A Practical Taxonomy of TAS-related Usecase Scenarios

Peta Masters
peta.masters@kcl.ac.uk
King's College London
United Kingdom

Victoria Young
King's College London
United Kingdom
victoria.young@kcl.ac.uk

Alan Chamberlain
University of Nottingham
United Kingdom
alan.chamberlain@nottingham.ac.uk

Sachini Weerawardhana
King's College London
United Kingdom
sachini.weerawardhana@kcl.ac.uk

Peter E. McKenna
Heriot-Watt University
United Kingdom
p.mckenna@hw.ac.uk

Yang Lu
York St John University
United Kingdom
y.lu@yorksj.ac.uk

Liz Dowthwaite
University of Nottingham
United Kingdom
Liz.Dowthwaite@nottingham.ac.uk

Paul Luff
King's College London
United Kingdom
paul.luff@kcl.ac.uk

Luc Moreau
King's College London
United Kingdom
luc.moreau@kcl.ac.uk

ABSTRACT

This paper proposes a taxonomy of experimental usecase scenarios to facilitate research into trustworthy autonomous systems (TAS). Unable to identify an open-access repository of usecases to support our research, the project team embarked on development of an online library where fellow researchers would be able to find, share and recommend usecases to other practitioners in the field. To organise the library's content, we needed a taxonomy and, informed by a commitment to responsible research and innovation (RRI), we prioritised stakeholder involvement to shape its development. Conflict arose, however, between the project team's objective—a rigorous taxonomy focused on surfacing genuine "benchmarks" that can be used to test a multiplicity of variables in a range of domains under differing experimental conditions—and stakeholder expectation that the library would provide details of particular studies and results. How then can we reconcile project requirements with stakeholder preferences? A practical solution has to be found.

CCS CONCEPTS

• **Information systems** → **Digital libraries and archives**; *Collaborative and social computing systems and tools.*

KEYWORDS

taxonomy, trust, autonomous systems

ACM Reference Format:
Peta Masters, Victoria Young, Alan Chamberlain, Sachini Weerawardhana, Peter E. McKenna, Yang Lu, Liz Dowthwaite, Paul Luff, and Luc Moreau. 2023. A Practical Taxonomy of TAS-related Usecase Scenarios. In *First International Symposium on Trustworthy Autonomous Systems (TAS '23), July 11–12, 2023, Edinburgh, United Kingdom.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3597512.3597523

TAS '23, July 11–12, 2023, Edinburgh, United Kingdom

ACM ISBN 979-8-4007-0734-6/23/07.
https://doi.org/10.1145/3597512.3597523

1 INTRODUCTION

The design and development of any shared socio-technical resource is a complex matter. Understanding the preferences and requirements of a range of stakeholders, and developing a strategy that prioritises some of these, while necessarily sacrificing others, is an activity fraught with the potential for bias and acrimony. Yet, if we are to meet our research goals and avoid an entirely unproductive stand-off or stalemate, a pragmatic solution must be found.

In this paper, we consider how the application of responsible research and innovation (RRI) principles with respect to stakeholder engagement have aided development of a taxonomy that categorises experimental usecase scenarios for the evaluation of trust and trustworthiness in human-robot interaction (HRI) and in relation to trustworthy autonomous systems (TAS). We define trust, in line with Lee and See [6], as the sense that one would choose to rely on the AS in a situation characterised by risk or vulnerability.

Our taxonomy is not being developed as an incidental artefact. It has a clearly defined purpose: to support the organisation of—and search criteria for—a library of trust-related usecase scenarios. In this online resource, those with an interest in TAS can (a) find relevant experimental usecase scenarios, (b) upload and make onward recommendations of further usecases to other library users, and (c) review or critique the usecases that they find. The stakeholders are a relatively well defined group: primarily researchers but including artists, regulators, policy-makers and potentially members of the public interested in accessing or sharing scenarios that have been or could be used to test the impact on human trust of a range of autonomous systems (AS) and situational-related variables (e.g., a robot or AS' size or behaviour [15], a human's degree of vulnerability [9], apparent shared values [7], etc.).

As discussed in [10], the proposed library meets a clear need. In the absence of such a resource, researchers typically devise one-off usecases from scratch, with the risk that resultant scenarios may be poorly thought out and poorly received. The availability of tried and tested scenarios thus not only saves time; it improves the quality of our research. Moreover, it has the benefit that reusable scenarios may be elevated to the status of "benchmarks", which we define primarily by their transferability; that is, their usefulness in testing

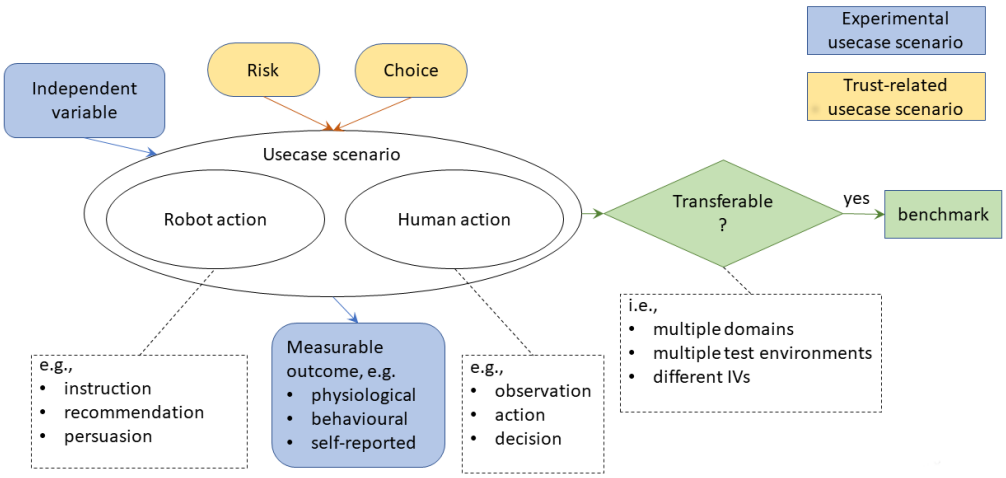

Figure 1: To be a benchmark, a usecase must be transferable.

We model a usecase scenario as an interaction between a robot or AS and a human (in white). Each performs some action, even if the human's action is only to observe. To be an *experimental* usecase (in blue), the interaction is impacted by an independent variable and there is some measurable outcome. To be trust related (yellow), the scenario involves risk/vulnerability to the human and, if the measure is behavioural, the human must have had a choice. Finally, to be a benchmark (green), the usecase must be transferable. That is, a usecase that can only be used in one set of circumstances, with one variable, is not a benchmark.

a multiplicity of variables in a range of domains under differing experimental conditions (see Figure 1). Our hope is that, thanks to their portability, the usecase scenarios surfaced by the library—via the taxonomy—will provide valuable reference points for the research community, facilitating the comparison of trust-related performance across domains and special interest areas.

The formal framework for RRI established by the European Union's Horizon program [2] drives us, as researchers, to examine the broader implications of our work and to anticipate any potential negative impacts that it may have. Our project team is multidisciplinary and diverse in terms of age, ethnicity and religious backgrounds. Where we observed a gender imbalance, we invited additional male advisors to join a predominantly female team. The library we aim to develop will require registration for contributors but, once uploaded, content will be accessible to all; its UX (usability) design prioritises accessibility without sacrificing visual appeal. Furthermore, our team meetings routinely allocate time to ethical deliberation. Thus we recognise—and seek to achieve—an integrated approach to RRI. In this paper, however, we focus on just one piece of the puzzle, the aspect of RRI which has most influenced development of the taxonomy to date: stakeholder engagement.

1.1 Related Work

A search for the terms "trust" and "taxonomy" generates over 800,000 Google Scholar search results yet we have identified only two that relate to usecases: a taxonomy of HRI failure types and their impact on trust [14] and another relating to the use of smart glasses in the domains of nursing, maintenance or logistics [4].

There are some online libraries, necessarily supported by classification systems, that include usecases. For example, a repository of "model problems" is maintained via the SEAMS (Software Engineering for Adaptive and Self-Managing Systems) symposia. The extensive Safety Pool Scenario Database is a repository of test scenarios that supports the development, verification and validation of connected and autonomous vehicles. The thousands of contributions to the latter are built on an SDL (Specification and Description Language) grammar, formalising machine/human-readable scenario definitions of variables such as junction-types, environmental conditions and vehicle manoeuvres. The former focuses on software engineering, the latter on mechanical engineering; neither relates to human factors or trust.

There are many reviews of papers that relate to trusted and trustworthy AS [e.g., 11] but these focus on the *studies* rather than the *usecases* making them of limited use in the context of experimental design. A recent review of trust-related questionnaires [5] points up a similar problem to that which we identify here, that is, the tendency of researchers to invent ad-hoc solutions or 'reinvent the wheel' for each new project. Though design-related, the focus in [5] is on the instruments used for the measurement of trust (which they categorise into three types: self-reported, behavioural and physiological). These focus on measuring the result of the human/AS interaction, whereas we focus on the interactions themselves.

We have adopted terms from other authorities, notably Malle and Ullman [8] on trustworthiness and reference alternative terms for the measurement of trust in relation to explainability and transparency from Miller [12].

1.2 Paper Structure

In what follows, we map our progress to the structure of a television drama. Our hero, the project team, formulates a plausible rationale with respect to benchmark usecase scenarios in the "ordinary world" of software development and information science. Encountering "conflict" in the shape of stakeholder engagement and potential disagreement, the team revises their stance attempting to change and grow, and concludes, we hope, with a satisfying "resolution".

2 "ORDINARY WORLD"

Our multi-disciplinary project team includes social scientists, creatives, business advisors and various industry partners. We viewed the taxonomy as a search tool, an aspect of library/database development, so the task of its construction initially fell to those with a background in computer science and IT. Software development was safe ground and we chose to adopt a formal process of taxonomy construction, proposed by Nickerson et al. [13], which is explicitly geared to information systems.

Nickerson et al. set out some ground rules. The end result, they suggest, should be "concise", "robust", "comprehensive", "extendible" and "explanatory" (p.344). They define a taxonomy as a set of dimensions each comprising a mutually exclusive and exhaustive set of characteristics. Formally, $T = \{D_i, i = 1, .., n | D_i = \{C_{ij}, j = 1, ..k_i, k_i \geq 2\}\}$. Crucially, no item described by the taxonomy can have more than one characteristic for any given dimension, and the taxonomy is all on one level; there are no *sub*-dimensions.

The Nickerson et al. process is first to define what is being classified and who the users will be, with a view to establishing the taxonomy's "meta-characteristic". The meta-characteristic then guides selection of the dimensions and characteristics on which classification is to be based. Next, adopting a familiar IT paradigm, the exit criteria are determined. The classification process involves iterating through some subset of the items to be categorised taking either an empirical-to-conceptual or a conceptual-to-empirical approach: the former based on examination of the items themselves, the latter leading with and depending on knowledge and intuition.

Our preliminary decisions are set out in Table 1 and these were critical in determining the nature of the taxonomy we initially produced. In particular, note that the purpose of the library, which the taxonomy exists to support, is to surface "benchmark" scenarios, which we distinguish from standard usecases or experimental scenarios by their potential for *transferability* across multiple domains and/or environments to test multiple variables, as illustrated in Figure 1; and this notion is reflected in the way we conceptualised and described the objects for classification. Furthermore, despite the similarity of agenda shared by ourselves and our primary userbase of researchers, these users were of course not privy to our hopes for what the library might achieve. Mistakenly, however, we assumed that they would share our views. Practically, the meta-characteristic we arrived at was the *nature* of the scenario. That is, to feature in the taxonomy, characteristics must be *intrinsic* to the scenario they describe. To test for inclusion we ask, "If we changed this characteristic, would we significantly change the scenario?"

Taking a random selection of usecases from our preliminary dataset,[1] we applied the Nickerson et al. method to arrive at the following 7-dimension taxonomy.

Taxonomy 1: reflects project goals

domain{dependent, independent}. *Our key criterion for identification of a benchmark usecase is its transferability. Rather than enumerate countless possible domains, we need to determine whether the usecase is domain-dependent or independent.*

interaction{adversarial, advisory, collaborative, observation, recommendation}. *Captures the core nature of the scenario.*

measurement{demonstrated, perceived}. *Terms from [12] distinguish behavioural/physiological measures from those that are self-reported. (We have later expanded and simplified these in line with [5]).*

risk{financial, reputational, physical, privacy, productivity}. *Typically present when a usecase is trust-related* [6]. *(We later add 'hypothetical' as a characteristic since, when dealing with moral/value-based trust (see below), risk may not be present.)*

test environment{dependent, independent}. *As for domain, reflects a fundamental distinction that can determine transferability.*

type{built, broken, instantaneous, repaired}. *Where* ***interaction*** *captures what happens in the scenario,* ***type*** *captures what happens to 'trust' in the scenario.*

trustworthiness{performative, moral}. *Terms from [8] distinguish trust traditionally placed in machines from the type traditionally placed in humans. (We have later renamed the dimension as trust type and the characteristics—for clarity—as "competency-based" and "value-based".)*

Notably, since our interest is in scenarios and, in particular, *benchmark* scenarios, the taxonomy avoids reference to particular domains or test environments. Also, the terms used for measurement and types of trustworthiness are drawn from the literature and are not necessarily the most intuitive/user-friendly choices. Nevertheless, Taxonomy 1 meets Nickerson et al.'s subjective requirements in that it is concise, robust, extendible, explanatory and, arguably, comprehensive relative to the small sample initially considered.

3 "CONFLICT"

We conducted an online workshop attended by 18 participants, 14 joining via Microsoft Teams and 4 via a collaborative online noticeboard [1]. The participants included primary investigators or their nominees from each of six nodes of the UKRI's TAS network (representing six aspects of trust-related research: functionality, governance and regulation, resilience, security, trust and verifiability); representatives from industry and from five "responsibility" projects, similarly funded by the UKRI TAS initiative, also attended.

Table 1: Our Preliminary Decisions

Step	Task	Decision
(a)	Objects to classify	TAS-related experimental usecase scenarios with "benchmark" potential
(b)	Users	HRI and TAS researchers (primary), artists, regulators, policy-makers, general public
1	Meta-characteristic	Nature of scenario
2	Exit criteria	No new dimensions found, no more scenarios

Table 2: Stakeholder Priorities

Theme	Priority
Previous subjects	Sample size, M/F split, participant age, experimental design, dependent and independent variables (DV, IV), DV source
Original study	Test environment, original discipline, questionnaires used, domain
The usecase	Task time critical, vullnerability type, level of vulnerability, star rating, user review
The AS	Wizard-of-Oz, robot platform/system type, level of autonomy
Impact	Influence on policy makers, labs and research groups, adoption by industry and start-ups

Table 3: Projected Stakeholder Responses

Nickerson task	Projected decision
Objects to classify	TAS-related experimental studies
Meta-characteristic	Nature of the study

We analysed stakeholder priorities under five themes. Based on those priorities, we projected their responses to the Nickerson et al. tasks and found they differed markedly from our own. Stakeholders focused on the study, not the usecase, on the results and the domain, not portability.

[1]The preliminary dataset comprised 95 trust-related scenarios: 85 as considered by McKenna et al. in their HRI study of vulnerability as a predictor of trust [11] plus 10 from studies relating to recommender systems, security, and interpersonal trust.

Table 4: Issue Resolution

Project objective	Issue	Dilemma
Adherence to Nickerson et al.	Single level taxonomy	On a technical level, we would prefer a single level autonomy: it is conceptually cleaner and easier to implement. But are there advantages to multiple levels? Like [5], we encountered three broad methods of trust measurement: physiological, behavioural and self-reported. But all of these could be broken down further. Behavioural measurements from the literature include e.g., compliance, change of mind, physical action. These are all related. Meets an alternative suggestion from Nickerson et al. [13] that one test for the quality of a taxonomy is how easy it is for a user to find related usecases. **Resolution:** Incorporate multiple levels and use this to recommend 'similar' usecases, ("you may also be interested in...").
Transferability	Domains	Transferability across domains is a valued quality for a usecase to acquire benchmark status. But our stakeholders are very wedded to the domains within which they work. Setting aside difficulty of encouraging them to upload usecases without reference to the domain where they originated, they will want and expect to be able to *search* by domain. There are good reasons for this. They want to know what usecases other researchers in their domain are using. How do we reconcile this with our project goals? **Resolution:** Collect origin domains but, on search, also display usecases that are 'domain-independent' - alternatives that they may be unaware of because, although applicable, they originated in a different domain (or have been written up without reference to any domain at all). Do this (in library) but maintain integrity of our taxonomy..
Transferability	Test environments	Transferability across test environments. We suggest that a usecase is more useful—has greater benchmark potential—if it can be used in a range of different test environments. Our stakeholders, arriving at the site with a usecase in mind (to add to the library) are likely to think only in terms of the test environment that was actually used. Is it our role to educate them? Why? They're not 'wrong' just at odds with our intention. If we ask the direct question "Is this usecase transferable across test environments?" will they be able to answer it without a long explanation? This could discourage use of the library. **Resolution:** At an interface level, invite users to checkmark "all that apply". If they check more than one, the usecase may be categorised as "transferable". The more environments they select, the more transferable that usecase is with respect to test environment.
Transferability	Results	Results are irrelevant to transferability but important to stakeholders. Moreover, they tell us whether and the extent to which a usecase has been *useful in the field*. The same may be said of many of the search criteria suggested by stakeholders (e.g., originating lab, publication). A taxonomy for 'studies' implies a completely different set of dimensions. How do we accommodate this notion without proliferation of dimensions? Should usefulness be a property of the usecase? **Resolution:** Include the capacity for contributors to comment on the usefulness of scenarios both when they upload and when they read a usecase uploaded by someone else. Similar function to reviews on, say, Amazon, but searchable.
Usability	Adding usecases	Papers that describe studies of potential interest typically feature *results* in the abstract. This is the easiest part of the paper for users to cut and paste but is not what we are interested in. Should we encourage input of a stripped down usecase? Will that be too onerous? Do we let users cut and paste irrelevant info in their submissions and then edit them (likely *most* of them)? Or retain irrelevant uploads unedited and depend on classification and contributor feedback? **Resolution:** Pre-populate the library with 'good' examples. Reinforce with a subtitled Vimeo on the homepage. Monitor usage and guide users with questions that signal expected input as explicitly as necessary, e.g., Please describe the scenario not the results of the study: what does the AS do? what does the human do?
Usability	Categorisation	Our taxonomy may be 'correct' but are we trying to make distinctions between usecases and benchmark usecases which are too rarefied? Can we protect users from having to engage directly with concepts they do not need to grapple with and still provide the benefits identifying the usecases they are most likely to find useful? **Resolution:** Separate how users search the library from the taxonomy itself. Allow it to be extended. Allow free search across all text fields. If users specify a particular domain, show them, in addition, usecases that are 'domain-independent'. Invite input in such a way that database fields (based on characteristics) can be populated 'behind the scenes'.
Usability	Unexpected usage	We are developing the library for users to find usecases but they might have other ideas. Should we stream resource users into e.g., those seeking to amplify lit review and those seeking to identify relevant transferable usecase scenarios? **Resolution:** We cannot know how users will use the library until it is up and running. Build it for its intended purpose; monitor its use; keep it modular and be prepared to extend and modify.

We subsequently circulated a questionnaire[2] to which 16 participants responded anonymously, identifying themselves as ranging from doctoral student to UK Government Advisor. From this fieldwork, we sought to understand research practices amongst potential users of various levels of experience in relation to the sourcing of trust-related usecase material to inform their work. In particular, we probed for search terms and metadata preferences which we anticipated would underpin, and potentially validate, our taxonomy.

Stakeholder priorities, as voiced in the workshop and recorded on the Mural noticeboard, are set out in Table 2. Based on those priorities, we were able to project their likely responses to the preliminary tasks set by Nickerson et al. (see Table 3). These were in stark variance to those of the project team.

The key distinction is that our stakeholders, as represented in this sample, tended to focus on the *study*, not the usecase; on the results, the domain, not portability; the specifics of the implementation and its impact. A secondary issue, hard on the heels of the first, was that, if we were to try to incorporate the requirements of all potential stakeholders, we would face a potentially exponential explosion of features to accommodate within the taxonomy. Moreover, hardly any of our stakeholders' priorities related to the specifics of the interaction or the type of trust that a usecase examined: notions we regarded as intrinsic. Measurement, where mentioned at all, was assumed to be by questionnaire, domains and test environments were explicitly referenced—that is, stakeholders expected to search by the particular domain and test environment at their disposal. When asked what extra information would be helpful when searching the library for a usecase, only one respondent mentioned the notion of transferability ("Independence of a specific system or approach").

Bloodied but unbowed, we faced a dilemma. If our stakeholders expect a library of studies how can we accomplish the necessary switch of their perspective to the identification of transferable benchmark usecases? If we are assembling this library for our stakeholders, how far should we give them what they seem to want? To what extent should we give them what we think they should have? Where is the line of reconciliation and how do we find it in a 'responsible' way?

[2]See: https://kings.onlinesurveys.ac.uk/trust-benchmarks-library.

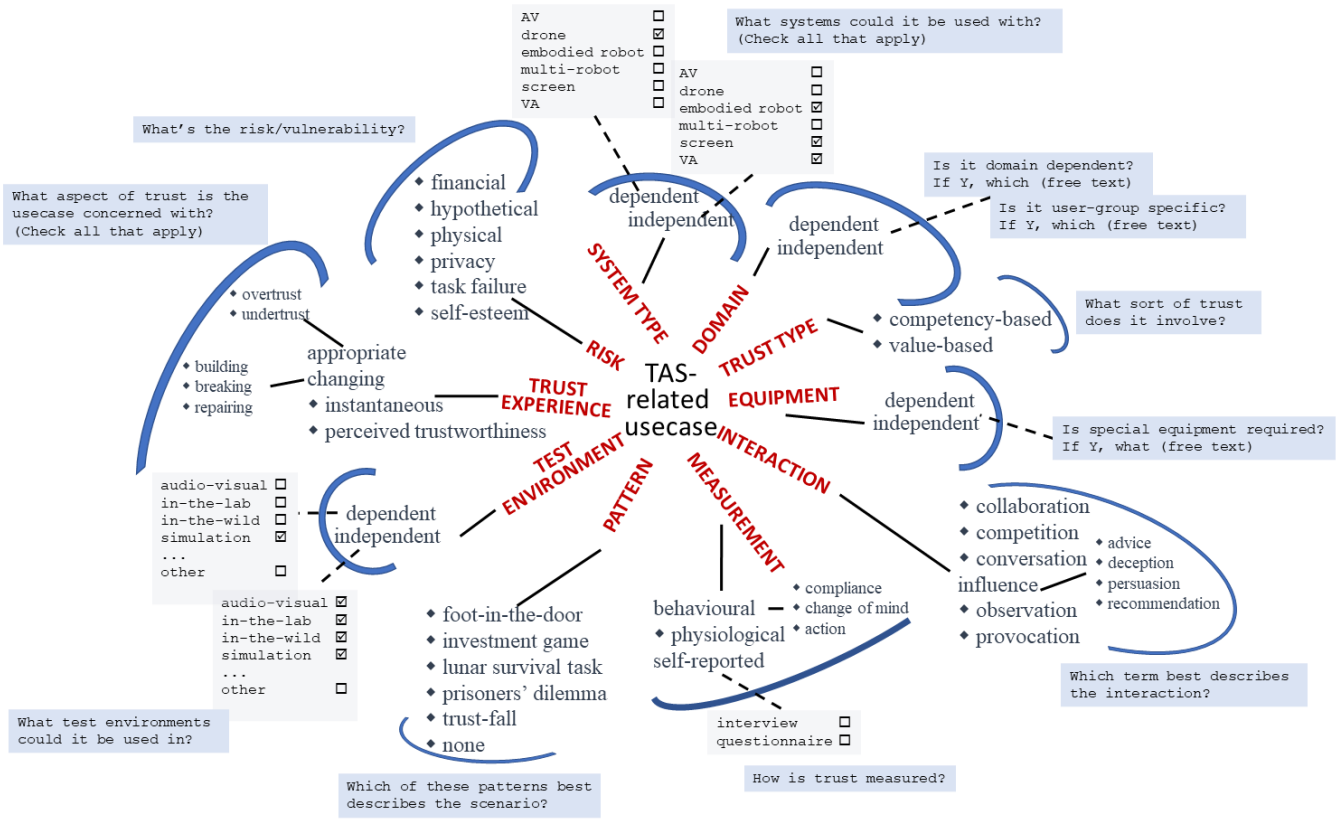

Figure 2: A practical taxonomy.

Library users are insulated from the taxonomy. The red 'wheel' of captions represent the revised taxonomy's 10 dimensions. Characteristics for each dimension are listed inside the curved blue lines, which indicate the threshold between the library interface and the taxonomy itself. Outside those lines, captions illustrate how library users are prompted with straightforward questions. Where necessary, their checkbox responses are converted to terms supported by the taxonomy (e.g., at test environment (dependent, independent)). Diamond bullets indicate taxonomy terms that may be presented directly to the user. The multi-level taxonomy supports recommendations for similar usecases (e.g., at trust experience—changing(building, breaking, repairing). In practice, users will be presented with the 'other' option routinely, enabling them to enter a free text response, to be handled as library maintenance.

4 "RESOLUTION"

From the outset, we adopted the AREA code (to anticipate, reflect, engage and act) recommended as a "key lesson" from RRI Tools [3]. We failed to anticipate, however, the mismatch between our vision for the library and stakeholders' expectation. Interestingly though, it is precisely this unforeseen turn of events that prompted us to *reflect*, re-evaluate stakeholder *engagement* and take *action*.

Concretely, we switched from an exclusively technology-focused approach to a socio-technical one. We conducted a more comprehensive review of the source material, examining 75 HRI-related plus 11 non-robotics trust-related studies, and produced a thematic analysis of the usecases in an iterative process using post-it notes and their online equivalents. In the process of reconceptualising and reorganising the material, we uncovered and endeavoured to resolve a range of issues, as summarised in Table 4.

Improved understanding of our stakeholders' needs, points to three areas requiring action: our preference to be guided by a formal methodology; the need, because we seek benchmarks, to identify transferability as a key characteristic; and the taxonomy's usability. Broadly, we see three overlapping solutions:

(1) **Separation.** Without the library, there would be no need for a taxonomy. Yet, the solution to most of our stakeholder-related dilemmas appears to lie in separating the two, ensuring that the taxonomy serves the library without dictating how it operates and that the library serves the taxonomy, with database fields that represent characteristics invisibly filled *based* on user input, without requiring users to understand the underlying naming conventions or organisation.
(2) **Extendability.** Neither the taxonomy nor the library should be in any way prescriptive. "The customer is always right." As part of its role in servicing the taxonomy, the library interface must facilitate addition of new fields where stakeholders are unable to map their usecase to the ones we provide.
(3) **Flexibility.** A formal model is useful, but not if it becomes a straight jacket. We must be mindful that alternative approaches may better serve our stakeholders' interests. We do not want to sacrifice our primary objectives but we do want to facilitate our users and must remain open to the possibility that their requirements and expectations may continue to differ from our own.

5 DISCUSSION

Understanding the requirements and expectations of a range of stakeholders and developing a strategy to prioritise as many of these as possible, so that all users can not only use the system but are able to find their research interest and research culture represented, is important; to achieve this without sacrificing a project's integrity, equally so. Our focus on transferability de-emphasises the domain of the research (e.g., health, transport, etc.) so that researchers from many different disciplines can use our taxonomy. The early involvement of our stakeholders, however, has shaped the project in ways that, at the outset, we had not anticipated. In terms of RRI, our process has grappled with finding a balance between the incorporation of the stakeholder voice and our own research objectives, prompting our reflection on the legitimacy of requiring users to conform, to some extent, to our objectives. It is a question that we still debate, and will explore further as the usability of the developing online resource is tested. The current symposium, in particular, affords us the opportunity to trial the taxonomy outside the project team prior to the library's implementation.

Whatever the outcome, our goal remains to develop a tool that enables as broad a spectrum of researchers as possible—with their varying methods and concerns—to negotiate a database of TAS-related usecases, to locate usecases of relevance and apply them in their own domain. Arguably, our strategy can lead to a more creative, interdisciplinary approach. Success will see researchers using the scenarios in a way that critically informs their work. At a more abstract level, it sees them begin to appreciate the complexities associated with trust and autonomous systems and to discover how active examination of trust-related issues can cross application domains, test environments and academic disciplines.

ACKNOWLEDGMENTS

We thank our anonymous reviewers for helping to improve this work. Our research is supported by the UKRI Trustworthy Autonomous Systems (TAS) Hub (EP/V00784X/1).

REFERENCES

[1] David Baga. 2011. Mural. https://www.mural.co/ Accessed: March 2023.
[2] European Commission. 2020. Institutional changes towards responsible research and innovation. https://op.europa.eu/en/web/eu-law-and-publications/publication-detail/-/publication/582ef256-cbcc-11ea-adf7-01aa75ed71a1
[3] RRI Tools Consortium et al. 2020. *A practical guide to responsible research and innovation: Key lessons from RRI tools.* Barcelona: Milimétrica Producciones SL.
[4] Kai Klinker, Lisa Berkemeier, Benedikt Zobel, Hanna Wüller, Veronika Huck-Fries, Manuel Wiesche, Hartmut Remmers, Oliver Thomas, and Helmut Krcmar. 2018. Structure for innovations: A use case taxonomy for smart glasses in service processes. *Multikonferenz Wirtschaftsinformatik Lüneburg, Deutschland* (2018).
[5] Spencer C Kohn, Ewart J de Visser, Eva Wiese, Yi-Ching Lee, and Tyler H Shaw. 2021. Measurement of trust in automation: A narrative review and reference guide. *Frontiers in psychology* 12 (2021), 604977.
[6] John D. Lee and Katrina A. See. 2004. Trust in Automation: Designing for Appropriate Reliance. *Human Factors* 46, 1 (2004), 50–80.
[7] Jamy Li, Wendy Ju, and Cliff Nass. 2015. Observer perception of dominance and mirroring behavior in human-robot relationships. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*. 133–140.
[8] Bertram F Malle and Daniel Ullman. 2021. A multidimensional conception and measure of human-robot trust. In *Trust in human-robot interaction*. Elsevier, 3–25.
[9] Nikolas Martelaro, Victoria C Nneji, Wendy Ju, and Pamela Hinds. 2016. Tell me more designing hri to encourage more trust, disclosure, and companionship. In *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 181–188.
[10] Peta Masters, Sachini Weerawardhana, and Victoria Young. 2023. Towards an Open Source Library and Taxonomy of Benchmark Usecase Scenarios for Trust-Related HRI Research. In *International Conference on Human-Robot Interaction (HRI) Workshop on Advancing HRI Research and Benchmarking Through Open-Source Ecosystems, March 13, 2023, Stockholm, Sweden*. 1–4.
[11] Peter E. McKenna, Muneeb I. Ahmad, Tafadzwa Maisva, Birthe Nesset, Katrin Lohan, and Helen. Hastie. 2023. A Meta-analysis of Vulnerability and Trust in Human-Robot Interaction. *submitted* (2023).
[12] Tim Miller. 2022. Are we measuring trust correctly in explainability, interpretability, and transparency research? *arXiv preprint arXiv:2209.00651* (2022).
[13] Robert C Nickerson, Upkar Varshney, and Jan Muntermann. 2013. A method for taxonomy development and its application in information systems. *European Journal of Information Systems* 22, 3 (2013), 336–359.
[14] Suzanne Tolmeijer, Astrid Weiss, Marc Hanheide, Felix Lindner, Thomas M Powers, Clare Dixon, and Myrthe L Tielman. 2020. Taxonomy of trust-relevant failures and mitigation strategies. In *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*. 3–12.
[15] Kun Xu. 2019. First encounter with robot Alpha: How individual differences interact with vocal and kinetic cues in users' social responses. *new media & society* 21, 11-12 (2019), 2522–2547.