Watch Me Improve—Algorithm Aversion and Demonstrating the Ability to Learn

Benedikt Berger, Martin Adam, Alexander Rühr, Alexander Benlian

Business & Information Systems Engineering (2020)

Appendix (available online via http://link.springer.com)

Literature Overview

Table A1: Synopsis of selected studies on algorithm aversion

Article	**Tasks**	**Decision**	**Disclosed information**	**Main Findings**
Dietvorst et al. (2015)	Predict student performance or rank of U.S. states in terms of airline passengers (objective tasks)	Rely on an algorithm or make an estimate by yourself/rely on another person's estimate for an incentivized forecast (nominal choice)	Before the decision, the participants either gather no experience in the task, gather experience by themselves, observe the algorithm, or both gather experience by themselves and observe the algorithm in ten training predictions.	The participants becoming familiar with the algorithm are less likely to rely on it for the incentivized forecast.
Prahl and Van Swol (2017)	Predict performance indicators in operating room management (objective tasks)	Take into account advice after an initial estimate (weight of advice)	The advice comes either from a human or from an algorithm; the participants gather experience in two training predictions before making 14 incentivized predictions.	After receiving an advice that turns out to be worse than the participants' estimate, utilization of advice from the IT system decreases significantly more than utilization of advice from the human.
Dietvorst et al. (2018)	Predict student rank in a math test (objective tasks)	Rely on an algorithm or make an estimate by yourself for an incentivized forecast (nominal choice)	The participants know about the algorithm's average deviation and are allowed to alter the algorithm forecasts to varying extents.	The participants are more likely to rely on an imperfect algorithm if they can adjust its forecasts.

Castelo et al. (2019)	Tasks from widely varying domains (objective and subjective tasks)	Indications of trust in or reliance on algorithmic conduct of the tasks (continuous measures)	The participants receive varying information about the tasks and the performance of algorithms depending on the study.	The participants trust in and rely on algorithms more for tasks that appear objective than for tasks that appear subjective in nature.
Logg et al. (2019)	Various estimation tasks (objective and subjective tasks)	Take into account advice after an initial estimate (weight of advice)	The advice comes either from a human or from an algorithm.	The participants adhere more to advice from an algorithm than to advice from another human. Overconfidence and expertise in the task decrease reliance on algorithms.
Longoni et al. (2019)	Medical analysis and treatment (objective tasks)	Deciding between two advice/service providers or indicating the likelihood of following/utilizing advice/service (nominal choice or continuous measures)	The participants are offered human and/or automated conduct of the healthcare service. Depending on the study, the automated conduct is stated to be of equal or better performance than that of the human conduct.	People prefer human medical advice or service to automated medical advice or service because people believe that algorithms cannot account for their unique characteristics.
Yeomans et al. (2019)	Predict which jokes people find funny (subjective task)	Take into account an algorithm’s advice when recommending a joke or rate a recommended joke (continuous measures)	The algorithm’s performance and processing are revealed to differing degrees.	The participants are reluctant to rely on algorithmic advice when recommending jokes and prefer to receive joke recommendations from humans.

Calculation of the Number of Calls

Overview

We calculated the number of calls underlying the estimation by multiplying an average number of calls per day with the values of seven factors determining deviations from the average. We assumed the average number of calls per day to be 5,000. Table A2 lists the influencing factors. The participants in the experiment knew about the first six factors, while the random influence was unknown to them to ensure it was impossible to somehow determine the number of calls to be estimated.

Table A2: Factors influencing the number of calls on a specific day

Factor	**Levels**
Quarter of the year	Q1 to Q4
Day of the month	1 to 31
Weekday	Monday to Friday
Promotions	Yes or No
Recent sales	-10% below average to 10% above average
Website traffic	-10% below average to 10% above average
Random influence	-0.5% or +0.5%

Quarter of the Year

To incorporate seasonal trends, we included a seasonal factor using the revenues of a large German telecommunications provider as a proxy (see Table A3). The participants in the experiments saw a curve displaying the development of that factor before the estimations but the precise factor values remained unknown.

Table A3: Values of the factor quarter of the year

Quarter of the year	**Factor values**
Q1	0.951
Q2	0.964
Q3	0.997
Q4	1.087

Day of the Month

To include trends occurring over the course of a month, we included a respective factor (see Table A4) in accordance with the call center frequency analysis by Nielsen (2010). The participants in the experiments saw a curve displaying the development of that factor before the estimations but the precise factor values remained unknown.

Table A4: Values of the factor day of the month

Day of the month	Factor values	Day of the month	Factor values	Day of the month	Factor values
1	1.093	11	0.941	21	0.956
2	1.161	12	0.941	22	0.937
3	0.927	13	1.093	23	1.015
4	1.060	14	0.966	24	0.878
5	1.060	15	0.888	25	0.976
6	1.142	16	1.007	26	0.976
7	0.878	17	0.878	27	1.191
8	0.869	18	0.967	28	0.937
9	0.995	19	0.967	29	1.005
10	0.820	20	1.093	30	1.171
				31	1.210

Weekday

To include daily trends during the week, we included a weekday factor (Table A5) in accordance with the call center frequency analysis of Nielsen (2010). The participants in the experiments saw a curve displaying the development of that factor before the estimations but the precise factor values remained unknown.

Table A5: Values of the factor weekday

Weekday	Factor values
Monday	1.154
Tuesday	0.971
Wednesday	0.934
Thursday	1.047
Friday	0.894

Promotions, Current Sales, and Website Visits

To account for current business developments potentially influencing the number of calls, we incorporated the occurrence of promotions, the recent sales development, and recent developments in the website visits in the derivation of the call capacity. We randomized the recent sales developments and website visits drawing from the options displayed in Table A6.

The likelihood of average sales/website visits (~43%) was larger than the likelihood of deviations from the average (~14% each). To incorporate another layer of complexity, the effect of sales/website visit developments on the call volume was disproportionate. Recent sales developments affected the number of calls with a 10% discount. Recent website visits affected the number of calls with a 10% surplus. The occurrence of a promotion increased the number of calls by 20%. The participants in the experiment knew neither the exact effect of promotions on the call volume nor the discount factors for sales and website visit development.

Table A6: Values of the factors promotions, recent sales, and website traffic

Factor	**Factor values**					**Weight**
Promotions	no	yes				120%
Recent sales	-10%	-5%	average	+5%	+10%	90%
Website traffic	-10%	-5%	average	+5%	+10%	110%

Random influence

To further hinder the exact predictability of the number of calls, we included a random influence between -0.5% and +0.5% on the number of calls per day.

Constructs and Items

Table A7: Comprehension questions

Question	Possible answers
What do you have to estimate in this study?	(a) The sales figures of important companies in 2019. (b) The scores of upcoming football matches. (c) The number of incoming calls in a call center.
How much information do you receive as basis for your estimations?	(a) 2 variables (b) 6 variables (c) 15 variables
What is the source of the advice you receive in this study?	(a) Industry Expert (b) Customer Survey (c) Prediction Software

Table A8: Measurement scales for controls

Construct	Items
Trusting disposition (Gefen and Straub 2004) 7-point Likert-type scale Cronbach's $\alpha = 0.96$	(1) I generally trust others. (2) I generally have faith in others. (3) I feel that others are generally well meaning. (4) I feel that others are generally trustworthy.
Personal innovativeness (Agarwal and Prasad 1998) 7-point Likert-type scale Cronbach's $\alpha = 0.79$	(1) If I heard about a new information technology, I would look for ways to experiment with it. (2) Among my peers, I am usually the first to try out new information technologies. (3) In general, I am hesitant to try out new information technologies. (reversed) (4) I like to experiment with new information technologies.
Experience in working for call centers (self-developed)	(1) Are you currently working for a call center or did you do so in the past?
Experience in calling hotlines (self-developed)	(1) How often do you call hotlines?
Product knowledge—call center (Flynn and Goldsmith 1999) 7-point Likert-type scale Cronbach's $\alpha = 0.85$	(1) I know quite a lot about working in call centers. (2) I do not feel very knowledgeable about call centers. (reversed) (3) When it comes to call centers, I really do not know a lot. (reversed)

Table A9: Measurements for demographics

Construct	**Question (and possible answers)**
Age	What is your age in years?
Gender	What is your gender? (male; female)
Education	What is the highest degree or level of school you have completed? (no degree; school to a certain extent; high school; associate degree; Bachelor’s degree; Master’s degree; Professional degree; Doctorate degree)

Table A10: Attention check

Question	**Possible answers**
Getting meaningful and useful responses from participants in a study depends on a number of important factors. Thus, we are interested in knowing certain things about you. Specifically, we are interested in seeing whether you take the time to read survey directions and questions carefully prior to providing an answer. So, in order to demonstrate that you have read these instructions carefully, please ignore the question below and click the next button without providing an answer. Thank you for your cooperation and participation in this study. What is your favorite sport?	(a) Football (b) Soccer (c) Tennis (d) Rugby (e) I don’t play sports (f) (Nothing)

Table A11: Manipulation checks and scenario realism

Construct	Items
Perceived learning (Alavi et al. 2002) 7-point Likert-type scale Cronbach's α = 0.91	Throughout the 8 estimations in the training phase... (1) ... the Prediction Software/Industry Expert gained a good understanding of how to properly estimate the number of calls. (2) ... the Prediction Software/Industry Expert learned to properly estimate the number of calls. (3) ... the Prediction Software/Industry Expert developed the ability to properly estimate the number of calls. (4) ... the Prediction Software's/Industry Expert's ability to properly estimate the number of calls has improved.
Familiarity (Gefen 2000; Kim et al. 2009) 7-point Likert-type scale Cronbach's α = 0.87	(1) I am familiar with the Prediction Software/Industry Expert providing information. (2) I am familiar with the process of the Prediction Software/Industry Expert providing estimations. (3) I am familiar with receiving estimations from the Prediction Software/Industry Expert. (4) Overall, I am familiar with the Prediction Software/Industry Expert.
Anthropomorphism (Bartneck et al. 2009; Benlian et al. 2020) 5-point polarity profile Cronbach's α = 0.92	Please rate the characteristics of the source of the advice (i.e. the one helping you with the estimations): (1) Automated … Human (2) Machinelike … Humanlike (3) Fake … Natural (4) Artificial … Lifelike
Perceived realism (self-developed) 7-point Likert-type scale	(1) The simulation was realistic.

References

Agarwal R, Prasad J (1998) A conceptual and operational definition of personal innovativeness in the domain of information technology. Inf Syst Res 9(2):204-215. doi:10.1287/isre.9.2.204

Alavi M, Marakas GM, Yoo Y (2002) A comparative study of distributed learning environments on learning outcomes. Inf Syst Res 13(4):404-415. doi:10.1287/isre.13.4.404.72

Bartneck C, Kulić D, Croft E, Zoghbi S (2009) Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. Int J Soc Robotics 1(1):71-81. doi:10.1007/s12369-008-0001-3

Benlian A, Klumpe J, Hinz O (2020) Mitigating the intrusive effects of smart home assistants by using anthropomorphic design features: A multimethod investigation. Inf Syst J 30(6):1010-1042. doi:10.1111/isj.12243

Castelo N, Bos MW, Lehmann DR (2019) Task-dependent algorithm aversion. J Market Res 56(5):809-825. doi:10.1177/0022243719851788

Dietvorst BJ, Simmons JP, Massey C (2015) Algorithm aversion: People erroneously avoid algorithms after seeing them err. J Exp Psychol Gen 144(1):114-126. doi:10.1037/xge0000033

Dietvorst BJ, Simmons JP, Massey C (2018) Overcoming algorithm aversion: People will use imperfect algorithms if they can (even slightly) modify them. Manag Sci 64(3):1155-1170. doi:10.1287/mnsc.2016.2643

Flynn LR, Goldsmith RE (1999) A short, reliable measure of subjective knowledge. J Bus Res 46(1):57-66. doi:10.1016/S0148-2963(98)00057-5

Gefen D (2000) E-commerce: the role of familiarity and trust. Omega 28(6):725-737. doi:10.1016/S0305-0483(00)00021-9

Gefen D, Straub DW (2004) Consumer trust in B2C e-Commerce and the importance of social presence: Experiments in e-Products and e-Services. Omega 32(6):407-424. doi:10.1016/j.omega.2004.01.006

Kim DJ, Ferrin DL, Rao HR (2009) Trust and satisfaction, two stepping stones for successful e-commerce relationships: A longitudinal exploration. Inf Syst Res 20(2):237-257. doi:10.1287/isre.1080.0188

Logg JM, Minson JA, Moore DA (2019) Algorithm appreciation: People prefer algorithmic to human judgment. Organ Behav Hum Decis Process151:90-103. doi:10.1016/j.obhdp.2018.12.005

Longoni C, Bonezzi A, Morewedge CK (2019) Resistance to medical artificial intelligence. J Consumer Res 46(4):629-650. doi:10.1093/jcr/ucz013

Nielsen TB (2010) Call center capacity planning. Technical University of Denmark, Kongens Lyngby

Prahl A, Van Swol L (2017) Understanding algorithm aversion: When is advice from automation discounted? J Forecast 36(6):691-702. doi:10.1002/for.2464

Yeomans M, Shah A, Mullainathan S, Kleinberg J (2019) Making sense of recommendations. J Behav Decis Mak 32(4):403-414. doi:10.1002/bdm.2118