Application and Analysis

The questionnaire consists of six subscales, which are listed together with the corresponding items in Table 1. The underlying theoretical model is described in Körber (2019) and visualized in Figure 1. The responses to inverted items (item 5, 7, 10, 15, and 16) must be recoded prior to the analysis such that a higher agreement results in a lower score in the respective dimension. For example, a response of 5 to item 16 must be recoded to 1, an answer of 4 to 2 (and vice versa). The category *no response* must not be encoded as 0 but as "no response" or as a missing value.

Each of these subscales should be analyzed on its own. Accordingly, it is possible to use solely one or more subscales without the complete questionnaire. It is still advised to use the complete questionnaire to assess trust in automation in its entirety. A total sum score derived from all items together is, given the multidimensionality, not unambiguously interpretable with the current state of knowledge.

Table 1
Scales and items of the questionnaire

#	Item	Subscale
		Reliability/Competence
1	R/C1	The system is capable of interpreting situations correctly
6	R/C2	The system works reliably
10	R/C3*	A system malfunction is likely
13	R/C4	The system is capable of taking over complicated tasks
15	R/C5*	The system might make sporadic errors
19	R/C6	I am confident about the system's capabilities
		Understanding/Predictability
2	U/P1	The system state was always clear to me
7	U/P2*	The system reacts unpredictably
11	U/P3	I was able to understand why things happened
16	U/P4*	It's difficult to identify what the system will do next
		Familiarity
3	F1	I already know similar systems
17	F2	I have already used similar systems
		Intention of Developers
4	I1	The developers are trustworthy
8	I2	The developers take my well-being seriously
		Propensity to Trust
5	Pro1*	One should be careful with unfamiliar automated systems
12	Pro2	I rather trust a system than I mistrust it
18	Pro3	Automated systems generally work well
		Trust in Automation
9	T1	I trust the system
14	T2	I can rely on the system

Note. *: inverse item.

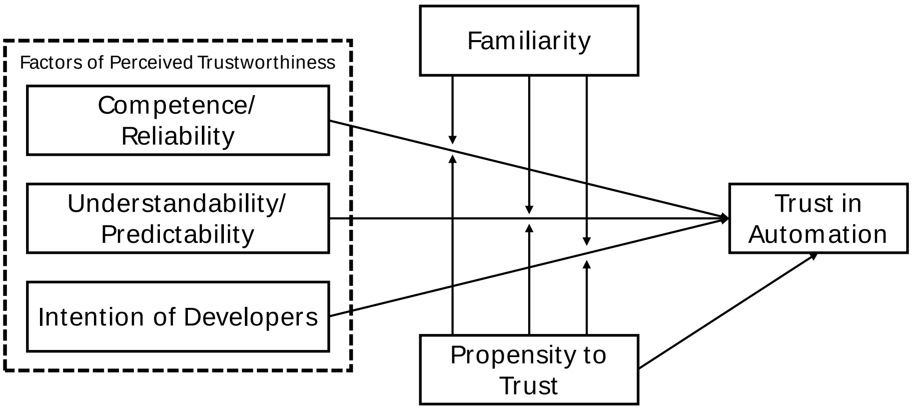

Figure 1. The theoretical model of trust in automation (Körber, 2019).

Psychometric quality

In the following, the results of two studies conducted to assess the psychometric quality and construct validity are presented. A more detailed account, which also includes the development process, can be found in Körber (2018) and Körber (2019). All reported statistics refer to the questionnaire's German version.

In an online study, a total of $n = 58$ participants (aged between 17 and 72 years, $M = 34.00$, $SD = 15.10$, 59 % male, 38 % female) watched two videos of a conditionally automated highway featuring either a perfectly functioning (*reliable* condition) automated driving system (ADS) or the same videos extended by a take-over request (*unreliable* condition). The resulting indices of reliability of the scales ranged between .77 and .95 and are listed in Table 2.

Table 2
Indices of the internal consistency of each scale

	Omega Total	Revelle's Omega
Familiarity	.83[a]	-
Intention of Developers	.79[a]	-
Propensity to Trust	.78	.77
Reliability/Competence	.92	.95
Understanding	.81	.88

Note. [a]: since Omega total and Revelle's omega cannot be calculated for scales with fewer than three items, the Spearman-Brown coefficient according to Eisinga et al. (2013) was calculated.

The resulting pattern matrix of an exploratory factor analysis (principal axis factoring) with oblique rotation (oblimin) shows a clear structure of four factors, where each factor is represented by multiple variables that show high pattern coefficients (>.50) while each of them does not load substantially (>.35) onto other factors, as well as medium to high communalities. More detailed statistics are reported in Körber (2019). Participants in the *reliable* condition rated as more reliable ($t(41.32) = 3.76$, $p < .001$, $d = 1.05$). All scales correlated positively with different strength with the rating of the item "I trust this system" on a 5-point rating scale (1 = strongly disagree, 5 = strongly agree; lowest: *Familiarity*: $r = .33$, highest: *Reliability*: $r = .85$).

Predictive validity has been confirmed in a driving simulator study by Körber et al. (2018). In this study, $n = 40$ participants experienced three critical situations while driving in a conditionally automated vehicle (SAE Level 3). Eye-tracking was used to assess how much the participants rely on the ADS. Furthermore, the instruction for the ADS was varied between two groups with participants receiving either trust-promoting or trust-lowering introductory information. The *Trust in Automation* (TiA) questionnaire was administered three times: 1) after an introductory video, 2) after an introductory drive, 3) after the experimental drive. The analysis comprised the whole TiA as well as just the subscales *Trust in Automation* and *Competence*. The scale *Trust in Automation* exhibited an internal consistency of $\alpha = .63$ after the introductory video, $\alpha = .70$ after the introductory drive, and $\alpha = .85$ after the experimental drive. The subscale *Competence* showed an internal consistency of $\alpha = .71$ (introductory video), $\alpha = .71$ (introductory drive) und $\alpha = .83$ (experimental drive). Participants with higher reported trust consistently showed stronger reliance on the ADS in all behavioral measurements: The reported trust in automation correlated positively with take-over time ($r = .27$ to $r = .33$) and negatively with minimum time-to-collision ($r = -.29$ to $r = -.35$). With increasing trust, participants spent more time looking at the NDRT ($r = .34$ to .35), while at the same time they monitored the road and the instrument cluster shorter ($r = -.33$ to $-.42$) and less frequently ($r = -.40$ to $-.44$). Furthermore, participants who overruled the ADS in close but non-critical situations reported lower trust ($d = 0.23$ to $d = 0.51$).

Use and Further Development

Feel free to use the questionnaire abiding by the license (CC BY-SA 4.0) and to develop the questionnaire further. Körber (2019) and the README of the corresponding Github repository (github.com/moritzkoerber/TiA_Trust_in_Automation_Questionnaire) list some potential objectives for future work. A more detailed discussion on the use of a single item compared to a total questionnaire can be found in Körber (2018) and Körber (2019). The reference to cite the questionnaire is as follows (APA 7th edition):

Körber, M. (2019). Theoretical considerations and development of a questionnaire to measure trust in automation. In S. Bagnara, R. Tartaglia, S. Albolino, T. Alexander, & Y. Fujita (Eds.), *Proceedings of the 20th Congress of the International Ergonomics Association (IEA 2018): Volume VI: Transport Ergonomics and Human Factors (TEHF), Aerospace Human Factors and Ergonomics* (1st ed., pp. 13–30). Springer.

References

Eisinga, R., Grotenhuis, M. t., & Pelzer, B. (2013). The reliability of a two-item scale: Pearson, Cronbach, or Spearman-Brown? *International Journal of Public Health*, *58*(4), 637–642. https://doi.org/10.1007/s00038-012-0416-3

Körber, M. (2018). *Individual differences in human-automation interaction: A driver-centered perspective on the introduction of automated vehicles* [Doctoral dissertation]. Technical University of Munich, Munich. https://mediatum.ub.tum.de/1432904

Körber, M. (2019). Theoretical considerations and development of a questionnaire to measure trust in automation. In S. Bagnara, R. Tartaglia, S. Albolino, T. Alexander, & Y. Fujita (Eds.), *Proceedings of the 20th Congress of the International Ergonomics Association (IEA 2018): Volume VI: Transport Ergonomics and Human Factors (TEHF), Aerospace Human Factors and Ergonomics* (1st ed., pp. 13–30). Springer.

Körber, M., Baseler, E., & Bengler, K. (2018). Introduction matters: Manipulating trust in automation and reliance in automated driving. *Applied Ergonomics*, *66*, 18–31. https://doi.org/10.1016/j.apergo.2017.07.006