“There Is Not Enough Information”: On the Effects of Explanations on Perceptions of Informational Fairness and Trustworthiness in Automated Decision-Making

Appendix

JAKOB SCHOEFFER, Karlsruhe Institute of Technology, Germany
NIKLAS KUEHL, Karlsruhe Institute of Technology, Germany
YVETTE MACHOWSKI, Karlsruhe Institute of Technology, Germany

A ABBREVIATIONS

Tab. 3 contains our most commonly used abbreviation.

Table 3. Summary of commonly used abbreviations.

Abbreviation	Explanation
ADS	Automated decision system(s)
AILIT	AI literacy
AMTIN	Amount of information
(Base)	Baseline treatment without explanations
(F)	Treatment with disclosure of factors
(FFI)	Treatment with disclosure of factors and factor importance
(FFICF)	Treatment with disclosure of factors, factor importance, and counterfactual explanations
INFF	Informational fairness (dependent variable)
SEM	Structural equation model
SP	Study participant(s)
TRST	Trustworthiness (dependent variable)
XAI	Explainable AI

B CONSTRUCTS AND MEASUREMENT ITEMS

All items within the following constructs were measured on a 5-point Likert scale and mostly drawn (and adapted) from previous studies.

(1) **Informational Fairness (INFF)**
- The automated decision system explains decision-making procedures thoroughly. [6]
- The automated decision system’s explanations regarding procedures are reasonable. [6]
- The automated decision system tailors communications to meet the applying individual’s needs. [6]
- I understand the process by which the decision was made. [2]
- I received sufficient information to judge whether the decision-making procedures are fair or unfair.

(2) **Trustworthiness (TRST)**

Authors’ addresses: Jakob Schoeffer, Karlsruhe Institute of Technology, Germany, jakob.schoeffer@kit.edu; Niklas Kuehl, Karlsruhe Institute of Technology, Germany, niklas.kuehl@kit.edu; Yvette Machowski, Karlsruhe Institute of Technology, Germany, yvette.machowski@alumni.kit.edu.

- Given the provided explanations, I trust that the automated decision system makes good-quality decisions. [11]
- Based on my understanding of the decision-making procedures, I know the automated decision system is not opportunistic. [5]
- Based on my understanding of the decision-making procedures, I know the automated decision system is trustworthy. [5]
- I think I can trust the automated decision system. [3]
- The automated decision system can be trusted to carry out the loan application decision faithfully. [3]
- In my opinion, the automated decision system is trustworthy. [3]

(3) **(Self-Assessed) AI Literacy (AILIT)**

- How would you describe your knowledge in the field of artificial intelligence?
- Does your current employment include working with artificial intelligence?
- I am confident interacting with artificial intelligence. [21]
- I understand what the term *artificial intelligence* means.

C EXPLANATION STYLES FOR ONE EXEMPLARY SETTING

Condition *(F)*

A finance company offers loans on real estate in urban, semi-urban and rural areas. A potential customer first applies online for a specific loan, and afterwards the company assesses the customer's eligibility for that loan.
An individual applied online for a loan at this company. The company denied the loan application. The decision to deny the loan was made by an automated decision system and communicated to the applying individual electronically and in a timely fashion.

The automated decision system explains that the following factors (in alphabetical order) on the individual were taken into account when making the loan application decision:

- Applicant Income: $3,069 per month
- Co-Applicant Income: $0 per month
- Credit History: Good
- Dependents: 0
- Education: Graduate
- Gender: Male
- Loan Amount: $71,000
- Loan Amount Term: 480 months
- Married: No
- Property Area: Urban
- Self-Employed: No

Condition *(FFI)*

A finance company offers loans on real estate in urban, semi-urban and rural areas. A potential customer first applies online for a specific loan, and afterwards the company assesses the customer's eligibility for that loan.
An individual applied online for a loan at this company. The company denied the loan application. The decision to deny the loan was made by an automated decision system and communicated to the applying individual electronically and in a timely fashion.

The automated decision system explains …

- …that the following factors (in alphabetical order) on the individual were taken into account when making the loan application decision:
  - Applicant Income: $3,069 per month
  - Co-Applicant Income: $0 per month
  - Credit History: Good
  - Dependents: 0
  - Education: Graduate
  - Gender: Male
  - Loan Amount: $71,000
  - Loan Amount Term: 480 months
  - Married: No
  - Property Area: Urban
  - Self-Employed: No
- …that different factors are of different importance in the decision. The following list shows the order of factor importance, from most important to least important: Credit History > Loan Amount > Applicant Income > Co-Applicant Income > Property Area > Married > Dependents > Education > Loan Amount Term > Self-Employed > Gender

Condition *(FFICF)*

A finance company offers loans on real estate in urban, semi-urban and rural areas. A potential customer first applies online for a specific loan, and afterwards the company assesses the customer's eligibility for that loan.
An individual applied online for a loan at this company. The company denied the loan application. The decision to deny the loan was made by an automated decision system and communicated to the applying individual electronically and in a timely fashion.

The automated decision system explains …

- …that the following factors (in alphabetical order) on the individual were taken into account when making the loan application decision:
  - Applicant Income: $3,069 per month
  - Co-Applicant Income: $0 per month
  - Credit History: Good
  - Dependents: 0
  - Education: Graduate
  - Gender: Male
  - Loan Amount: $71,000
  - Loan Amount Term: 480 months
  - Married: No
  - Property Area: Urban
  - Self-Employed: No
- …that different factors are of different importance in the decision. The following list shows the order of factor importance, from most important to least important: Credit History > Loan Amount > Applicant Income > Co-Applicant Income > Property Area > Married > Dependents > Education > Loan Amount Term > Self-Employed > Gender
- …that the individual would have been granted the loan if—everything else unchanged—one of the following hypothetical scenarios had been true:
  - The Co-Applicant Income had been at least $800 per month
  - The Loan Amount Term had been 408 months or less
  - The Property Area had been Rural

D MEASUREMENT MODEL

In order to assess the validity and the reliability of our constructs, we conduct a confirmatory factor analysis and assess the results w.r.t. multiple measures. As measures for convergent reliability, we examine average variance extracted (AVE) and composite reliability (CR). For the constructs of informational fairness and trustworthiness, AVE is above the recommended threshold of 0.5, whereas the AVE of AI literacy is 0.41. According to Fornell and Larcker [8], if AVE is low, convergent validity of a construct can still be sufficient if composite reliability (CR) is above 0.6, which is the case for all three constructs, including AI literacy (see Tab. 4). In fact, the CR of our three main constructs, informational fairness (0.88), trustworthiness (0.94), and AI literacy (0.72) is above the recommended threshold of 0.7 [1], indicating that our convergent validity is adequate for AI literacy as well, despite the lower AVE measure.

Cronbach's alpha (CA) values for our constructs are larger than the recommended threshold of 0.7, thus showing good reliability for all constructs [7]. Validity and reliability measures are summarized in Tab. 4. Our matrix of factor loadings, demonstrated in Tab. 5, shows that all items load highly (>0.5) on one factor each with low cross-loadings, and the correlations between factors are all below 0.7 (see Tab. 4). Furthermore, the AVE value of each of our constructs is larger than the squared correlation of that construct with every other construct, which is a discriminant validity measure suggested by Chin [4] and Fornell and Larcker [8]. Therefore, convergent validity and discriminant validity are sufficiently satisfied. We test for multicollinearity by determining the variance inflation factors (VIF). According to a rule of thumb, the VIF has to be lower than 10, otherwise, multicollinearity might be a serious problem [20]. All VIFs in our model are less than 2, which indicates that there are no issues of multicollinearity.

Table 4. Correlations and measurement information for latent factors.

Factor	M	SD	CA	CR	AVE	INFF	TRST	AILIT
INFF	3.15	0.87	0.87	0.88	0.60	1.00		
TRST	3.26	0.84	0.94	0.94	0.73	0.67	1.00	
AILIT	2.87	0.61	0.71	0.72	0.41	0.25	0.18	1.00

Notes: M = Mean; SD = Standard deviation

Table 5. Standardized loadings of measurement items on constructs.

Measurement item	INFF	TRST	AILIT
INFF1	**0.95**	-0.11	-0.03
INFF2	**0.65**	0.21	0.01
INFF3	**0.52**	0.10	0.05
INFF4	**0.79**	0.01	0.03
INFF5	**0.76**	0.01	0.00
TRST1	0.24	**0.66**	-0.05
TRST2	0.20	**0.51**	-0.08
TRST3	0.01	**0.90**	-0.01
TRST4	-0.08	**0.97**	0.06
TRST5	0.02	**0.90**	0.05
TRST6	-0.09	**1.01**	0.00
AILIT1	0.08	-0.11	**0.73**
AILIT2	0.06	-0.03	**0.53**
AILIT3	-0.12	0.17	**0.67**
AILIT4	0.00	-0.02	**0.58**

E SEM MODEL: RESULTS OF MODEL ESTIMATION

Detailed information on the results of the SEM model estimation, including path estimates, standard errors (SE), z-values, p-values, and standardized estimates (Std.lv) are reported in Tab. 6. A breakdown of direct and indirect effects of independent variables on trustworthiness (TRST) is given in Tab. 7.

F SOFTWARE AND TOOLS

Tab. 8 contains all employed software and tools.

Table 6. Results of model estimation.

Path	Estimate	SE	z-value	p-value	Std.lv
AILIT → INFF	0.59***	0.08	7.01	<0.001	0.31
AMTIN → INFF	0.37***	0.03	14.25	<0.001	0.47
INFF → TRST	0.78***	0.05	15.30	<0.001	0.78
AILIT → TRST	-0.02	0.07	-0.24	0.81	-0.01
AMTIN → TRST	-0.09*	0.04	-2.55	0.01	-0.11
Notes: *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$					

Table 7. Decomposition of effects on perceived trustworthiness.

	Direct effect	Indirect effect	Total effect
AMTIN on TRST	-0.09*	0.37·0.78=0.29***	0.20***
AILIT on TRST	-0.02	0.59·0.78=0.46***	0.44***
Notes: *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$			

Table 8. Software and tools.

Task(s)	Software/tool	Source
Data processing (general)	`Python`	Van Rossum and Drake Jr [19]
ML for training ADS and predictions	`Python` package `scikit-learn`	Pedregosa et al. [14]
Crowdsourcing study participants	`Prolific`	Palan and Schitter [13]
Questionnaires	`SoSci Survey`	Leiner [12]
Survey data processing, statistical analyses	`R`	R Core Team [15]
CFA, model fit, measurement model, SEM	`R` package `lavaan`	Rosseel [18]
Fit measures, reliability measures	`R` package `cSEM`	Rademaker and Schuberth [16]
Cross-loadings table, correlations	`R` package `psych`	Revelle [17]
VIF	`R` package `car`	Fox and Weisberg [9]
Qualitative analysis	`MAXQDA`	Kuckartz and Rädiker [10]

REFERENCES

[1] Donald Barclay, Christopher Higgins, and Ronald Thompson. 1995. *The Partial Least Squares (PLS) Approach to Causal Modeling: Personal Computer Adoption and Use as an Illustration.*

[2] Reuben Binns, Max Van Kleek, Michael Veale, Ulrik Lyngs, Jun Zhao, and Nigel Shadbolt. 2018. 'It's reducing a human being to a percentage' – Perceptions of justice in algorithmic decisions. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–14.

[3] Lemuria Carter and France Bélanger. 2005. The utilization of e-government services: Citizen trust, innovation and acceptance factors. *Information Systems Journal* 15, 1 (2005), 5–25.

[4] Wynne W Chin. 1998. The partial least squares approach to structural equation modeling. *Modern Methods for Business Research* 295, 2 (1998), 295–336.

[5] Chao-Min Chiu, Hua-Yang Lin, Szu-Yuan Sun, and Meng-Hsiang Hsu. 2009. Understanding customers' loyalty intentions towards online shopping: An integration of technology acceptance model and fairness theory. *Behaviour & Information Technology* 28, 4 (2009), 347–360.

[6] Jason A Colquitt and Jessica B Rodell. 2015. Measuring justice and fairness. (2015).

[7] Jose M Cortina. 1993. What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology* 78, 1 (1993), 98–104.

[8] Claes Fornell and David F Larcker. 1981. Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research* 18, 1 (1981), 39–50.

[9] John Fox and Sanford Weisberg. 2019. *An R Companion to Applied Regression* (3 ed.). Sage, Thousand Oaks CA. https://socialsciences.mcmaster.ca/jfox/Books/Companion/

[10] Udo Kuckartz and Stefan Rädiker. 2019. *Analyzing Qualitative Data with MAXQDA*. Springer.

[11] Min Kyung Lee. 2018. Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management. *Big Data & Society* 5, 1 (2018), 1–16.

[12] D J Leiner. 2019. SoSci Survey. *München: SoSci Survey GmbH* (2019).

[13] Stefan Palan and Christian Schitter. 2018. Prolific.ac—A subject pool for online experiments. *Journal of Behavioral and Experimental Finance* 17 (2018), 22–27.

[14] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. 2011. Scikit-learn: Machine learning in Python. *The Journal of Machine Learning Research* 12 (2011), 2825–2830.

[15] R Core Team. 2017. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.r-project.org/

[16] Manuel E Rademaker and Florian Schuberth. 2020. *cSEM: Composite-Based Structural Equation Modeling*. https://m-e-rademaker.github.io/cSEM/

[17] William Revelle. 2020. *psych: Procedures for Psychological, Psychometric, and Personality Research*. Northwestern University, Evanston, Illinois. https://cran.r-project.org/package=psych

[18] Yves Rosseel. 2012. lavaan: An R package for structural equation modeling. *Journal of Statistical Software* 48, 2 (2012), 1–36. http://www.jstatsoft.org/v48/i02/

[19] Guido Van Rossum and Fred L Drake Jr. 1995. *Python Tutorial*. Vol. 620. Centrum voor Wiskunde en Informatica Amsterdam.

[20] Eric Vittinghoff, David V Glidden, Stephen C Shiboski, and Charles E McCulloch. 2011. *Regression Methods in Biostatistics: Linear, Logistic, Survival, and Repeated Measures Models*. Springer Science & Business Media.

[21] Ann Wilkinson, Julia Roberts, and Alison E While. 2010. Construction of an instrument to measure student information and communication technology skills, experience and attitudes to e-learning. *Computers in Human Behavior* 26, 6 (2010), 1369–1376.