Trust in AI:
Far-Away yet Near-Today

Peter Hancock

Department of Psychology, and
Institute for Simulation and Training,
University of Central Florida,
Orlando, FL USA

Presentation Given to the National Academy of Sciences, Engineering, Medicine.
Human-AI Teaming Workshop (Sponsored by AFRL), July 28-29, 2021.


JOB FAIR
All 7.85 Billion of Them?

Back to 'Reality' of Science, Let's Start with Definitions:

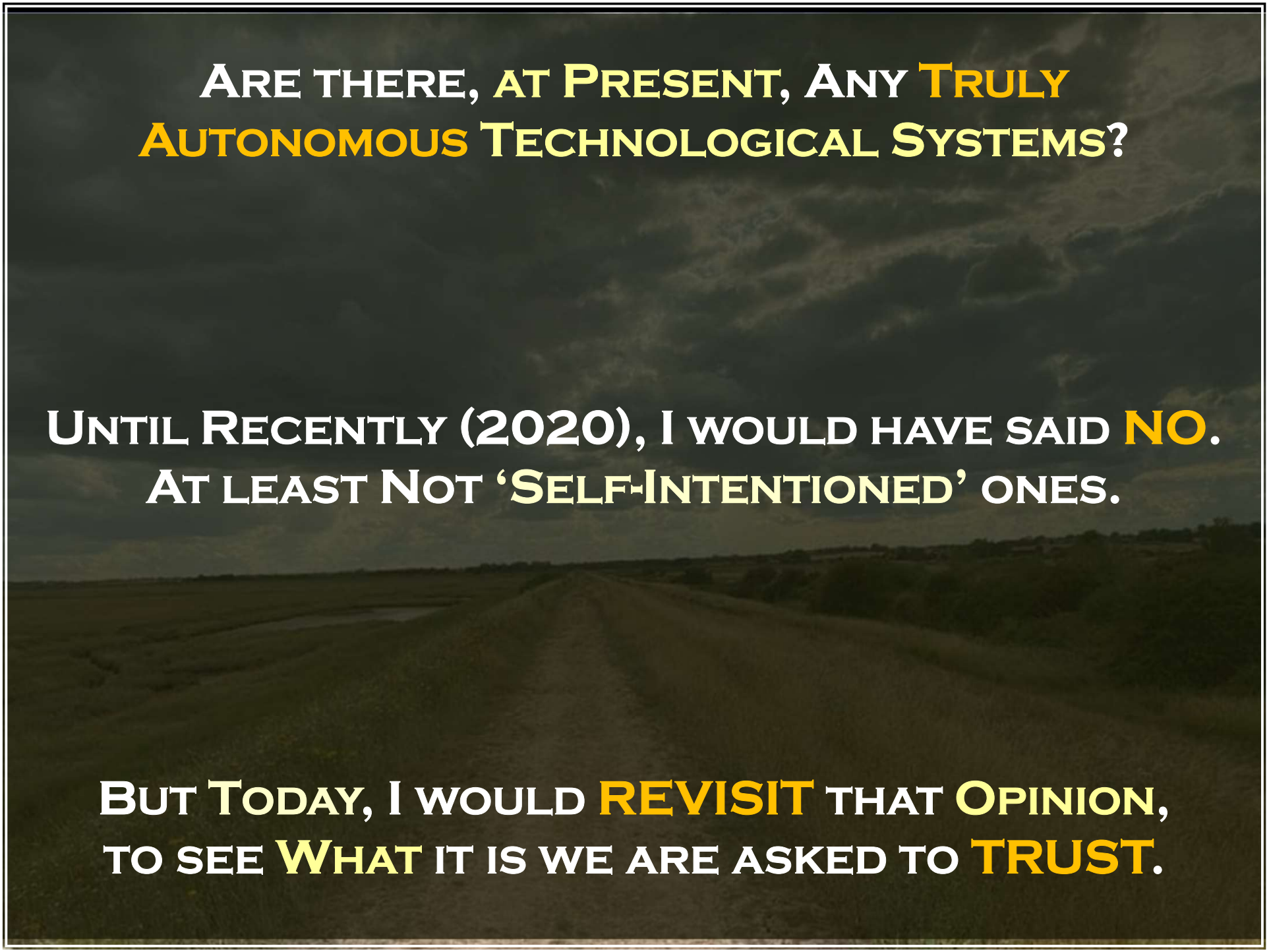
Automated Systems are designed to accomplish a specific set of largely deterministic steps (often in a repeating pattern) in order to achieve one of a limited set of pre-defined goals.

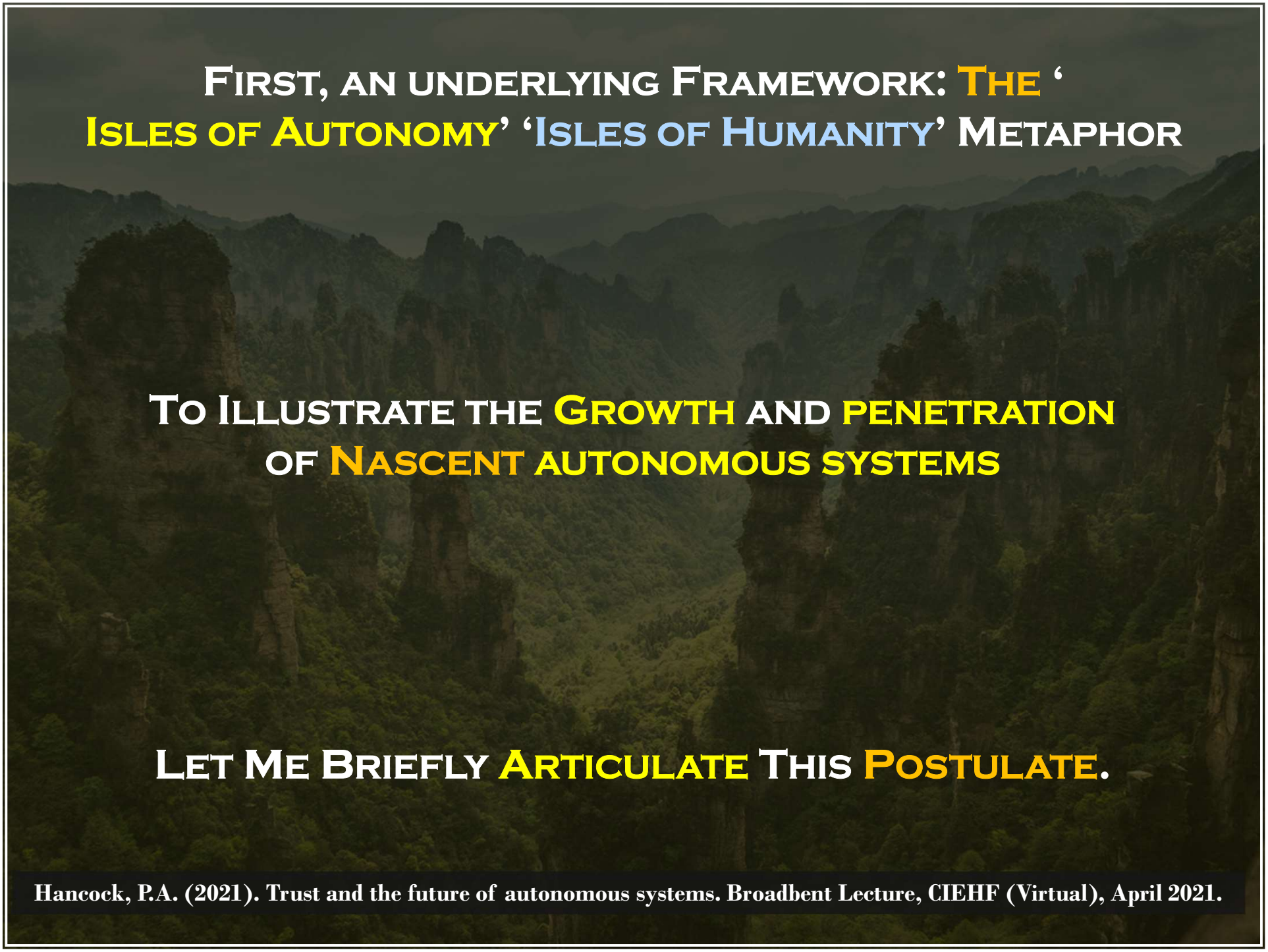
The Foundation of Autonomy Derives from the Etymological Bases: Nomos – 'Law' and Auto - 'Self'. Thus, autonomous systems are Laws Unto Themselves.

Autonomous Systems, are generative; they learn and evolve through feedback of operational and contextual information. Their actions necessarily become more indeterminate across time. (Hancock, 2017).

However, I Now Believe There is no Necessary, disruptive 'Threshold' between these Two forms.

Hoffman, R.R., & Hancock, P.A. (2014). Words matter. *Human Factors and Ergonomics Society Bulletin*, *57* (8), 3-7.

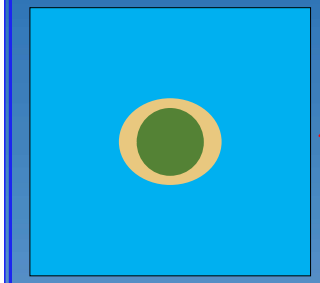

Are there, at Present, Any Truly Autonomous Technological Systems?
Until Recently (2020), I would have said NO.
At least Not 'Self-Intentioned' ones.
But Today, I would REVISIT that Opinion,
to see What it is we are asked to TRUST.

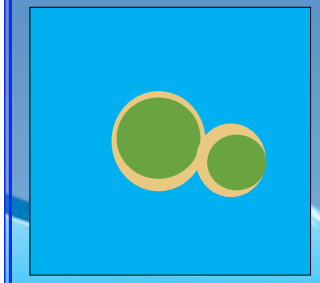


Hancock, P.A. (2021). Trust and the future of autonomous systems. Broadbent Lecture, CIEHF (Virtual), April 2021.

The Isles of Autonomy Metaphor

(Represented as a Series of Phase Transitions)





At First, there is a **SINGULAR ISLAND** of Autonomy which is **SURROUNDED** by a **SWATH** of Human Supporters, Shown as the **BEACH**. This forms a **LITTORAL** (Literal) Collaboration

A Critical Watershed Occurs when Discrete Autonomous Systems **LINK TOGETHER**. The **ISTHMUS** is First Composed of **HUMANS**

Here, Autonomy is 'Nominally' Proceeding From **TOOL** to **TEAM-MATE**

Sea Level = Relative Ratio of Human to Autonomous Capacities

Early Suggestions of this 'Linking Stage' can be Illustrated by a recent demonstration

Here, SOPHIA the Robot Communicates with JACK the Automated Car

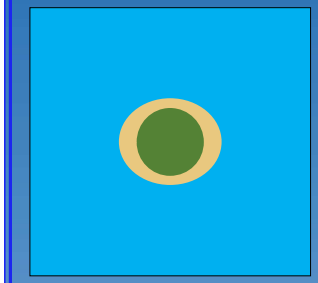
The Video: https://www.youtube.com/watch?v=vtX-qVUfCKI.

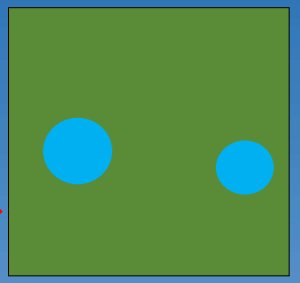
The Human as an "Intelligent Intermediary"

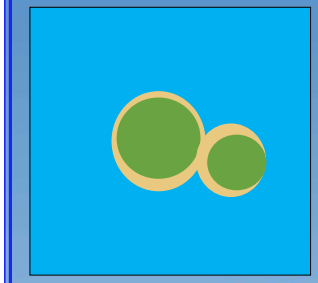
Sophia the Robot meets Jack the Car! https://www.youtube.com/watch?v=vtX-qVUfCKI

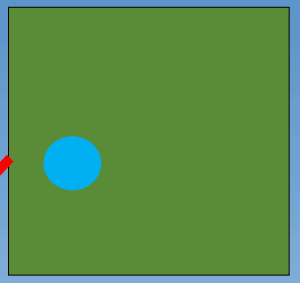
FORESEEABLE PROGRESS

(THE OTHER PHASE TRANSITIONS ARE)









Slowly, Autonomous Systems Begin to Dominate the Electronic Ecosphere. Humans are Sequentially Squeezed Out

Eventually, ANY Remaining "ISLE OF HUMANITY" Becomes Palimpsestual, Residual, Deliquescent, and then Extinct.

Hancock, P.A. (2017). Imposing limits on autonomous systems. *Ergonomics*, *60* (2), 284-291.

Hancock, P.A. (2014). Automation: How much is too much? *Ergonomics*, *57* (3), 449-454.

The Following is Based on Kaplan et al., (2021).

(Copies of the Paper are Authorized for Release and Publicly Available).

Thanks go to Dr. J. Christopher Brill (AFRL/711th Human Performance Wing) for his leadership and graduate student mentorship throughout the project. Thanks also go to Dr. Joseph Lyons for all his help and providing insightful reviews of this work. This research was originally sponsored by the AFRL Human Insight and Trust (HIT) Program (Chris Brill, Manager).

Trust in Artificial Intelligence: Meta-Analytic Findings

Alexandra D. Kaplan, University of Central Florida, Orlando, Florida, USA, **Theresa T. Kessler**, Georgia Tech Research Institute, Atlanta, Georgia, USA, **J. Christopher Brill**, Air Force Research Laboratory, Dayton, Ohio, USA, and **P. A. Hancock**, University of Central Florida, Orlando, Florida, USA

Objective: The present meta-analysis sought to determine significant factors that predict trust in artificial intelligence (AI). Such factors were divided into those relating to (a) the human trustor, (b) the AI trustee, and (c) the shared context of their interaction.

Background: There are many factors influencing trust in robots, automation, and technology in general, and there have been several meta-analytic attempts to understand the antecedents of trust in these areas. However, no targeted meta-analysis has been performed examining the antecedents of trust in AI.

Method: Data from 65 articles examined the three predicted categories, as well as the subcategories of human characteristics and abilities, AI performance and attributes, and contextual tasking. Lastly, four common uses for AI (i.e., chatbots, robots, automated vehicles, and nonembodied, plain algorithms) were examined as further potential moderating factors.

Results: Results showed that all of the examined categories were significant predictors of trust in AI as well as many individual antecedents such as AI reliability and anthropomorphism, among many others.

Conclusion: Overall, the results of this meta-analysis determined several factors that influence trust, including some that have no bearing on AI performance. Additionally, we highlight the areas where there is currently no empirical research.

Application: Findings from this analysis will allow designers to build systems that elicit higher or lower levels of trust, as they require.

Keywords: artificial intelligence, trust, human–automation interaction, meta-analysis

Address correspondence to Alexandra D. Kaplan, Department of Psychology, University of Central Florida, 4111 Pictor Lane, Orlando, FL 32826, USA; e-mail: adkaplan@knights.ucf.edu

HUMAN FACTORS
Vol. 00, No. 0, Month XXXX, pp. 1-25
DOI:10.1177/00187208211013988

ARTIFICE IN AN INTELLIGENT AGE

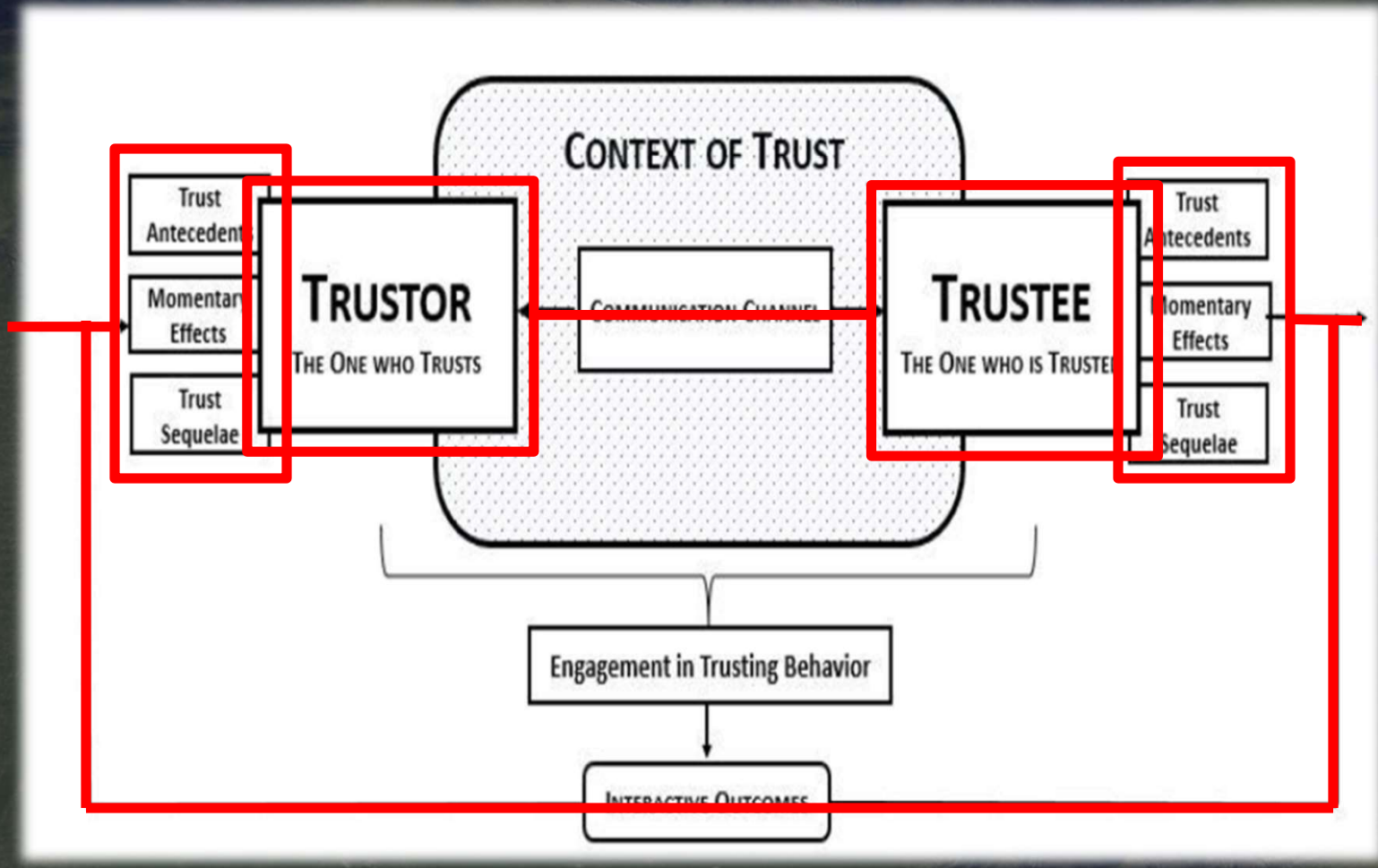
Artificial intelligence (AI) is the software-based technology that permits automated machines to sense their surroundings and intelligently make decisions based on the available data. With those summated information inputs, they are able to decide which actions are most likely to lead to their success in achieving a goal (Poole et al., 1998). Although some forms of AI employ distinctly inhuman thought processes, others are considered to be imitations of natural human cognitive abilities. McCarthy et al. (1955) noted that an artificially intelligent computer had to possess some form of an abstract model of its environment. It had to be somewhat imaginative, it had to display originality and common sense, and finally, it had to deal with "randomness." In humans, these qualities require some degree of intelligent rationalization. In automation, it can be much the same. To make optimal decisions, automation, like humans, must factor in aspects of the surrounding environment and take actions based on whatever choice leads to the highest probability of success. Even so, automation is most often typified by brittle determinism that renders it distinctly unintelligent, which is what distinguishes it from truly intelligent artificial systems. In the following analysis, we distinguish between AI and automation by categorizing unmanned systems that perform repetitive or rote tasks based on static rules or human commands as automation, while unmanned systems that can deal with uncertainty and make a "decision" in novel or semi-novel circumstances, as AI.

AI can be applied to many circumstances. Some of these include a degree of embodiment, while others are wholly virtual in nature. Some autonomous robots use AI to determine

The Views Expressed here are that of the Author(s) and Do Not Necessarily Represent that of the Supporting Agency in Any Way.

Kaplan, A.D., Kessler, T.T., Brill, J.C., & Hancock, P.A. (2021). Trust in Artificial Intelligence: Meta-analytic findings. *Human Factors*, in press. https://journals.sagepub.com/doi/abs/10.1177/00187208211013988

Trust is a Dynamic and Cybernetic Process

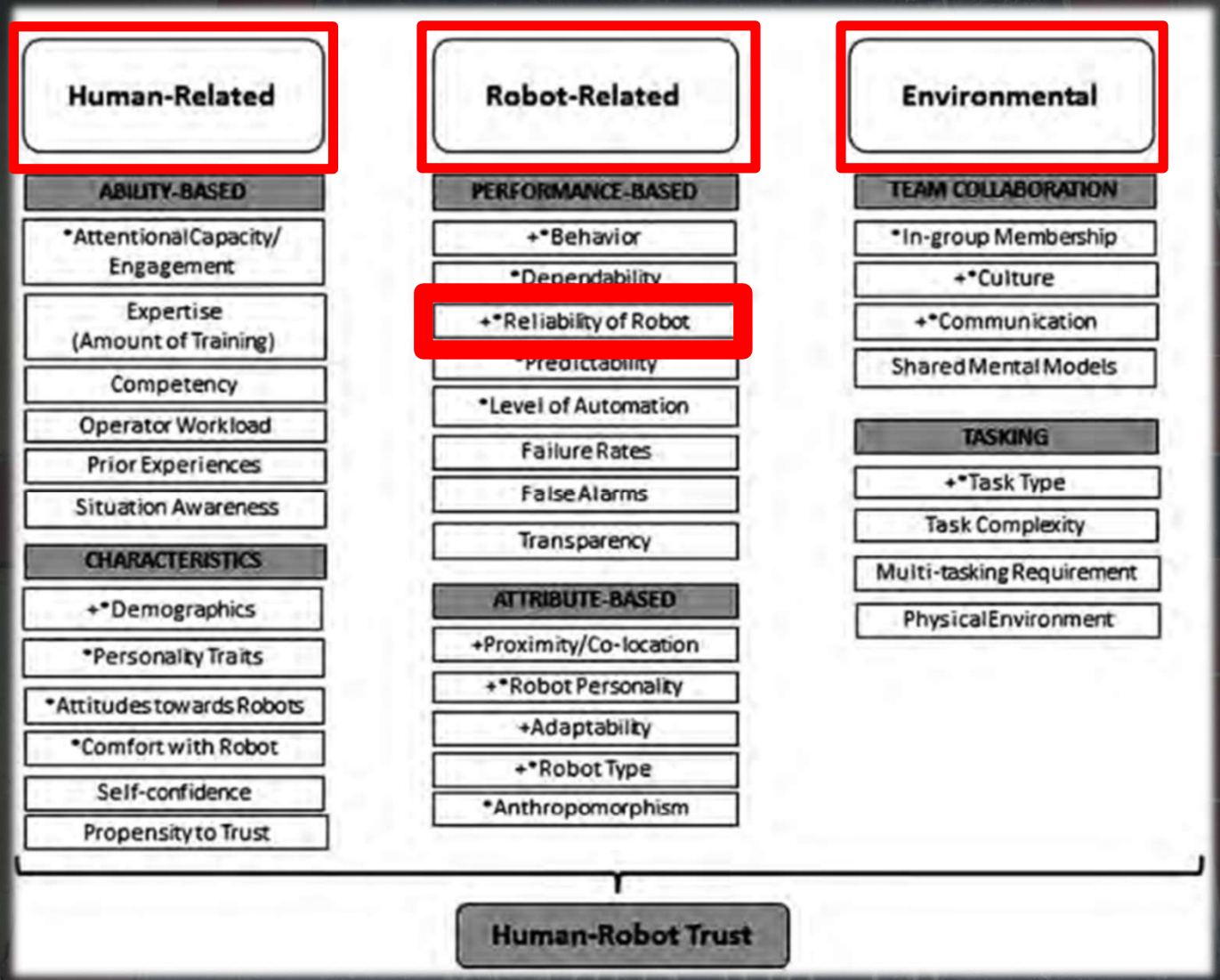


It Necessarily Requires At Least Two Entities

So, If People Tell You They Trust Themselves – Then Don't Trust Them!

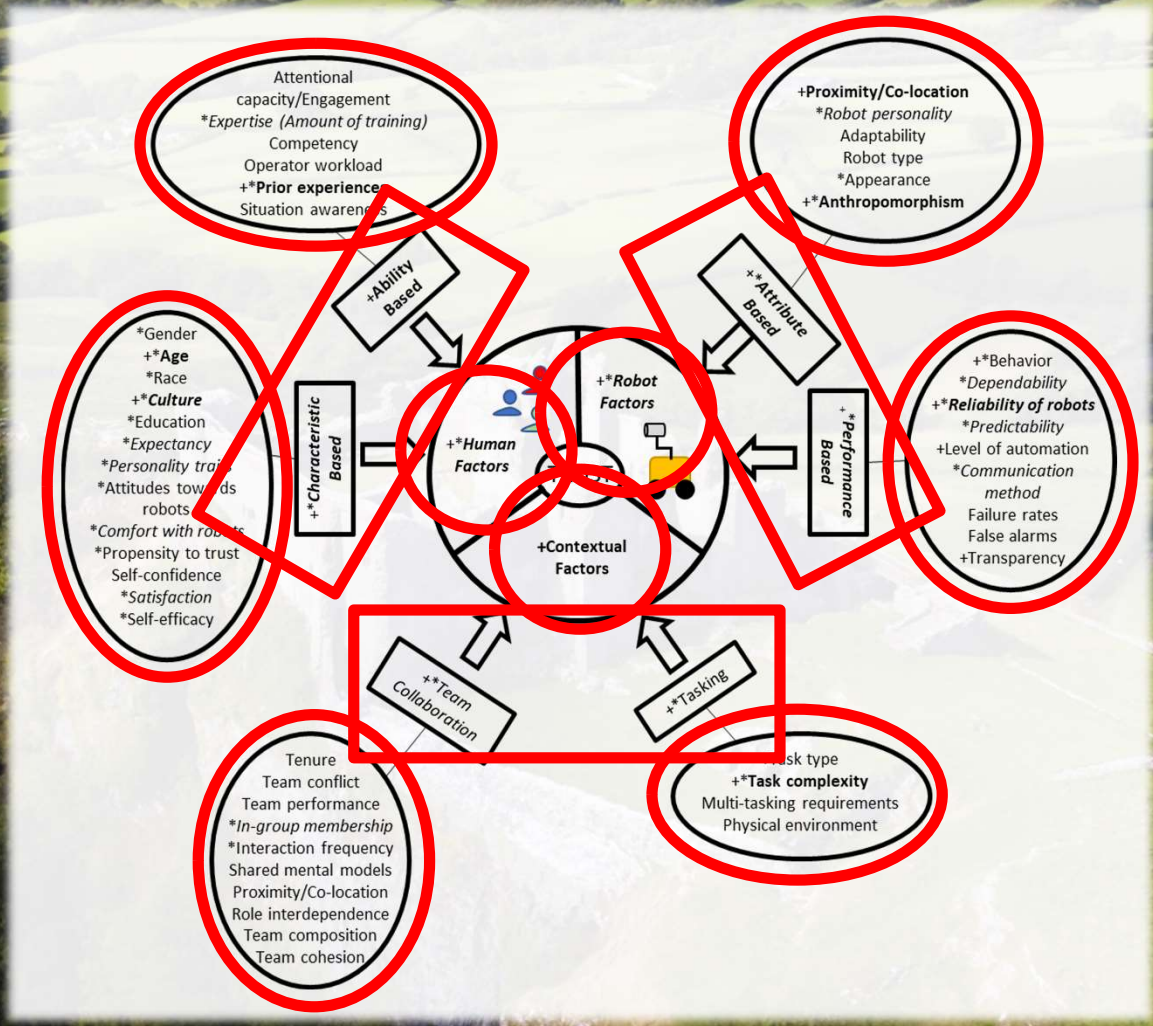
and some Communication Channel Between Them

The Foundational 'Triad' Model of Trust : 2011



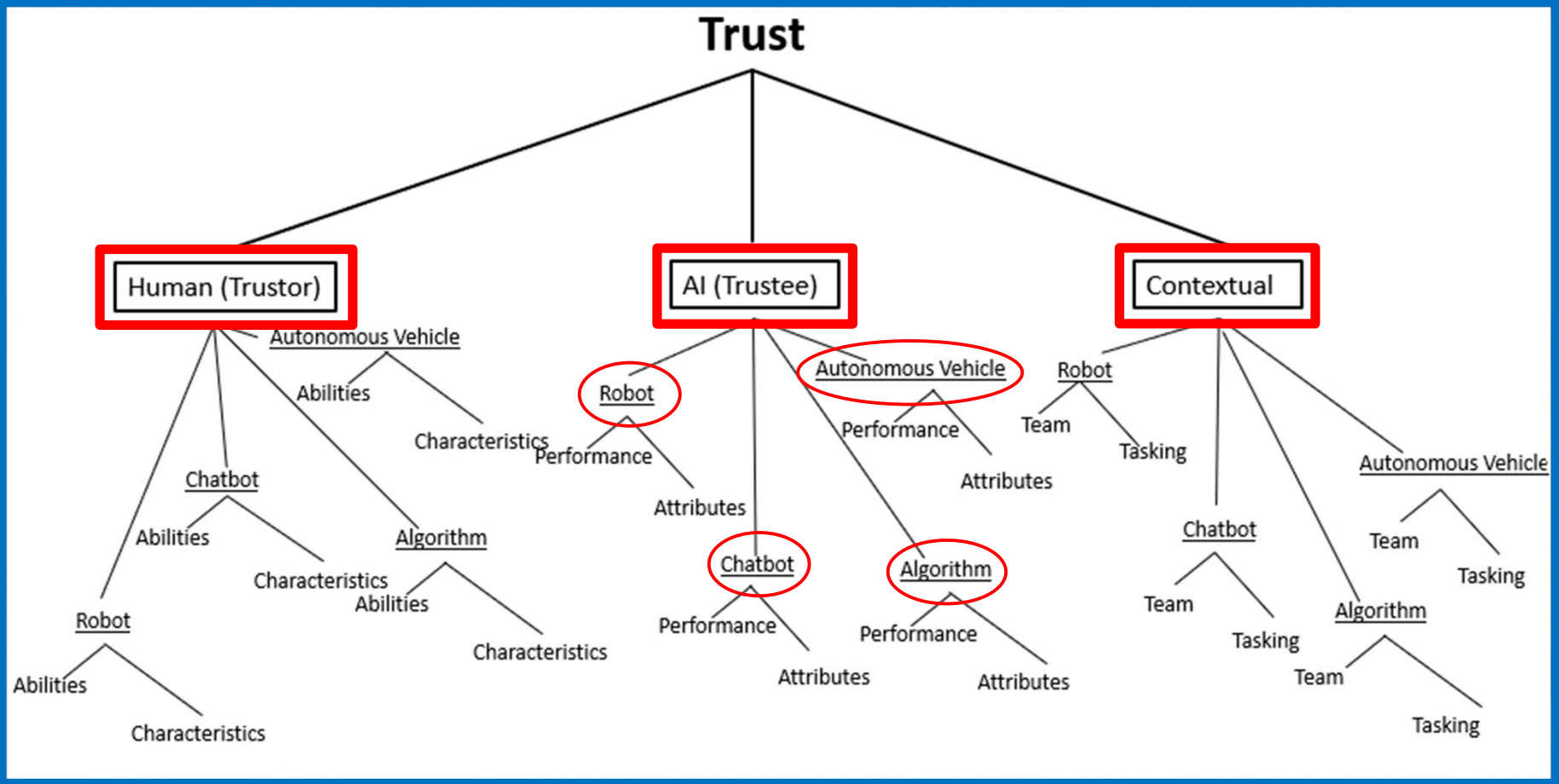
Hancock, P.A., Billings, D.R., Olsen, K., Chen, J.Y.C., de Visser, E.J., & Parasuraman, R. (2011). A meta-analysis of factors impacting trust in human-robot interaction. *Human Factors*, *53* (5), 517-527.

The Evolved 'Flower' Model of Trust: 2021



Hancock, P.A., Kessler, T.T., Kaplan, A.D., Brill, J.C. & Szalma, J.L. (2021). Evolving trust in robots: Specification through sequential and comparative meta-analyses. *Human Factors*, in press.

The Base AI-Trust Model



The CATEGORIES are Driven by the Prior MODEL Organization.
The DOMAINS are driven by Emergence in the LITERATURE.

Inclusion-Exclusion Criteria

1) The article must have come from a peer-reviewed journal or the proceedings of a conference.

2) Trust-in-AI had to be the dependent variable in at least one included analysis.

3) Sufficient statistical data had to be included to determine an effect size (e.g., *r, t, F,* means and standard deviations, *Cohen's d*, or percent).

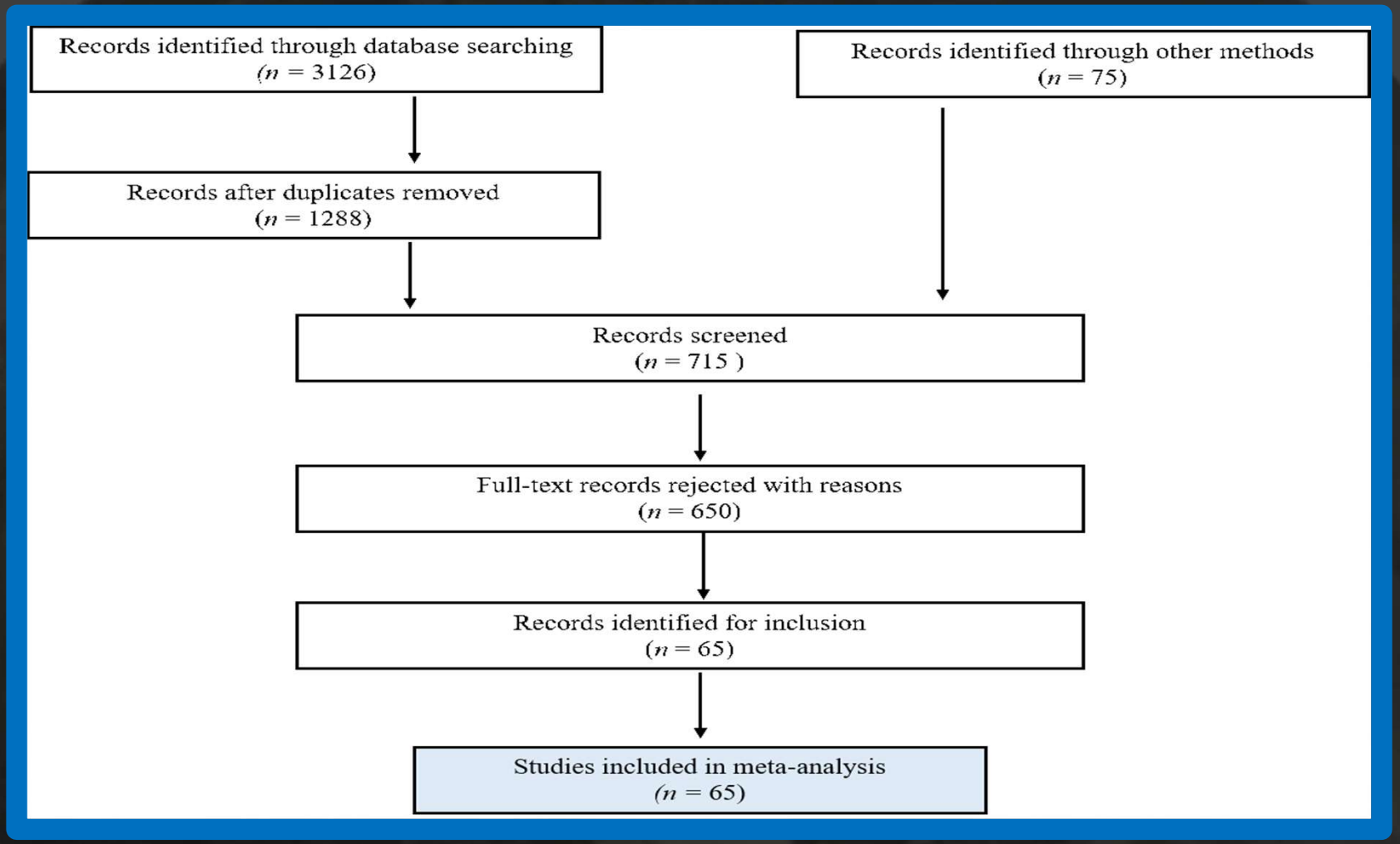
4) The examined sample could not have been derived from any vulnerable population, such as a medical population or individuals under the age of 18 years old.

5) The article had to be written in English or have an English translation available.

6) The data from the study could be included only once in the analysis. For instance, data from a dissertation could not be included if those same data were also reported in a conference publication or refereed article.

The Starting PRISMA

Preferred Reporting Items for Systematic Reviews and Meta-Analysis



Moher, D., Liberati, A.,Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic \reviews and meta-analyses: the PRISMA statement. *Annals of Internal Medicine*, *151* (4), 264-269.

Trust in AI: Human Dimensions

Antecedent of Trust	*k*	Effect size *d*	s_g^2	s_e^2	95% Confidence Interval Upper	Lower
Global	23	0.26*	0.34	0.09	0.14	0.38
Ability-based	8	0.32*	0.60	0.02	0.21	0.42
Competency/Understanding	2	1.02*	0.05	0.18	0.43	1.61
Expectancy	8	0.29	0.63	0.33	-0.11	0.69
Expertise	4	0.47*	0.14	0.20	0.03	0.91
Operator Performance	2	0.76	0.02	1.46	-0.92	2.43
Prior experience	4	-0.19	0.69	0.06	-0.43	0.05
Workload	2	-1.19	0.23	1.38	-2.82	0.44
Characteristic-based	20	0.38*	0.18	0.11	0.24	0.53
Age	2	0.09	0.02	0.02	-0.08	0.26
Attitudes Towards AI	5	1.05	0.30	3.61	-0.61	2.72
Comfort with AI	1	-0.37	---	---	---	---
Culture	2	0.51*	0.04	0.07	0.15	0.87
Education	1	0.17	---	---	---	---
Gender	3	0.42*	0.05	0.05	0.17	0.67
Personality Traits	4	0.25*	0.47	0.02	0.12	0.37
Propensity to Trust	1	0.70	---	---	---	---
Satisfaction	1	1.04	---	---	---	---

Kaplan, A.D., Kessler, T.T., Brill, J.C., & Hancock, P.A. (2021). Trust in Artificial Intelligence: Meta-analytic findings. *Human Factors*, in press. https://journals.sagepub.com/doi/abs/10.1177/00187208211013988

Trust in AI: AI Dimensions

Antecedent of Trust	*k*	Effect size *d*	s_g^2	s_e^2	95% Confidence Interval Upper	Lower
Global	48	0.62*	1.10	0.09	0.54	0.70
Performance-based	22	1.47*	1.34	0.17	1.30	1.64
Dependability	2	0.80	0.15	2.02	-1.18	2.77
Performance	13	1.48*	1.41	0.16	1.26	1.70
Predictability	2	1.42	0.67	1.85	-0.46	3.31
Reliability	5	2.70*	0.33	0.37	2.16	3.23
Attribute-based	35	0.31*	0.55	0.07	0.22	0.39
AI Personality	4	0.63*	2.39	0.04	0.42	0.83
Anthropomor-phism	10	0.30*	0.29	0.12	0.08	0.52
Appearance	1	-0.05			---	---
Behavior	6	0.81*	0.38	0.09	0.57	1.04
Communication	9	0.06	0.15	0.05	-0.08	0.20
Level of Auto-mation	2	0.03	0.00	0.01	-0.10	0.17
Reputation	5	0.68*	0.04	0.12	0.38	0.99
Transparency	9	0.24*	0.26	0.06	0.08	0.40

*Denotes significance at the $p < .05$ level.

** s_e^2 = sampling error variance; s_g^2 = observed variance

Kaplan, A.D., Kessler, T.T., Brill, J.C., & Hancock, P.A. (2021). Trust in Artificial Intelligence: Meta-analytic findings. *Human Factors*, in press. https://journals.sagepub.com/doi/abs/10.1177/00187208211013988

Trust in AI: Contextual Effects

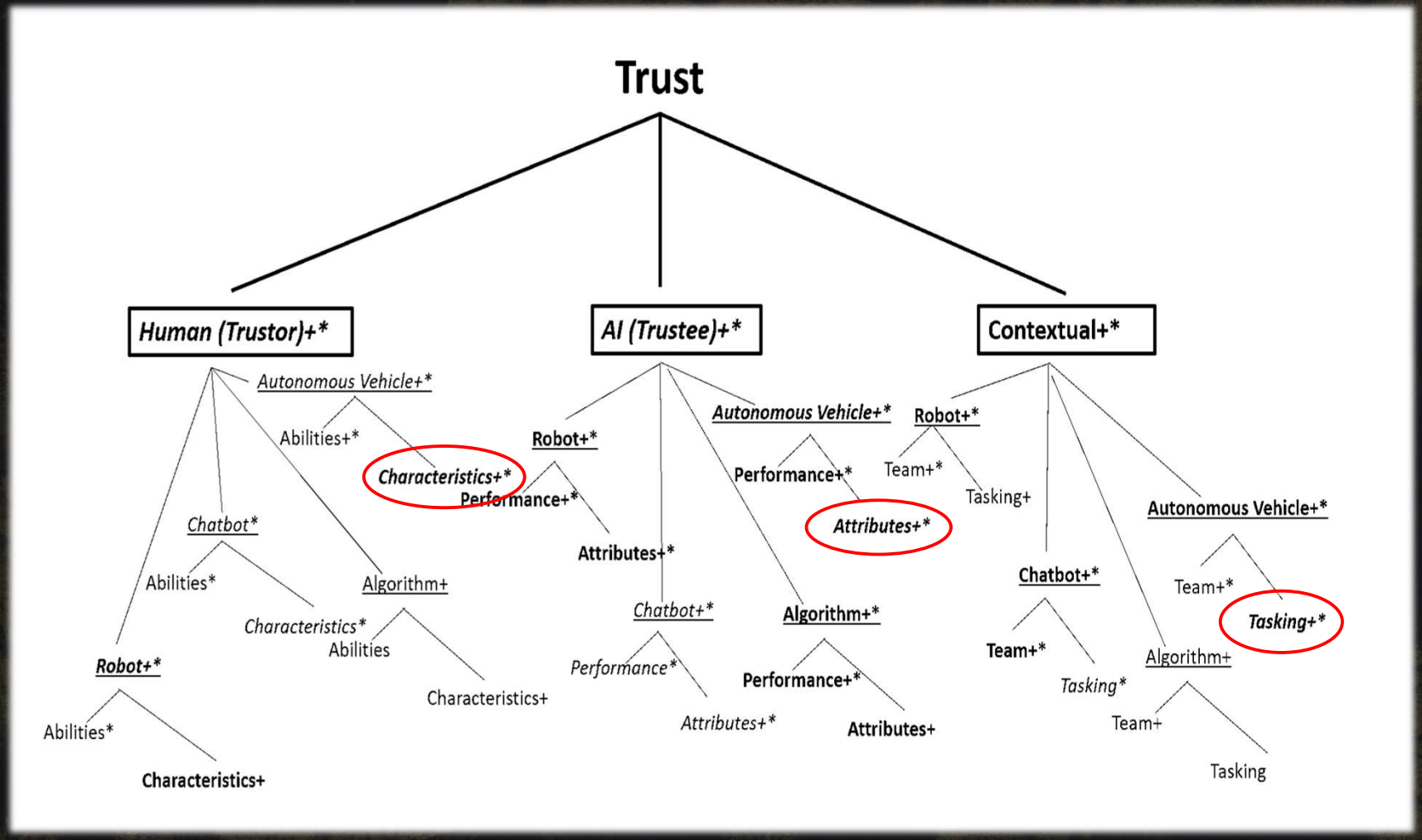
Antecedent of Trust	*k*	Effect size *d*	s_g^2	s_e^2	95% Confidence Interval	
					Upper	Lower
Global	25	0.34*	1.36	0.04	0.26	0.42
Teaming-related	15	0.39*	0.34	0.12	0.22	0.56
Communication	3	0.39*	0.6	0.03	0.19	0.60
Interaction Frequency	2	0.66	0.06	0.97	-0.71	2.03
Shared Mental Models	1	2.41	---	---	---	---
Tenure	6	0.62*	0.14	0.42	0.10	1.14
Tasking-related	12	-0.43*	0.67	0.18	-0.68	-0.19
Risk	8	-0.41*	0.25	0.33	-0.80	-0.01
Task Complexity	3	0.19	0.53	0.10	-0.16	0.54
Task Type	1	0.28	---	---	---	---

*Denotes significance at the $p < .05$ level.

** s_e^2 = sampling error variance; s_g^2 = observed variance

Kaplan, A.D., Kessler, T.T., Brill, J.C., & Hancock, P.A. (2021). Trust in Artificial Intelligence: Meta-analytic findings. *Human Factors*, in press. https://journals.sagepub.com/doi/abs/10.1177/00187208211013988

Trust in AI: An Evolving Model



The Evolving Model of the Dactors influencing Trust in AI. **Bold indicates significant pairwise findings.** ***Italics indicate significant correlational findings.*** **Asterisk (*) indicates that the subject was examined in the correlational literature. A plus sign (+) indicates that the subject was examined in the pairwise literature.**

Summary Thoughts and Concerns

Any Meta-Analysis is On hich Compose It.

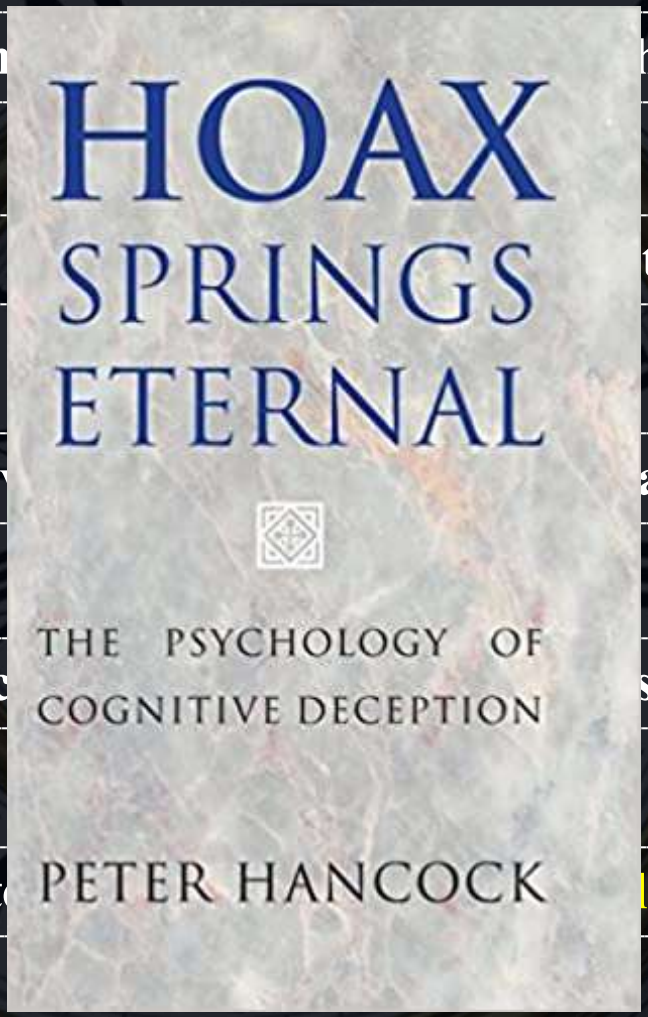
The Panoply of Rigorous t Present, Threadbare.

Perhaps Most Under-Ser and Contextual Effects.

Reliability remains Critic stablish Objective Levels?

Our Model can be Used t levolent Systems!

Studies are Needed to Stabilize Results and Elaborate on Influential Factors.



And Always, Let us Beware The Behemoth TIME

The Human Concept of "Real-Time" is an Exercise in Hubris not Reality

The Critical Issue is TEMPORAL DISSOANCE

With such Dissonance there will be no Time to Either Trust or Distrust

Hancock, P.A. (2021). Avoiding adverse autonomous agent action. *Human-Computer Interaction*, in press.

Appeasing our Forthcoming Rulers!

The Sad Truth is That (Sooner Rather than Later) Humans Will be the Rate-Limiting Element (Try to Come to Terms With It).

A Way Ahead, ... If Way There Is

The Sheepdog ...

Epitomizing Human Interaction with Capable, but Cognitively-Limited Systems. Higher-Level Goals Expressed by the Human, Lower-Level Actions Subsumed by the Dog.

Human

Technology

Context

... and the Japanese Garden

Harmonizing the Context of Operations by Intentionally Shaping Boundary Constraints. (Understanding the Limits of Adaptation and the Nature of Resilience).

Hancock, P.A. (1987). The Sheepdog and the Japanese Garden. *Essays on the Future of Human-Machine Systems*. Eden Prairie: MN Banta. [A Third of a Century Ago: And the Metaphor Story].

Questions Specific to Contributory Studies to the Meta-Analysis can be Directed TO:

Dr. Theresa Kessler

Georgia Tech Research Institute

Dr. Alexandra Kaplan

All Data Used in the Reported Meta-Analysis are Open for Public Inspection.
Work Continues to Add Further Results as they appear in the Archival Literature.
This Latter, Important Work Remains in Need of Funding.