T., Shoben, A., Klein, E.G., & Peters, E. (2020). Numeracy and memory for risk probabilities and risk outcomes depicted on cigarette warning labels. *Health Psychology*. <https://doi.org/10.1037/hea0000879>
- Sobków, A., Figol, J., & Traczyk, J. (2020). Zdolności numeryczne jako kluczowe zdolności poznawcze w procesie podejmowania decyzji. *Decyzje*, 33, 25–53. <https://doi.org/10.7206/DEC.1733-0092.139>
- Sobkow, A., Fulawka, K., Tomczak, P., Zjawiony, P., & Traczyk, J. (2019). Does mental number line training work? The effects of cognitive training on real-life mathematics, numeracy, and decision making. *Journal of Experimental Psychology: Applied*, 25(3), 372–385. <https://doi.org/10.1037/xap0000207>
- Sobkow, A., Garrido, D., & Garcia-Retamero, R. (2020). Psychological Perspectives on Financial Decision Making. In T. Zaleskiewicz & J. Traczyk (Eds.), *Psychological Perspectives on Financial Decision Making*. Springer International Publishing. <https://doi.org/10.1007/978-3-030-45500-2>
- Sobkow, A., Olszewska, A., & Traczyk, J. (2020). Multiple numeric competencies predict decision outcomes beyond fluid intelligence and cognitive reflection. *Intelligence*, 80, 101452. <https://doi.org/10.1016/j.intell.2020.101452>
- Sobkow, A., Traczyk, J., Kaufman, S.B., & Nosal, C. (2018). The structure of intuitive abilities and their relationships with intelligence and Openness to Experience. *Intelligence*, 67, 1–10. <https://doi.org/10.1016/j.intell.2017.12.001>
- Traczyk, J., & Fulawka, K. (2016). Numeracy moderates the influence of task-irrelevant affect on probability weighting. *Cognition*, 151, 37–41. <https://doi.org/10.1016/j.cognition.2016.03.002>
- Traczyk, J., Fulawka, K., Lenda, D., & Zaleskiewicz, T. (2021). Consistency in probability processing as a function of affective context and numeracy. *Journal of Behavioral Decision Making*, 34(2), 228–246. <https://doi.org/10.1002/bdm.2206>
- Traczyk, J., Lenda, D., Serek, J., Fulawka, K., Tomczak, P., Strizyk, K., Polec, A., Zjawiony, P., & Sobkow, A. (2018). Does Fear Increase Search Effort in More Numerate People? An Experimental Study Investigating Information Acquisition in a Decision From Experience Task. *Frontiers in Psychology*, 9(1203). <https://doi.org/10.3389/fpsyg.2018.01203>
- Traczyk, J., Sobkow, A., Fulawka, K., Kus, J., Petrova, D., & Garcia-Retamero, R. (2018). Numerate decision makers don't use more effortful strategies unless it pays: A process tracing investigation of skilled and adaptive strategy selection in risky decision making. *Judgment and Decision Making*, 13(4), 372–381. <http://www.sjdm.org/journal/17/17208/jdm17208.pdf>
- Traczyk, J., Sobkow, A., Matukiewicz, A., Petrova, D., & Garcia-Retamero, R. (2020). The experience-based format of probability improves probability estimates: The moderating role of individual differences in numeracy. *International Journal of Psychology*, 55(2), 273–281. <https://doi.org/10.1002/ijop.12566>

Weller, J.A., Dieckmann, N.F., Tusler, M., Mertz, C.K., Burns, W.J., & Peters, E. (2013). Development and Testing of an Abbreviated Numeracy Scale: A Rasch Analysis Approach. *Journal of Behavioral Decision Making*, 26(2), 198–212. <https://doi.org/10.1002/bdm.1751>

APPENDIX

Algorithm aversion: Sensitivity to interventions and the relationship with numeracy

**Table S1**  
*Bayesian Pearson Correlations among predictors in the statistical model*

Variable		1	2	3	4	5	6
1. Score	Pearson's r	—					
	BF <sub>10</sub>	—					
2. Gender	Pearson's r	0.044	—				
	BF <sub>10</sub>	1.536	—				
3. Games	Pearson's r	-0.237 ***	0.219 ***	—			
	BF <sub>10</sub>	> 100	> 100	—			
4. Reading	Pearson's r	0.362 ***	-0.109 ***	-0.091 ***	—		
	BF <sub>10</sub>	> 100	> 100	> 100	—		
5. Class repeat	Pearson's r	-0.148 ***	0.052	0.050	-0.060 **	—	
	BF <sub>10</sub>	> 100	8.880	5.713	69.060	—	
6. Status	Pearson's r	0.407 ***	0.003	-0.094 ***	0.229 ***	-0.099 ***	—
	BF <sub>10</sub>	> 100	0.019	> 100	> 100	> 100	—

\* BF<sub>10</sub> > 10, \*\* BF<sub>10</sub> > 30, \*\*\* BF<sub>10</sub> > 100

**Table S2**  
*Posterior mean, standard error, 95% credible interval, and Rhat statistic for each parameter of the model (with varying intercept by subject) predicting decision to use algorithm in Study 1*

Parameter	Mean	SE	CrI lower	CrI higher	Rhat
Intercept	-0.9	0.19	-1.28	-0.54	1
BNT	-0.29	0.15	-0.59	-0.01	1
Gender	-0.02	0.12	-0.25	0.21	1
Class repeat	0.02	10.03	-19.78	19.7	1
Reading	-0.12	0.04	-0.19	-0.05	1
Games	0	0.04	-0.08	0.07	1
Status	-0.13	0.06	-0.24	-0.02	1
σ <sub>subject</sub>	1.73	0.16	1.44	2.08	1

**Table S3**

*Posterior mean, standard error, 95% credible interval, and Rhat statistic for each parameter of the model (with varying intercept by subject) predicting performance in the estimation task in Study 1*

Parameter	Mean	SE	CrI lower	CrI higher	Rhat
Intercept	23.24	0.34	22.58	23.89	1
BNT	0.01	0.23	-0.44	0.45	1
Decision (to use the statistical model predictions)	-6.34	0.58	-7.47	-5.21	1
Gender	-4.37	0.67	-5.69	-3.07	1
Class repeat	0.02	9.96	-19.64	19.48	1
Reading	1.13	0.21	0.71	1.54	1
Games	-2.05	0.21	-2.46	-1.63	1
Status	-0.55	0.32	-1.19	0.08	1
$\sigma_{\text{subject}}$	0.45	0.33	0.02	1.2	1
$\sigma_e$	14.76	0.19	14.39	15.14	1

**Table S4**

*Posterior mean, standard error, 95% credible interval, and Rhat statistic for each parameter of the model (with varying intercept by subject) predicting decision to use algorithm in Study 2*

Parameter	Mean	SE	CrI lower	CrI higher	Rhat
Intercept	-1.56	0.34	-2.24	-0.89	1
BNT	-0.35	0.18	-0.7	0.01	1
SNS	0.04	0.03	-0.03	0.11	1
Description condition	0.28	0.48	-0.67	1.23	1
Feedback condition	-0.02	0.47	-0.96	0.93	1
Gender	0.08	0.09	-0.1	0.27	1
Class repeat	0.02	9.9	-19.32	19.34	1
Reading	-0.11	0.03	-0.17	-0.06	1
Games	-0.04	0.03	-0.1	0.02	1
Status	-0.11	0.05	-0.2	-0.02	1
$\sigma_{\text{subject}}$	2.46	0.18	2.12	2.84	1

**Table S5**

*Posterior mean, standard error, 95% credible interval, and Rhat statistic for each parameter of the model (with varying intercept by subject) predicting performance in the estimation task in Study 2*

Parameter	Mean	SE	CrI lower	CrI higher	Rhat
Intercept	23.27	0.4	22.5	24.06	1
BNT	-0.12	0.2	-0.52	0.27	1
Decision (to use the statistical model predictions)	-6.28	0.46	-7.18	-5.38	1
SNS	0.05	0.04	-0.03	0.12	1
Description condition	-1.14	0.54	-2.19	-0.09	1
Feedback condition	0.05	0.52	-0.95	1.06	1
Gender	-4.7	0.51	-5.7	-3.7	1
Class repeat	0	9.99	-19.7	19.55	1
Reading	1.14	0.16	0.82	1.46	1

<b>Parameter</b>	<b>Mean</b>	<b>SE</b>	<b>CrI lower</b>	<b>CrI higher</b>	<b>Rhat</b>
Games	-2.04	0.17	-2.36	-1.71	1
Status	-0.42	0.25	-0.91	0.07	1
$\sigma_{\text{subject}}$	0.56	0.37	0.03	1.35	1
$\sigma_e$	15.31	0.15	15.03	15.61	1