ARL-TR-9875 ● FEB 2024

Re-examining the Trust in Autonomy (TIA) Scale as Trust and Distrust Factors

by Julia L Wright, Stephen M Fiore, Joshua Foldes, and Gita Sukthankar

NOTICES

Disclaimers

The research reported herein was sponsored by the US Army Combat Capabilities Development Command (DEVCOM) Army Research Laboratory (ARL) and was accomplished under Cooperative Agreement W911NF-21-2-0279 with the University of Central Florida. The views and conclusions contained in this report are those of the authors and should not be interpreted as representing the official policies or positions, either expressed or implied, of ARL, the US Government, or the University of Central Florida. The US Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein.

ARL-TR-9875 ● FEB 2024

Re-examining the Trust in Autonomy (TIA) Scale as Trust and Distrust Factors

Julia L Wright
DEVCOM Army Research Laboratory

Stephen M Fiore, Joshua Foldes, and Gita Sukthankar
University of Central Florida

REPORT DOCUMENTATION PAGE

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED START DATE	3. DATES COVERED END DATE
February 2024	Technical Report	8/1/2022	7/30/2023

4. TITLE AND SUBTITLE
Re-examining the Trust in Autonomy (TIA) Scale as Trust and Distrust Factors

5a. CONTRACT NUMBER	5b. GRANT NUMBER	5c. PROGRAM ELEMENT NUMBER
W911WF-21-2-0279		
5d. PROJECT NUMBER	**5e. TASK NUMBER**	**5f. WORK UNIT NUMBER**
ARL 23-068		

6. AUTHOR(S)
Julia L Wright, Stephen M Fiore, Joshua Foldes, and Gita Sukthankar

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)	8. PERFORMING ORGANIZATION REPORT NUMBER
DEVCOM Army Research Laboratory ATTN: FCDD-RLA-FD Aberdeen Proving Ground, MD 21005	ARL-TR-9875

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
DISTRIBUTION STATEMENT A. Approved for public release: distribution unlimited.

13. SUPPLEMENTARY NOTES
ORCID ID: Julia L Wright, 0000-0003-3026-1538

14. ABSTRACT
This report describes an exploratory factor analysis (EFA) of the Jian et al. “Trust in Automation” questionnaire and re-analysis of human–machine trust data using a two-factor model of the items (Jian J-Y, Bisantz AM, Drury CG. Foundations for an empirically determined scale of trust in automated systems. Int J Cogn Ergon. 2000;4(1):53–71). The purpose was to examine the degree to which two distinct factors—separately measuring trust and distrust—were identifiable. Data was gathered from six similar experiments that all used the questionnaire. Two datasets were created for analysis: a dataset wherein there is one data point for each participant (Single, N = 91), and a dataset that included repeated submissions by each participant (Full, N = 503). First, an EFA was performed to determine the robustness of the two factors. After verifying the two factors, we reanalyzed an existing dataset that examined trust based on levels of transparency when interacting with an agent. We show how separating out the trust and distrust subcomponents of the Jian et al. questionnaire can provide additional insights into human–machine teaming research.

15. SUBJECT TERMS
Trust in Automation, factor analysis, scale validation, distrust of automation, human–autonomy teaming, Humans in Complex Systems

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	C. THIS PAGE	17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UU	26

19a. NAME OF RESPONSIBLE PERSON	19b. PHONE NUMBER *(Include area code)*
Julia L Wright	407-208-3348

STANDARD FORM 298 (REV. 5/2020)
Prescribed by ANSI Std. Z39.18

Contents

List of Figures

List of Tables

1. Introduction

Understanding how attitudes affect performance in human–machine teams is increasingly important for the modern battlefield. As automation and autonomy increase, the military must understand the attitudinal, behavioral, and cognitive effects on human performance. Trust in machine-aiding continues to be a critical issue, particularly as machine intelligence increases. One under-studied issue associated with trust is the degree to which trust and distrust are separable constructs versus opposite ends of a hypothetical continuum. In the study of human–human teaming, Wildman et al. (2009) argued that trust in collaboration should be distinguished between high and low levels of trust as well as high and low levels of distrust. Specifically, rather than viewing distrust as merely the lowest form of trust, they argued that both trust and distrust can manifest in higher and lower levels. Wildman et al. (2024) showed that, across a number of distinct collaborative contexts, this distinction is both separable and differentially predictive of process and performance outcomes.

In the study of human–machine teams, trust research has rarely examined this distinction. This is surprising given that scales developed to understand trust in technology have included, at least implicitly, assessments of distrust. For example, the Trust in Automation (TIA) scale by Jian et al. (2000), assessed trust and distrust as associated concepts. Later research more closely examined the scale using confirmatory factor analysis (CFA) to determine if a single-factor model of trust or a two-factor model (trust and distrust) was superior (Spain et al. 2008). They found that the two-factor oblique model demonstrated superior fit; thus, verifying the scale assesses two distinct but related factors: trust and distrust.

Twenty years later, automation and autonomy are ubiquitous in today's society, which should affect the public's perception of autonomous systems. As such, the utility of the original TIA scale may have lessened since its inception. In this report, we re-examine the scale using data from several recent studies. First, using exploratory factor analysis (EFA) we verify the two-factor model. Then, we re-examine data from a recent study using the separate factors to understand better how using a two-factor model may enhance our understanding of trust and distrust in automation.

2. Background

When approaching the development of a scale to assess trust in automation, Jian et al. (2000) began with a series of studies identifying words that describe trust and distrust, and assigning them to categories (i.e., "general trust," "trust between people," and "trust between human and automated systems"). They then had participants rate how closely each word was related to "trust" or "distrust" using seven-point Likert scales. Analysis showed that these ratings of trust/distrust were highly negatively correlated in each of the three categories, which Jian et al. interpreted as indicating these terms were opposite, rather than different, factors. Their further analysis showed that the ratings in the human–human trust category were distinct from those in both the general trust and human–machine trust categories, indicating that people perceive trust in regard to other humans differently than they do for the other categories.

The original Jian et al. (2000) TIA scale has 12 questions designed to assess human trust in automation. They concluded that distrust and trust are opposing constructs along a single dimension, as such, the survey from Jian et al. comprises five questions that assess distrust and seven questions that assess trust. Critical for the purpose of this report, Jian et al. state that the five items assessing distrust should be reverse scored. From this, an overall sum of the 12 total items yields an overall score rating one's trust in automation. Thus, rather than viewing trust and distrust as separate, the reverse scoring simply lowers the perceptions of trust captured in the measure (e.g., a high score on a question designed to capture distrust is reversed to create a low score of trust). Further, although the scale was empirically developed, and subsequently used in numerous studies of trust in automation, no assessments of convergent or discriminant validity were run.

In 2008, Spain et al. proposed to evaluate the psychometric properties and validate the scale. To that end, they conducted a repeated measures study (N = 60) wherein students performed a monitoring task with the assistance of an automated aid and completed the Jian et al. TIA scale after each of three trials. The experimenters then reverse-scored the distrust questions and conducted CFAs comparing three models: a single-factor and two two-factor, one orthogonal, and one oblique rotation. All CFAs used the "Maximum Likelihood" method. They concluded the two-factor oblique model best fit the data, suggesting there are two factors, trust and distrust, which are distinct but related.

The Jian et al. (2000) TIA scale is still widely used in research (as evidenced by over 1400 citations on Google Scholar). However, since its inception, automation and autonomy are ubiquitous in many aspects of our daily lives. This raises the question of how to best use the scale when assessing human trust and distrust toward automation. To this end, we re-evaluated the scale using data collected within the past 5 years to determine the value that comes from separately considering trust and distrust as distinct concepts.

3. Method

For our analysis, we used data from six human–agent teaming studies conducted by the US Army Combat Capabilities Development Command (DEVCOM) Army Research Laboratory (ARL) (Drnec et al. 2018; Wright et al. 2021, 2024; Cox et al. 2022; Gremillion et al. 2022, 2024). These studies had similar tasking, in which the participants worked with an autonomous agent to identify targets in a simulated environment and supervised the vehicle navigating through the environment. All studies were repeated measures design, wherein each participant completed the Jian et al. TIA scale after each trial. Each completed TIA scale is considered a "data point." See Table 1 for detailed information about the studies and their data.

Table 1 Participant and data point counts for each study. The final dataset column shows the number of usable data points for each study. The single data column shows the number of data points with only one entry per participant.

Study	No. of participants	No. of times the questionnaires were administered	Total potential data points	Total actual data points in raw data	Removed by cleaning	Final dataset	Single data
Cox et al. 2022	13	17	167	111	2	109	13
Drnec et al. 2018	19	4	76	72	3	69	19
Gremillion et al. 2024	28	22	476	308	95	213	27
Gremillion et al. 2022	5	2	10	6	0	6	3
Wright et al. 2024	20	4	80	80	1	79	20
Wright et al. 2021	9	3	27	27	0	27	9
Totals	**94**	**NA**	**836**	**604**	**101**	**503**	**91**

For our analyses we conducted the following data cleaning steps. First, we removed incomplete data points. Next, we removed entries made by confederates in the Gremillion et al. (2024) study (this was the only study that used confederates to fill in for missing participants). Then, entries with the same answer for all 12 items (a.k.a., straightlining), and entries where at least 3 of the first 5 answers (as answered by the participant) were the same as at least 4 of the last 7 answers (contradictory answers), were removed. Finally, when elapsed time for completing the scale was available, it was reviewed and times that were less than 6 s, or more than 2 SD lower than the mean time, were removed.

Software

Analyses were performed using IBM's SPSS[*] Statistics version 28.

4. Factor Analyses

4.1 Data

All studies were repeated measures; however, the number of trials between studies was not consistent. As such, there was concern that data from a single study could sway the results. Also, there was some concern about potential multicollinearity issues with the complete dataset, or that the larger studies would have a greater influence on the findings. To address these concerns, a dataset was created wherein each participant was represented by one data point (the first instance of completing the questionnaire). Two datasets were examined: a "Full" dataset (N = 503) and a "Single" dataset (N = 91). The first five questions in each dataset were reverse scored before further analysis, as was done in Spain et al. (2008).

We began with the intent to replicate Spain et al.'s 2008 procedure. Although we followed their steps closely, they ran a series of CFAs using the maximum likelihood estimation and specified the number of factors at each iteration—using both orthogonal and oblique rotations—followed by a series of goodness-of-fit tests to determine which model fit best. We ran EFAs using the maximum likelihood estimation for the data extraction method, as the EFA uses correlations between the variables to define the number of factors rather than impose a set number of factors. Similar to Spain et al.'s work, both orthogonal and oblique rotations were examined.

4.2 Sample Correlation Matrix

A bivariate correlation analysis was conducted to examine the dataset for potential multicollinearity issues. As shown in Tables 2 and 3, all items are significantly correlated. Several correlations (i.e., Q9 with Q's 7 and 8; Q10 with Q's 7–9; Q11 with Q's 6–10) are over r = 0.8, indicating there may be issues with multicollinearity between these items. There were similar issues with the single data point set shown in Table 3.

[*] SPSS is a registered trademark of the IBM Corporation.

Table 2 Bivariate correlations of questions 1–12. Full dataset, N = 503.

Correlations N = 503													
Question (Q)		**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
1	Pearson Correlation	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
	Sig. (two-tailed)	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
2	Pearson Correlation	.758[a]	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
	Sig. (two-tailed)	0.000	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
3	Pearson Correlation	.720[a]	.674[a]	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
	Sig. (two-tailed)	0.000	0.000	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
4	Pearson Correlation	.577[a]	.540[a]	.646[a]	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
	Sig. (two-tailed)	0.000	0.000	0.000	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
5	Pearson Correlation	.615[a]	.541[a]	.733[a]	.606[a]	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
	Sig. (two-tailed)	0.000	0.000	0.000	0.000	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
6	Pearson Correlation	.362[a]	.349[a]	.415[a]	.265[a]	.392[a]	. . .	. . .	. . .	. . .	. . .	. . .	. . .
	Sig. (two-tailed)	0.000	0.000	0.000	0.000	0.000	. . .	. . .	. . .	. . .	. . .	. . .	. . .
7	Pearson Correlation	.351[a]	.348[a]	.369[a]	.254[a]	.393[a]	.768[a]	. . .	. . .	. . .	. . .	. . .	. . .
	Sig. (two-tailed)	0.000	0.000	0.000	0.000	0.000	0.000	. . .	. . .	. . .	. . .	. . .	. . .
8	Pearson Correlation	.374[a]	.345[a]	.429[a]	.305[a]	.463[a]	.777[a]	.774[a]	. . .	. . .	. . .	. . .	. . .
	Sig. (two-tailed)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	. . .	. . .	. . .	. . .	. . .
9	Pearson Correlation	.504[a]	.483[a]	.510[a]	.378[a]	.521[a]	.783[a]	.800[a]	.811[a]	. . .	. . .	. . .	. . .
	Sig. (two-tailed)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	. . .	. . .	. . .	. . .
10	Pearson Correlation	.482[a]	.458[a]	.468[a]	.370[a]	.471[a]	.775[a]	.801[a]	.801[a]	.901[a]	. . .	. . .	. . .
	Sig. (two-tailed)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	. . .	. . .	. . .
11	Pearson Correlation	.470[a]	.454[a]	.506[a]	.376[a]	.535[a]	.806[a]	.814[a]	.803[a]	.886[a]	.878[a]	. . .	. . .
	Sig. (two-tailed)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	. . .	. . .
12	Pearson Correlation	.402[a]	.361[a]	.419[a]	.286[a]	.434[a]	.566[a]	.569[a]	.577[a]	.652[a]	.638[a]	.680[a]	. . .
	Sig. (two-tailed)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	. . .

[a] Correlation is significant at the 0.01 level (two-tailed).

Table 3 Bivariate correlations of questions 1–12. Single dataset, N = 91.

		Correlations N = 91											
Question (Q)		**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
1	Pearson Correlation	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
	Sig. (two-tailed)	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
2	Pearson Correlation	.753[a]	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
	Sig. (two-tailed)	0.000	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
3	Pearson Correlation	.731[a]	.681[a]	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
	Sig. (two-tailed)	0.000	0.000	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
4	Pearson Correlation	.668[a]	.622[a]	.789[a]	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
	Sig. (two-tailed)	0.000	0.000	0.000	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
5	Pearson Correlation	.654[a]	.652[a]	.752[a]	.698[a]	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
	Sig. (two-tailed)	0.000	0.000	0.000	0.000	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
6	Pearson Correlation	.468[a]	.433[a]	.568[a]	.474[a]	.558[a]	. . .	. . .	. . .	. . .	. . .	. . .	. . .
	Sig. (two-tailed)	0.000	0.000	0.000	0.000	0.000	. . .	. . .	. . .	. . .	. . .	. . .	. . .
7	Pearson Correlation	.507[a]	.437[a]	.585[a]	.529[a]	.596[a]	.845[a]	. . .	. . .	. . .	. . .	. . .	. . .
	Sig. (two-tailed)	0.000	0.000	0.000	0.000	0.000	0.000	. . .	. . .	. . .	. . .	. . .	. . .
8	Pearson Correlation	.574[a]	.460[a]	.592[a]	.530[a]	.626[a]	.768[a]	.774[a]	. . .	. . .	. . .	. . .	. . .
	Sig. (two-tailed)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	. . .	. . .	. . .	. . .	. . .
9	Pearson Correlation	.599[a]	.499[a]	.600[a]	.559[a]	.649[a]	.829[a]	.880[a]	.810[a]	. . .	. . .	. . .	. . .
	Sig. (two-tailed)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	. . .	. . .	. . .	. . .
10	Pearson Correlation	.561[a]	.427[a]	.562[a]	.516[a]	.574[a]	.814[a]	.875[a]	.800[a]	.934[a]	. . .	. . .	. . .
	Sig. (two-tailed)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	. . .	. . .	. . .
11	Pearson Correlation	.566[a]	.516[a]	.646[a]	.587[a]	.690[a]	.849[a]	.918[a]	.809[a]	.918[a]	.910[a]	. . .	. . .
	Sig. (two-tailed)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	. . .	. . .
12	Pearson Correlation	.420[a]	.430[a]	.509[a]	.489[a]	.582[a]	.580[a]	.573[a]	.607[a]	.678[a]	.647[a]	.695[a]	. . .
	Sig. (two-tailed)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	. . .

[a] Correlation is significant at the 0.01 level (two-tailed).

4.3 Sample Adequacy

Each dataset was examined for its suitability for factor analysis. Barlett's test of sphericity determines whether the variables create an "identity matrix," meaning they would be highly correlated, making factor analysis meaningless. The Full dataset Bartlett's statistic was X^2 (66) = 5702.93, $p < 0.001$, and the Single dataset was X^2 (66) = 1209.37, $p < 0.001$, indicating neither dataset is an identity matrix; thus, both are suitable for factor analysis.

The Kaiser–Meyer–Olkin (KMO) measure of sampling adequacy was used to determine whether it is appropriate to examine the datasets using factor analysis. The Full dataset KMO statistic was 0.939, and the Single dataset was 0.934. KMO values between 0.5 and 1.0 indicate factor analysis is appropriate for the examined data.

4.4 Communalities

Each dataset was examined for shared variance between variables. Communalities are shown in Table 4. Values can be between 0 and 1; the closer the value is to 1 the better the factor explains that variable. Small values (<0.50) indicate a variable is less significant and could potentially be dropped. In both datasets Q12 is below 0.5 after extraction, which indicates this item does not fit well with the others and should (most likely) be considered for removal. As this examination primarily determines the content of an existing measure, we retained Q12.

Table 4 Communalities for both datasets, all questionnaire items are included

Question (Q)	Full		Single	
	Initial	Extraction	Initial	Extraction
1	0.677	0.713	0.724	0.676
2	0.622	0.629	0.646	0.634
3	0.701	0.767	0.758	0.808
4	0.485	0.520	0.665	0.703
5	0.617	0.613	0.702	0.710
6	0.722	0.728	0.777	0.766
7	0.738	0.761	0.876	0.868
8	0.744	0.753	0.723	0.720
9	0.867	0.892	0.913	0.921
10	0.856	0.874	0.905	0.909
11	0.861	0.886	0.927	0.935
12	0.484	0.482	0.557	0.493

Note: Extraction method = maximum likelihood.

4.5 Eigenvalues

The threshold for eigenvalue determination for factors was 1.0. As shown in Table 5, analysis of the datasets indicate there are two factors, which account for over 71% of the total variance.

Table 5 Eigenvalues and explained variance for the full and single datasets

Factor	Full			Single		
	Initial eigenvalues			Initial eigenvalues		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	7.20	59.99	59.99	8.13	67.79	67.79
2	1.98	16.48	76.47	1.47	12.27	80.05
3	0.56	4.69	81.16	0.54	4.51	84.56
4	0.52	4.32	85.48	0.43	3.61	88.17
5	0.39	3.29	88.76	0.31	2.56	90.73
6	0.28	2.30	91.06	0.29	2.41	93.14
7	0.24	1.98	93.04	0.24	2.03	95.17
8	0.22	1.81	94.84	0.20	1.69	96.86
9	0.21	1.73	96.57	0.15	1.25	98.11
10	0.21	1.71	98.28	0.11	0.93	99.04
11	0.11	0.92	99.21	0.06	0.52	99.56
12	0.10	0.79	100.00	0.05	0.44	100.00
Factor	Extraction sums of squared loadings			Extraction sums of squared loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	6.85	57.07	57.07	7.75	64.57	64.57
2	1.77	14.76	71.83	1.39	11.62	76.20

Note: Extraction method = maximum likelihood

4.6 Goodness of Fit

The goodness-of-fit test results show the two-factor model fits the data well for both datasets (Table 6).

Table 6 Goodness-of-fit test results

Dataset	Goodness-of-fit test		
	Chi-square	df	Sig.
Full	197.184	43	0.000
Single	67.596	43	0.010

4.7 Factor Analysis

For the factor analyses, both datasets were examined: 1) using an orthogonal rotation (Table 7), wherein the factors do not correlate; and 2) using an oblique rotation (Table 8), wherein the factors are allowed to correlate. Factors with values less than 0.3 are not shown. The factor correlation matrix generated with the oblique rotation indicates the two factors are moderately correlated. Note that all the original questions were included, but Q12 was somewhat lower than the others.

Table 7 Factor analysis results using varimax (orthogonal) rotation

Question (Q)	Rotated factor matrix			
	Full Factor		Single Factor	
	1	2	1	2
1	. . .	0.81	. . .	0.75
2	. . .	0.76	. . .	0.76
3	. . .	0.84	0.35	0.83
4	. . .	0.70	0.31	0.78
5	. . .	0.72	0.43	0.73
6	0.83	. . .	0.82	0.30
7	0.86	. . .	0.88	0.32
8	0.84	. . .	0.75	0.40
9	0.88	. . .	0.89	0.37
10	0.88	. . .	0.91	. . .
11	0.88	. . .	0.88	0.40
12	0.62	. . .	0.59	0.38

Note: Extraction method = maximum likelihood. Rotation method = varimax with Kaiser normalization.

Table 8 Factor analysis results using Oblim (oblique) rotation, with factor correlations

Question (Q)	Pattern matrix			
	Full Factor		Single Factor	
	1	2	1	2
1	. . .	0.84	. . .	0.79
2	. . .	0.79	. . .	0.85
3	. . .	0.87	. . .	0.88
4	. . .	0.75	. . .	0.84
5	. . .	0.71	. . .	0.71
6	0.88	. . .	0.90	. . .
7	0.93	. . .	0.97	. . .
8	0.89	. . .	0.76	. . .
9	0.89	. . .	0.95	. . .
10	0.91	. . .	1.01	. . .
11	0.90	. . .	0.92	. . .
12	0.62	. . .	0.56	. . .
Factor	**Factor correlation matrix**			
	Full Factor		**Single Factor**	
	1	**2**	**1**	**2**
1	1.000	0.559	1.000	0.691
2	0.559	1.000	0.691	1.000

Note: Extraction method = maximum likelihood. Rotation method = Oblimin with Kaiser normalization.

4.8 Factor Analysis Summary

Factor analysis was used to examine the Jian et al. (2000) TIA survey, with the objective to determine whether it assesses a single factor or two factors. Similar to Spain et al. (2008), we compared both orthogonal and oblique rotation methods in our analysis and found the oblique rotation model fit the data slightly better than the orthogonal model. Our analysis indicates there are indeed two distinct, but related, factors represented in this scale. In both models, Q12 had the lowest loadings; however, these were higher than in Spain et al.'s original examination (0.34). Considering the high correlations shown in the original examination and the low loadings for Q12, it may be prudent to re-examine the content of this scale. However, our original concern (that the scale may be too dated to still be useful) appears to have been addressed satisfactorily. As such, we next report a re-analysis of data from a recent study to see if using the scale as a two-factor measure yields more informative results.

5. Using Jian et al.'s TIA Scale as a Two-Factor Measure

To examine analyses of trust and distrust as separable factors, we analyzed data from a recent study conducted by ARL (Wright et al. 2024).

5.1 Experimental Design

That study was a 2 × 3 × 2 fractional–factorial design experiment. Within-subjects evaluations compared differences in performance and attributions regarding the agent across levels of transparency of agent reasoning and learning, and task load on a threat classification task. Between-subjects variables are not relevant to these analyses and are not discussed.

5.2 Independent Variables

The relevant independent variables were Aided Target Recognition (AiTR), which included reasoning (Reasoning Transparency: opaque vs. transparent), AiTR learning (Learning Transparency: implicit, explicit, human-directed), and task load (Task Load: low, high). Because it was not feasible to have a condition with reasoning transparency and learning transparency, or transparent learning with opaque reasoning, these were not included (see Table 9). In the baseline condition, Opaque Reasoning/Implicit Learning (T1), the agent conducts the task without sharing its reasoning for classifying persons. In the Transparent Reasoning/Implicit Learning (T2) condition, the agent shares its reasoning for classifying persons, but does not indicate how the participant's input is used in updating its underlying reasoning. In the Transparent Reasoning/Explicit Learning (T3) condition, the agent shares its reasoning for classifying persons and backfills the reasoning information based on participant input to indicate what it inferred from the human's input. Finally, in the Human-Directed Learning (T4) condition, the agent shares its reasoning for classifying persons and the participant completes any missing reasoning information before identifying whether the target is a potential threat. Participants completed four scenarios, one in each transparency combination condition. Scenarios were approximately 16 min long, and task load was varied in 4-min increments (e.g., high, low, high, low) by increasing/decreasing the event rate (all persons encountered).

Table 9 Reasoning by learning transparency condition matrix

		Learning		
		Implicit	**Explicit**	**Human-directed**
Reasoning	**Opaque**	T1	. . .	. . .
	Transparent	T2	T3	T4

5.3 Dependent Measures

For these analyses, the only components of Wright et al. (2024) that were considered were the trust ratings and perceived reliability. For trust, after each scenario, participants completed the modified Jian et al. (2000) TIA survey. For perceived reliability, this was assessed several times during the scenario using probes asking participants about the mission. A mean across the mission was used for the analyses.

6. Results

For our first analysis, we considered the Jian et al. measure as a single construct, that is, as a measure of trust overall. Following this, we conducted the same analyses but separated out trust and distrust as distinct measures.

6.1 Unidimensional Analyses

A one-way repeated measures analysis of variance (ANOVA) was conducted to compare scores on the TIA survey for each of the trials (i.e., T1, T2, T3, T4) to examine the effect of reasoning and learning transparency on trust. There was an overall significant effect of transparency on trust: Wilks' $\Lambda = 0.641$, $F(3, 17) = 3.18$, $p = 0.051$, *partial eta squared* = 0.36.

Paired t-tests showed that trust was consistent in trials T1, T2, and T3, indicating there was no significant difference in trust scores due to either reasoning or learning transparency. However, there was a significant difference in trust scores due to human-directed learning. Trust scores in T4 were significantly lower than those in T3 (transparent learning) ($p < 0.01$, $d_s = 0.62$). Surprisingly, trust scores in the human-directed learning condition were also significantly lower than those in T2 ($p < 0.05$, $d_s = 0.45$), which indicates that human-directed learning was more damaging to trust in the agent than opaque agent learning.

6.2 Bidimensional Analyses

The Jian et al. TIA survey was partitioned such that the five distrust items and seven trust items were averaged separately. This created a trust score and a distrust score. A one-way repeated measures ANOVA was conducted to compare trust and distrust scores for each of the trials (i.e., T1, T2, T3, T4) to examine the effect of reasoning and learning transparency on trust versus distrust. The means and SDs are presented in Table 10.

Table 10 Descriptive statistics for average trust and distrust scores across trials

Trial	Trust_Type	N	Mean	SD	SE
Trial_1	Trust	20	4.671	1.175	0.263
	Distrust	20	2.420	1.124	0.251
Trial_2	Trust	20	4.971	1.008	0.225
	Distrust	20	2.230	1.193	0.267
Trial_3	Trust	20	5.164	1.020	0.228
	Distrust	20	2.140	1.105	0.247
Trial_4	Trust	20	4.500	0.923	0.206
	Distrust	20	2.550	1.159	0.259

There was no significant effect across trials—that is, no effect of transparency on trust, $F < 1$. However, there was a significant effect of trust type: $F(1, 19) = 47.19$, $p < 0.001$. Further, there was a significant interaction between trial and trust type: $F(3, 19) = 3.07, p < 0.05$ (see Fig. 1). Paired t-tests showed that trust was consistent in trials T1, T2, and T3, indicating there was no significant difference in trust scores due to either reasoning or learning transparency. However, there was a marginally significant difference in trust scores due to human-directed learning. Trust scores in T4 were marginally significantly lower than those in T3 (transparent learning) ($p < 0.06$). Similarly, paired t-tests showed that distrust was consistent across all trials.

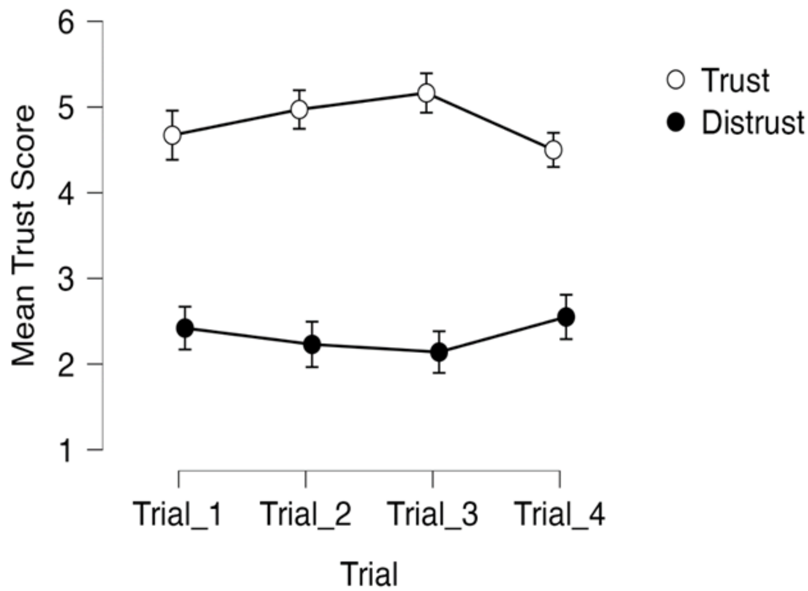

Fig. 1 Interaction between transparency trial and trust type

6.3 Correlational Analyses

As a form of convergent validity, to further test the differential utility of separating out trust and distrust, we ran correlational analyses. Thus, we additionally considered the associative strength of any relationship between trust and distrust across the trials. Table 11 provides the correlations. The results show that trust and distrust are significantly negatively correlated across T1, T2, and T3. For T4, the trial showing an increase in distrust and decrease in trust in the agent, the correlation between trust and distrust is not significant.

To examine a form of discriminant validity for the distinction, an additional proxy for trust was used. Specifically, as part of the primary experiment, one of the items to measure situation awareness on each trial assessed perceived reliability of the agent. Because reliability is conceptually similar to trust, we correlated this item with both trust and distrust from the TIA. Note that this question is worded such that higher scores mean less perceived reliability (higher perceived errors on the part of the agent). When looking at trust, for T1 through T4, there was a significant negative correlation between trust and reliability. When considering distrust, the correlation with reliability was positive across trials, but only significant for T1. This suggests that the trust items in Jian et al. (2000) are associated with agent perceived reliability, but the distrust items are more conceptually distinct in that they are not correlated with perceptions of reliability. This provides further evidence that it is of value to separately consider trust and distrust in agent automation.

Table 11 Correlations between trust and distrust and reliability across trials

Variable	Trust_T1	Distrust_T1	Trust_T2	Distrust_T2	Trust_T3	Distrust_T3	Trust_T4	Distrust_T4	T1_Rel	T2_Rel	T3_Rel	T4_Rel
Trust_T1	—											
	—											
Distrust_T1	-0.664	—										
	0.001	—										
Trust_T2	0.706	-0.367	—									
	< .001	0.111	—									
Distrust_T2	-0.552	0.431	-0.622	—								
	0.012	0.058	0.003	—								
Trust_T3	0.634	-0.348	0.913	-0.527	—							
	0.003	0.133	< .001	0.017	—							
Distrust_T3	-0.449	0.589	-0.478	0.592	-0.646	—						
	0.047	0.006	0.033	0.006	0.002	—						
Trust_T4	0.217	-0.148	0.569	-0.162	0.549	-0.211	—					
	0.358	0.534	0.009	0.496	0.012	0.372	—					
Distrust_T4	-0.487	0.386	-0.486	0.641	-0.407	0.481	-0.386	—				
	0.029	0.093	0.030	0.002	0.075	0.032	0.093	—				
T1_Rel	-0.535	0.501	-0.290	-0.013	-0.247	0.060	-0.309	0.339	—			
	0.015	0.025	0.214	0.957	0.295	0.803	0.184	0.144	—			
T2_Rel	-0.256	0.197	-0.548	0.155	-0.488	0.178	-0.737	0.373	0.573	—		
	0.276	0.405	0.012	0.514	0.029	0.454	< .001	0.106	0.008	—		
T3_Rel	-0.061	-0.108	-0.610	0.307	-0.608	0.254	-0.518	0.244	0.139	0.624	—	
	0.798	0.650	0.004	0.189	0.004	0.279	0.019	0.300	0.560	0.003	—	
T4_Rel	0.076	-0.056	-0.424	0.136	-0.400	0.184	-0.593	0.351	0.165	0.677	0.727	—
	0.749	0.816	0.063	0.569	0.080	0.438	0.006	0.129	0.487	0.001	< .001	—

7. Conclusions

This study set out to replicate and extend the findings on the TIA scale Jian et al. (2000) developed to further our understanding of attitudes toward technology. First, we conceptually replicated the factor analyses and documented that the TIA scale is better conceptualized as two distinct subscales: trust and distrust (Spain et al. 2008). As with Spain et al., we show that the first five questions assess distrust in automation, the following six note trust in automation, and one shows familiarity with automation. This last question is categorized as a trust question.

Second, we extended these findings by using this distinction to analyze data from a study of human–machine teaming that examined variations in agent transparency. We show that, when considering trust and distrust as separable items, there is a significant interaction in participants' perceptions of trust versus distrust. Further, through correlational analyses, we provide some support for convergent validity by showing that the responses on trust and distrust are significantly negatively correlated. We also provide some evidence for discriminant validity by showing that response to the trust items of the TIA scale are significantly correlated with a measure of agent reliability while responses to the distrust items are not.

Our findings strongly suggest that future uses of the TIA scale need to consider the trust and distrust items as separate. Rather than aggregating the items into a single measure of trust, as recommended by Jian et al. (2000), future work should examine that against using this scale as two distinct assessments. Finally, it is recommended that the scale be examined with the goal of improvement. Several of the questions have high correlations, indicating they are too similar to add nuance, and the scale could potentially be condensed into fewer questions. The wording of Q12 should be examined so it is less ambiguous. However, it may be equally useful to examine the effect of removing the question entirely.

8. References

Cox KR, Rexwinkle JT, Gremillion GM, Perelman BS, Brooks JR, Dyer P, Pollard KA, Forster DE, Chhan D, Carter EC, et al. Effects of technology aids and crew size on a next generation combat vehicle platoon section. DEVCOM Army Research Laboratory (US); 2022. Report No.: ARL TR-9403.

Drnec K, Gremillion G, Donavanik D, Canady JD, Atwater C, Carter E, Haynes, BA, Marathe AR, Metcalfe JS. The role of psychophysiological measures as implicit communication within mixed-initiative teams. In Virtual, Augmented and Mixed Reality: Interaction, Navigation, Visualization, Embodiment, and Simulation: 10th International Conference, VAMR 2018, Held as Part of HCI International 2018; 2018 July 15–20; Las Vegas, NV, USA. Proceedings, Part I 10. p. 299–313. Springer International Publishing.

Gremillion GM, Cox KR, Perelman BS, Hill SG, Forster DE, Chann D, Fitzhugh SM, Pollard KM, Thurman SM, Rexwinkle JT, et al. Preliminary assessment of manpower needs and task-sharing in the next Generation Combat Vehicle (Summary Technical Report, Oct 2019–Nov 2020). DEVCOM Army Research Laboratory (US); 2022 Feb. Report No.: ARL-TR-9408.

Gremillion GM, Rexwinkle JT, Cox KR, Brooks JR, Dyer P, Kucukosmanoglu M, Giammanco CA, Hung CP, Napier SJ, Rohaly TR, et al. Technologies to cue and support team tasking and coordination in the Next Generation Combat Vehicle. DEVCOM Army Research Laboratory (US); Forthcoming 2024.

Jian J-Y, Bisantz AM, Drury CG. Foundations for an empirically determined scale of trust in automated systems. Int J Cogn Ergon. 2000;4(1):53–71.

Spain RD, Bustamante EA, Bliss JP. Towards an empirically developed scale for system trust: take two. Proc Hum Factors Ergon Soc Annu Meet. 2008;2(19):1335–1339.

Wildman JL, Fiore SM, Salas E. Trust versus distrust: separate constructs or ends of a continuum? Poster presented at the 4th annual meeting of the Interdisciplinary Network for Group Research; 2009; Colorado Springs, CO.

Wildman JL, Thayer AL, Warren C, Fiore SM, Salas E. Interpersonal trust and distrust at work: scale validation and theoretical validation. Unpublished Manuscript; 2024.

Wright JL, Camacho G, Lee J, Schreck JA, Lakhmani SG. HAT with LCAs: how real-time teaching affects human performance. DEVCOM Army Research Laboratory (US); Forthcoming 2024.

Wright JL, Wigness M, Rogers J. Pilot study-robot learning from human intervention. DEVCOM Army Research Laboratory (US); 2021. Unpublished Manuscript.

List of Symbols, Abbreviations, and Acronyms

AiTR	Aided Target Recognition
ANOVA	analysis of variance
ARL	Army Research Laboratory
CFA	confirmatory factor analysis
DEVCOM	US Army Combat Capabilities Development Command
EFA	exploratory factor analysis
KMO	Kaiser–Meyer–Olkin
N	number (of)
Q	question
SD	standard deviation
SE	standard error
TIA	Trust in Automation

1 (PDF) DEFENSE TECHNICAL
INFORMATION CTR
DTIC OCA

1 (PDF) DEVCOM ARL
FCDD RLB CI
TECH LIB

1 (PDF) DEVCOM ARL
FCDD HSI
J THOMAS

1 (PDF) DEVCOM AFC
D TAMILIO

1 (PDF) DAC
L LARIMER
T DAVIS

25 (PDF) DEVCOM ARL
FCDD RLA CH
BS PERELMAN
FCDD RLA FA
E CARTER
D FORSTER
KA POLLARD
C NEUBAUER
FCDD RLA FB
GM GREMILLION
JS METCALFE
L MARUSICH-COOPER
SM FITZHUGH
FCDD RLA FC
HP CROWELL
FCDD RLA FD
JT REXWINKLE
AS KRAUSMAN
KR COX
CA GIAMMANCO
J WRIGHT
E HOLDER
AL BAKER
R COHEN-HOFFING
SJ NAPIER
S LAKHMANI
A SCHARINE
FCDD RLA FE
D CHHAN
B CESAR-TONDREAU
FCDD RLA FF
T SMITH
FCDD RLA JA
R BREWER

3 (PDF) UCF
S FIORE
G SUKTHANKAR
J FOLDES