Developing and Validating the Perceived System Curiosity Scale (PSC): Measuring Users' Perceived Curiosity of Systems

Jan Leusmann
LMU Munich
Munich, Germany
jan.leusmann@ifi.lmu.de

Steeven Villa
LMU Munich
Munich, Germany
villa@posthci.com

Burak Berberoglu
LMU Munich
Munich, Germany
burakberberoglu1905@gmail.com

Chao Wang
Honda Research Institute Europe
Offenbach/Main, Germany
chao.wang@honda-ri.de

Sven Mayer
LMU Munich
Munich, Germany
info@sven-mayer.com

Abstract

Like humans, today's systems, such as robots and voice assistants, can express curiosity to learn and engage with their surroundings. While curiosity is a well-established human trait that enhances social connections and drives learning, no existing scales assess the perceived curiosity of systems. Thus, we introduce the Perceived System Curiosity (PSC) scale to determine how users perceive curious systems. We followed a standardized process of developing and validating scales, resulting in a validated 12-item scale with 3 individual sub-scales measuring explorative, investigative, and social dimensions of system curiosity. In total, we generated 831 items based on literature and recruited 414 participants for item selection and 320 additional participants for scale validation. Our results show that the PSC scale has inter-item reliability and convergent and construct validity. Thus, this scale provides an instrument to explore how perceived curiosity influences interactions with technical systems systematically.

CCS Concepts

• **Human-centered computing → HCI design and evaluation methods**.

Keywords

Human Computer Interaction, Scale, Questionnaire, Perceived Curiosity

ACM Reference Format:
Jan Leusmann, Steeven Villa, Burak Berberoglu, Chao Wang, and Sven Mayer. 2025. Developing and Validating the Perceived System Curiosity Scale (PSC): Measuring Users' Perceived Curiosity of Systems. In *CHI Conference on Human Factors in Computing Systems (CHI '25), April 26–May 01, 2025, Yokohama, Japan.* ACM, New York, NY, USA, 18 pages. https://doi.org/10.1145/3706598.3713087

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
CHI '25, Yokohama, Japan
© 2025 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 979-8-4007-1394-1/25/04
https://doi.org/10.1145/3706598.3713087

1 Introduction

With the recent advances in human-robot interaction (HRI) and artificial intelligence (AI) [21, 101, 119], it is inevitable that robots and other non-embodied systems will become even stronger daily companions in our lives than before, for which it is crucial to enable better interaction design with these systems [94]. Previous research shows that human-like characteristics or higher levels of anthropomorphism increase trust [43, 77]. Curiosity is a human-like trait that makes us want to learn more about others or the environment, allows us to better connect with others, especially through social curiosity [126], and allows for more natural interaction. Additionally, curiosity makes us learn more about our surroundings and inherently gives us the capability to learn and grow [65, 75, 102]. Over time, systems will learn more and may even inquire about knowledge gaps to be better companions, e.g., Doering et al. [34], Wu et al. [160]. Thus, we envision that curiosity will offer active learning systems the possibility to engage the user. However, we have no standardized tool to measure how users perceive the curiosity systems express toward the users.

Curiosity has been used in technical domains as a driver for exploration, e.g., in Reinforcement Learning [24]. Walker et al. [152] studied the effects of this non-goal-focused exploration and found that, indeed, users perceived such behavior as curious [34]. On the other hand, Ceha et al. [29] studied explicit verbal expressions from a social robot and found that participants could identify this trait. Finally, Doering et al. [34] developed a system to limit curiosity-based exploration to always fulfill its task but still found the curiosity beneficial. While we see a number of HRI systems employing curiosity, how they measure curiosity is not standardized and, as such, not comparable. They mainly used self-created questions or rephrased questionnaires measuring humans' own curiosity. While there is a large corpus of scales measuring humans' own curiosity [32, 71, 91], there are currently no scales to measure the perceived curiosity of systems. As a result, there is currently no way to study which system behaviors affect this perceived curiosity systematically and reliably.

In this work, we created a scale measuring perceived system curiosity following the approach to develop validated scales by Boateng et al. [16], see Figure 1. First, we followed a four-step approach to identify potential items on our scale: a) an open literature review, b) a systematic literature review on curiosity questionnaires, c) a systematic literature review on curious robotic systems, and d)

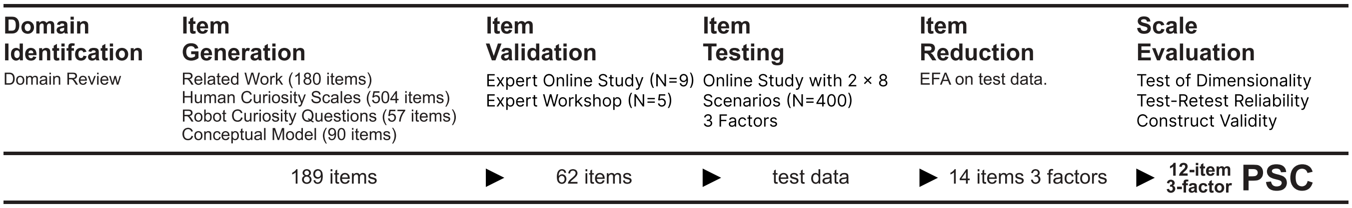

Figure 1: *PSC* **creation based on scale development process by Boateng et al. [16].**

generating items from prior conceptual models. This step returned 831 portal items, which we coded and merged based on their meaning, resulting in 189 items. Next, we asked experts (N=9) to rate the relevance of these items and discussed them with five additional experts in a workshop. This process resulted in 62 remaining items. We then performed an exploratory factor analysis with 400 participants and 16 scenarios ensuing generalizable items, which resulted in 12 remaining items. Afterward, we performed a confirmatory factor analysis with 320 participants. Finally, we validated the scale on internal consistency, reliability, and construct validity. This led to our 12-item, 3-factor scale named Perceived System Curiosity (PSC) scale, see Table 10.

We found that our three factors measuring perceived explorative curiosity, perceived investigated curiosity, and perceived social curiosity effectively measure the construct of perceived system curiosity with high consistency and reliability. We envision that this scale propels research on curious systems, leading to a more effective and natural way for systems to learn from users while still ensuring a nice interaction between the system and the user.

2 Related Work

In the following, we evaluate prior work on human curiosity and curiosity in systems.

2.1 Assessing Curiosity in Humans

Curiosity is often times compared but also distinguished from situational interest [35, 141, 142]. Berlyne [11] gives one of the first models to define curiosity, stating that curiosity is the balance between the two uncomfortable states of under-stimulation and over-stimulation, and proposed diversive curiosity being the search for anything that gives arousal and specific curiosity for when a stimulated person tries to gain understanding about something in order to reduce arousal. Loewenstein [92] later explains curiosity through the information gap theory, stating that people become curious when they realize that they lack desired knowledge. Human curiosity can also be classified into four categories. (1) Epistemic curiosity is the desire for knowledge [10, 87], (2) perceptual curiosity is the interest in novel perceptual stimuli [10, 32], (3) sensory curiosity is the search for novel and unusual sensory experiences [88], and (4) social curiosity is the interest in people [126]. Litman [84] describes two different types of curiosity as either the "need to know" or having fun while learning. However, these theories do not explain how humans express curiosity, but only how they become curious. Shin and Kim [141] propose that while curiosity and interest are connected, interest is related to more pleasurable learning, while curiosity is more of an innate drive. Kashdan et al. [71] propose the five-dimensional curiosity model based on previous models of curiosity with the five dimensions: deprivation sensitivity, joyous exploration, social curiosity, stress tolerance, and thrill-seeking and also developed a questionnaire measuring curiosity on these five dimensions.

Based on these theories, also many scales and questionnaires have been created to measure these different dimensions of curiosity [32, 66, 91, 126]. It has been shown that being curious positively benefits many areas, as curious people are perceived more positively [68], have advantages in social relationships and well-being [69, 114], are more accurate in judging personal traits of others [52], and are happier, open-minded, and empathic [157]. However, while there are many theories and models of what curiosity is and how to assess curiosity for oneself, to our knowledge, there is not a single scale measuring how curious we perceive others. As curiosity is described as something very innate, often coming from intrinsic motivation, these internal thoughts are also difficult to observe and, thus, also predict how curious another person currently is. However, there are also many expressions directly linked to curiosity, such as asking questions [93, 163], expressing explanation-seeking cues [82], non-verbal expressions, and verbal acknowledgments [51]. Perceptual curiosity has also been shown to activate measurable brain regions [58]. We also know that animals express curiosity by exploring novel objects [26, 48], spontaneous movement [81], or investigate behavior [44]. However, to our best knowledge, there are no validated questionnaires measuring the perceived curiosity of another human or entity.

2.2 Curious Systems

Prior work has given technical systems various humanlike traits [18], such as empathy [42, 112, 138], emotions [1, 19], creativity [47, 120], or humor [106]. Increasing the level of anthropomorphism has been shown to increase trust [43], acceptance [13], and general perception [39]. Curiosity could furthermore enhance transparency, as through the curious actions taken by the system, the user becomes informed on the system's goal [38].

Curiosity in this field can be distinguished between functional curiosity and observable curiosity. Functional curiosity has mainly been used to improve learning capabilities of systems [23, 107, 137] and enables open-ended learning. This allows systems to learn by observing and acquiring knowledge, similar to how children learn. This approach leads to rapid, lasting knowledge acquisition [102]. The motivation of curiosity as an intrinsic motivation to explore the unknown and fill knowledge gaps here is often used as a loss function, which motivates the system to take unknown actions to learn new things. On the other hand, in collaborative human-agent settings, it is crucial that the system expresses its curious behavior for the user to notice it. This can be described by the term observable curiosity. While functional curiosity involves actual learning [14, 104], observable curiosity describes how systems communicate their desire to learn. This is either implicitly by making visible actions to explore or explicitly by making internal processes observable and understandable to users [76], although the system might not need to perform an observable action to gather the information, which has been shown to improve user understanding [76, 149]. Active information gathering also improves the estimation of the human state over passive information gathering [133].

We envision systems becoming even more integrated into our lives, especially with advances in robotics. These systems will not only provide information but also offer physical support. However, in a world with limitless knowledge and skills, creating adaptive, personalized systems poses a challenge, as it is impossible to preprogram them with everything, making it indispensable for the system to still learn when deployed [36]. Here, we see two challenges; the development of systems that are capable of learning and ensuring that their curiosity rather enhances than disrupts interaction with humans. To build better systems, we must first identify behaviors perceived as curious and evaluate which ones are well-received.

However, it is still unclear how users perceive such curious behavior and how it can be optimally integrated. Although some effects of curious system behaviors have been studied [29, 152], the lack of a validated scale to measure user perceptions makes these findings difficult to compare. Researchers have used rephrased human curiosity questionnaires [143], related surveys [152], or custom metrics [34, 45], but consistent validation remains a challenge. Currently, it is not clear how a curious system affects user perception in detail. The question of "how curious a system should be" could previously not be solved, especially as there is a clear trade-off between task-oriented and curious behavior. Thus, in this work, we developed a scale measuring perceived curiosity.

3 Item Generation

Figure 1 shows the process we followed while developing the PSC scale, following Boateng et al. [16]. For item generation, we reviewed four sources of prior work. Then, we coded the items by similarity and removed duplicates. Finally, we conducted a study to determine the importance of the questions in evaluating the curiosity of systems.

3.1 Item Collection

First, we reviewed general related work to understand the field (Step A) and inform the next steps; here, we mainly looked at the curious HRI system and their motivations. With the insights from the first step, we performed two systematic searches to find general questionnaires to measure curiosity (Step B) and prior studies measuring the curiosity of systems (Step C). Finally, we reviewed conceptual models of curiosity (Step D).

Step A: From Related Work. We reviewed works using Google Scholar to find commonly used terms in the curiosity literature. Here, we did not specify a focus on questionnaires; however, we still found questionnaires like Loewenstein [92]. Our learning from this step was to split up the search into three parallel steps: curiosity questionnaires (Step B), curious systems (Step C), and looking into conceptual models (Step D).

Additionally, we used seven additional works to inspire the item creation that are not used in the later steps [10, 53, 70, 75, 92, 96, 111, 164]. This includes two questionnaires. Here, Harty and Beall [53] contributed 30 questions on science curiosity, and Pearson [111] with 90 questions on novelty seeking. We also used the remaining six papers for item generation (N=60), focusing on concepts like motivational force, network building, and epistemic and perceptual curiosity.

Step B: From WoS Questionnaire Search. We performed a systematic keyword search on the Web of Science to find questionnaires concerning curiosity. Based on Step A, we formulated the following search string:

```
TI=((measur* OR questionnaire OR scale) AND curiosity)
                          OR
AB=((measur* OR questionnaire OR scale) AND curiosity)
```

Thus, we limited the search to papers that clearly state that they fit out the target already in the title or abstract. This search led us to 2.849 papers. Next, we then went through all papers with the following exclusion criteria: (1) Not English, (2) Not Curiosity, (3) Not a Questionnaire, and (4) Missing Questionnaire. We found that 165 articles were not written in Englisch (1), 2.389 articles were not about curiosity (2), 264 articles about curiosity did not focus on measuring it and, thus, did not provide a questionnaire (3), and two questionnaires did not report their questionnaire (4), see Supplemental Material for the full list. Therefore, we found 29 articles that presented a novel questionnaire, see Table 1.

In the next step, we extracted all questions from the papers. We also extracted questions even if they were not in the final questionnaire to keep the diversity of our question pool high. With this step, we added 504 items from 29 questionnaires to our item pool.

Step C: From WoS Robot Papers Search. We also look into systems that can express curiosity. In detail, we investigated robotic systems that express curiosity. For this, we run a second Web of Science keyword search with the following search string:

```
(TI=("robot*" OR "cobot*") OR AB=("robot*" OR
"cobot*")) AND (TI=("curios*") OR AB=("curios*"))
```

This search resulted in 303 entries. We excluded six papers as they are not written in English, 214 papers that do not investigate curiosity, 76 papers that did not ask questions concerning curiosity or used an existing questionnaire, and one that was retracted and thus excluded. Therefore, we identified five papers that included 57 novel questions, adding to our questionnaire pool, see Table 1.

Step D: From Conceptual Models. Our final item generation search targets finding conceptual models of curiosity. While they typically do not directly result in usable questions, they allow us to understand the concepts and formulate targeted questions to address the underlying concepts. With the insights from the prior steps, we scoped the references and Google Scholar for conceptual models revolving around curiosity. In total, we identified 16

Table 1: The 29 questionnaires from our Web of Science keyword search on human curiosity and 5 questionnaires from our Web of Science keyword search on robot curiosity all present questions to measure different aspects of curiosity.

Year	Questionnaire	Reference	Items
	Curiosity Questionnaires for Humans		
1978	Academic Curiosity (AC) Scale	Vidler and Hansen [150]	8
2003	Epistemic Curiosity (EC) Scale	Litman and Spielberger [91]	7
2004	Curiosity As a Feeling of Deprivation (CFD) Scale	Litman and Jimerson [89]	27
2004	Perceptual Curiosity (PC) Scale	Collins et al. [32]	33
2004	Curiosity and Exploration Inventory (CEI)	Kashdan et al. [67]	7
2005	Sensory Curiosity (SC) Scale	Litman et al. [88]	14
2006	Social Curiosity Scale (SCS)	Renner [126]	14
2008	Interest and Deprivation in Epistemic Curiosity	Litman [86]	10
2009	Curiosity and Exploration Inventory-II (CEI-II)	Kashdan et al. [65]	10
2010	Sport Fan Exploratory Curiosity Scale (SFECS)	Park et al. [109]	18
2014	Sport Fan Specific Curiosity Scale (SFSCS)	Park et al. [108]	19
2014	Interest/Deprivation-Young Children (I/D-YC) Scale	Piotrowski et al. [115]	10
2015	Self-Curiosity Attitude-Interest (SCAI) Scale	Aschieri and Durosini [4]	18
2016	Gender-Free Curiosity Inventory	Byman [25]	17
2016	Science Curiosity in Learning Environments Scale	Weible and Zimmerman [154]	30
2017	Science Curiosity Scale	Kahan et al. [61]	12
2017	Elements of Nature Curiosity Scale (ENCS)	Próchniak [121]	10
2018	Five-Dimensional Curiosity Scale (5DC)	Kashdan et al. [71]	25
2018	Adolescent Smoking Curiosity Scale (ASCOS)	Khalil et al. [73]	7
2019	Children's Images of and Attitudes towards Curiosity (CIAC)	Post and Walma van der Molen [118]	24
2020	Five-Dimensional Curiosity Scale Revised (5DCR)	Kashdan et al. [64]	24
2020	Multidimensional Workplace Curiosity Scale	Kashdan et al. [66]	16
2021	Curiosity Towards STEM Education Questionnaire	Ahmad and Siew [2]	10
2021	Self-Curiosity Attitude-Interest Scale	Aschieri et al. [5]	7
2021	Morbid Curiosity Scale	Scrivner [139]	24
2022	Early Multidimensional Curiosity Scale	Lee et al. [78]	30
2023	Medical Curiosity Scale	Bugaj et al. [22]	10
2023	Curiosity of Climate Changes Scale (CCCS)	Próchniak and Ossowski [122]	15
2024	Intellectual Curiosity Scale (ICS)	Bluemke et al. [15]	6
	Curiosity Questionnaires for Robots		
2015	Social Robots and Science Curiosity	Shiomi et al. [143]	8
2019	Curiosity Based Learning Systems and Interactive Behavior	Doering et al. [34]	3
2019	Expression of Curiosity in Social Robots	Ceha et al. [29]	14
2020	Human Perceptions of a Curious Robot	Walker et al. [152]	12
2022	Impact of Task Order and Motive on Perceptions of a Robot	Carter et al. [27]	20

works Aristotle [3], Berlyne [10, 11], Cicero and Tullius, Markus [31], Freud [40], Harlow et al. [50], Hull [55], James [57], Kashdan et al. [71], Litman [83], Litman and Silvia [90], Litman and Spielberger [91], Loewenstein [92], McNary [99], Pavlov [110], Plutarch [116]. Based on these prior conceptual models, we created 90 items to reflect their perspective of curiosity in our context, e.g., PSC_1 "The system is driven towards novel sensations." from Berlyne [10], or Q_{027} "The system is motivated to understand the world." from Freud [40].

3.2 Item Coding

The item collection resulted in 831 items; see Supplemental Material for the full list. For coding, three authors made the following steps in hour-long sessions by discussing all the steps until unanimous consent was reached.

In the first step, we rephrased the perspective to one in which the user can evaluate the system. For example, we reworded "I see a challenge as a way to grow and learn" to "The system sees a challenge as a way to grow and learn." We then removed all 62 literal duplicates. Moreover, we removed three items[1] as they asked the user to evaluate their own interest in the system and not directly how they perceive the system. This left us with 767 items for the main coding.

During the coding, twelve core concepts emerged, which represent the questions; see Table 2. We iteratively discussed the concepts and adjusted their concepts during the process. We then went through these items in iterations to merge related items whenever the overarching concepts of two questions are the same. In total,

[1]Removed Items: "I enjoyed talking to the robot," "If I had the chance, I'd talk with the robot again," and "How interesting was your interaction with the robot?"

Table 2: The twelve code concepts of our coding process and the distribution of the 767 items across them. # I. = Number of items, # S. = Number of selected items.

Code	Description	# I.	# S.
Curiosity	Items asking directly about *curiosity*	18	2
Enjoyment	*Enjoying* things	145	20
Exploration	*Exploring* things and the environment	90	9
Expression	Observable *expressions* from the robot	34	15
Interaction	Verbal and physical *interaction*	59	13
Internal State	*Internal* processes	89	24
Learning	*Learning* new things and information	90	24
Problem Solving	*Solving complex problems*	63	14
Property	*Properties* of the robot	27	26
Reasoning	*Reasoning* processes from the robot	42	16
Reflective	*Reflections* from the robot	44	10
Social	*Social* interaction with at least 1 person	65	15

we merged 578 items into 189 final items representing all initial concepts of the initial item set.

4 Item Validation

Following the process of Boateng et al. [16], we validated that the items are indeed useful for evaluating systems. In our case, this is especially important as items may originate from a questionnaire designed for humans. Here, we followed a two-step approach. First, we asked nine experts to rate each item to calculate the Content Validity Index (CVI). Second, we conducted an expert panel with five additional experts to refine the remaining relevant items.

4.1 Expert Ratings: Online Study

We conducted an online expert study to calculate the CVI. We asked nine participants with expertise in HCI, HRI, or Psychology to rate all 189 items on a 4-point scale regarding their appropriateness to the topic.

Procedure. First, we informed participants that the study's goal is to develop a scale that measures if users perceive a system as curious. Next, participants gave their informed consent, filled in a demographic survey, and rated expertise in eight domains. We then asked the participants to rate each of the 189 items on a 4-point scale: "not relevant," "somewhat relevant," "quite relevant," and "very relevant," following the guidelines by Polit and Beck [117] and Yusoff [161]. We showed the participants each item individually in randomized order. For every item, participants also had the option to add comments via a free text field.

Participants. We recruited nine participants (2 female, 7 male) with strong expertise in HCI, HRI, or Psychology via convenience sampling. Participation was voluntary, and we made participants aware that they could stop the survey at any point. The average age of the participants was 32.89 years ($SD = 5.18$). Four participants had a doctoral degree, and five had a master's degree. We asked participants to self-assess their expertise in eight domains, c.f. Table 3: Human-Computer Interaction ($M = 4.44, SD = 0.73$), Human-Robot Interaction ($M = 3.33, SD = 1.58$), Psychology ($M = 2.56, SD = 1.24$), UX design ($M = 3.78, SD = 0.97$), Robotics ($M = 2.44, SD = 1.59$), system design ($M = 3.56, SD = 1.01$), human behavior theories ($M = 3.11, SD = 1.05$), and questionnaire development ($M = 3.11, SD = 1.27$). Furthermore, we asked participants to self-assess their curiosity ($M = 4.78, SD = 0.44$). Participants took, on average, 66.48 minutes ($SD = 44.03$) to complete the survey.

4.2 Expert Ratings: Results

Boateng et al. [16] suggest calculating the CVI per item to measure proportional agreement, resulting in an I-CVI [161]. Our 189 items had an S-CVI of 0.50 ($SD = 0.33, min = 0.00, max = 1.00$). We argue that our low S-CVI is due to the fact that the majority of our items are rephrased questions from human self-assessment curiosity questionnaires. Polit and Beck [117] suggest using a minimum I-CVI of 0.78 for six to ten experts. With nine participants, we are close to the upper bound of the suggested amount of six to ten experts [117]; thus, the diversity of opinions is also at the upper bound. Consequently, we chose to include all items with an I-CVI > 0.6 and move the discussion about the relevance of these items to the following expert workshop. This threshold resulted in 82 kept items. We removed all items below that threshold. The remaining set consisted of 15 items with an I-CVI of 1.0, 23 items with an I-CVI of 0.88, 26 items with an I-CVI of 0.77, and 18 items with an I-CVI of 0.66.

4.3 Expert Panel: Item Validation

We conducted an expert workshop with five participants to discuss the validity of the remaining 82 items. During the workshop, participants could reword, keep, remove, and add items. The goal of this process was to ensure consistent wording and understandability, remove ambiguous statements, and ensure that all remaining questions are in the domain of perceived curiosity.

Procedure. Two authors invited five participants to an in-person group discussion. This resulted in a total of seven researchers openly discussing the items. Participants first received a short introduction to the study, signed an informed consent form, filled in a demographic survey, and rated their own expertise in eight domains. We then handed out the 82 questions on printed paper. We divided the group discussion into two parts. First, we asked all participants to go through each item individually to get an overview and note down their thoughts on all items. Secondly, we started the open discussion round. Here, the two study conductors would guide the participants through each item one by one. We ensured that for every item, all participants could state their thoughts first and then vote on suggestions and comments in the group. In cases, where opinions about an item were indecisive, we would start a discussion about this item, and either reword the item in a way that all participants were satisfied with the outcome, remove the item in case we could not reach an agreement, or add a new item in cases were the group found that one item contributed to multiple parts of the construct. We also included comments on the individual items we received in the online study.

Participants. We recruited five new participants (1 female, 4 male) with strong expertise in HCI, HRI, or Psychology via convenience sampling. Participation was voluntary, and we made them aware that they could stop the survey at any point. Participants were 33.20

Table 3: Overview on demographics, background, and expertise of the nine participants who rated the relevance of 189 items in the expert panel for item validation. Includes the highest degree, self-assessed expertise in psychology (Psy.), Human-Robot Interaction (HRI), Human-Computer Interaction (HCI), their primary research domain, their self-assessed expertise in UX, robotics (rob.), system design (Sys.), behavioral theories (Beh.), questionnaire development (Dev.), their age, and gender.

Degree	Psy.	HRI	HCI	Domain	UX	Rob.	Sys.	Beh.	Dev.	Age	Gender
Master's	■■■□□	■■■□□	■■■■■	HCI	■■■■□	■■■□□	■■■■□	■■■□□	■■■■■	27	M
Doctoral	■■■□□	■■■□□	■■■■■	(Meta) HCI, AI	■■■■□	■□□□□	■■■■■	■■■■□	■■■■□	33	M
Doctoral	■■■□□	■■■■■	■■■■■	Gen. AI, XR	■■■■□	■■■■□	■■■■□	■■■□□	■■■□□	33	M
Master's	■□□□□	■□□□□	■■■■□	HCI	■■■□□	■□□□□	■■■□□	■□□□□	■□□□□	29	M
Master's	■□□□□	■□□□□	■■■■□	HCI	■■□□□	■□□□□	■■■□□	■■□□□	■■□□□	31	M
Doctoral	■■■■□	■■■□□	■■■■■	HCI	■■■■■	■□□□□	■■■□□	■■■■□	■■■■□	36	F
Doctoral	■■■■□	■■■■■	■■■□□	HRI	■■■□□	■■■■□	■■■□□	■■■■□	■■□□□	44	F
Master's	■□□□□	■■■■□	■■■■□	HCI	■■■■□	■■□□□	■■□□□	■■■□□	■■■□□	28	M
Master's	■■■□□	■■■■■	■■■■■	HRI	■■■■■	■■■■■	■■■■■	■■■■□	■■■■□	35	M

Table 4: Overview on demographics, background, and expertise of the five participants who participated in the expert panel for item validation. Includes the highest degree, self-assessed expertise in psychology (Psy.), Human-Robot Interaction (HRI), Human-Computer Interaction (HCI), their primary research domain, their self-assessed expertise in UX, robotics (rob.), system design (Sys.), behavioral theories (Beh.), questionnaire development (Dev.), their age, and gender.

Degree	Psy.	HRI	HCI	Domain	UX	Rob.	Sys.	Beh.	Dev.	Age	Gender
Master's	■■□□□	■■□□□	■■■■□	HCI, Privacy	■■■■□	■■□□□	■■■■□	■■■■□	■□□□□	28	F
Master's	■■□□□	■■□□□	■■■■■	Tech	■■■■□	■■□□□	■■■■□	■■■■□	■■□□□	32	M
Doctoral	■■□□□	■■■□□	■■■■■	HCI, XR, AI	■■■□□	■■□□□	■■■■□	■■■□□	■■□□□	37	M
Doctoral	■■■□□	■■■■■	■■■■■	HCI,AI	■■■■■	■■□□□	■■■■■	■■■□□	■■■□□	33	M
Master's	■■□□□	■■■■■	■■■■■	HRI	■■■■□	■■■■□	■■■■□	■■■□□	■■■■□	36	M

year old on average ($SD = 3.56$). Two participants had a doctoral degree, and three had a master's degree. We asked them to self-assess their expertise in eight domains, see Table 4. Furthermore, we asked participants to self-assess their curiosity ($M = 4.00, SD = 1.73$). The duration of the workshop was ~ 2 hours.

4.4 Expert Panel: Results

Reducing the items during the focus group resulted in 64 items. In the expert workshop, we kept 20 items, reworded 36 items, merged 12 items, removed 14 items, and added seven items.

We reworded 36 items where the experts agreed that the current version of the question was either too general, too narrow, had too difficult or technical language, could only be applied by physical systems, and streamlined language across all questions. Furthermore, we rephrased "people" to "others" when applicable, as systems might interact both with humans and other systems. For example, we rephrased item U276 "The system is open to trying new solutions when it is not able to solve a difficult problem." to "The system is open to trying new solutions." and item Q_{011} "The system actively seeks as much information as it can in new situations." to "The system actively seeks as much information as it can." to make the questions applicable to more scenarios. We rephrased some items to make them easier to understand. For example, we changed item PSC_4, "The system prefers tasks that are excitingly unpredictable." to "The system is excited by unpredictable tasks." and item PSC_2, "The system is driven toward seeking out novel stimuli." to "The system is driven towards novel sensations.". We rephrased items that involved actual physical interaction so they could be applied to any system, e.g., item Q_{039}: "The system often points, gestures, or talks about things that are new to it." to "The system often indicates that something is new to it."

All experts agreed to merge 12 items were another question was asking about the same concept. For example, we merged item Q_{072} "When the system encounters a new person, it is interested in learning more about them." into a rephrased item Q_{016} "The system is interested in learning more about new people." or item Q_{113} "The system has a strong drive to learn new things." into item Q_{041}: "The system likes learning new things.".

We removed items when the concept was not observable, e.g.,item Q_{063} "The system tries to figure out what might have caused it when something unexpected happens.", Q_{068} "The system finds it interesting to think about contradicting ideas.", or Q_{122} "The system has a strong drive to learn new things when there are positive rewards.".

Lastly, we added seven items based on discussions on other items and emerging concepts. For example, we added item Q050 "The system is nosy." based on discussions that a system could also be too curious when talking about item Q_{046} "The system is investigative." or item PSC_{11} "The system likes listening to others' conversations.". Furthermore, we added item Q_{047}, "The system invests substantial effort to collect new knowledge." based on item PSC_6, "The system invests substantial time to collect new knowledge." where we discussed that effort is clearly observable, while time might only sometimes be observable. All questions and changes can be found in the Supplemental Material.

Table 5: The Scenarios used for the item testing and item validation.

Context	No.	System	Audio	Ref.
Smart TV	1	**Curious:** Voice assistant asks about user preferences to help with personalized setup and makes suggestions based on user behavior.	02:23	[41, 74, 162]
	2	**Non-Curious:** Voice assistant guides the user through setup and provides general recommendations upon request.	01:20	
Cleaning Robot	3	**Curious:** The Cleaning robot asks about user preferences and explores home details itself, suggests decor ideas, and adapts to the user's schedule.	02:02	[63, 156]
	4	**Non-Curious:** Cleaning robot with default setup, maps users' home on request and follows a set cleaning routine based on user input.	01:14	
Assembly Robot	5	**Curious:** Humanoid robot assists with chair assembly by mimicking users' actions, asking questions to understand and improve, and learning and exploring.	02:19	[95, 103]
	6	**Non-Curious:** Humanoid robot assists with chair assembly by following direct user instructions and completing tasks users requests.	01:10	
Social Robot	7	**Curious:** Social robot assists with personalized meal suggestions, adapts tasks based on user preferences, and searches for additional organizational tips.	02:24	[72, 113]
	8	**Non-Curious:** A social robot completes tasks based on standard routines and follows specific instructions without further inquiry or adaptation.	01:11	
Smart Home	9	**Curious:** Smart home system engages in personalized conversations, learns user preferences, and provides proactive and tailored suggestions for comfort.	02:32	[80, 105]
	10	**Non-Curious:** Smart home system responds to user commands, adjusts temperature and alarms as instructed, and provides assistance when prompted.	01:21	
Design App	11	**Curious:** AI design tool actively engages in dialogue, asks for user preferences and reasoning behind choices, and explores creative options with the user.	02:43	[56, 153]
	12	**Non-Curious:** AI design tool responds to your commands, provides options for furniture and decor, and guides the user through the design process.	01:23	
Well-being App	13	**Curious:** Wellbeing app engages through questions about users stressors and coping habits, showing interest in preferences and offers personalized feedback.	02:49	[151]
	14	**Non-Curious:** Wellbeing app has a preset diary format. Users fill in sections about stressors and coping habits prompted with predefined topics.	00:51	
Ticket Machine	15	**Curious:** Robot ticket machine asks about travel purposes, offers route options and suggests personalized recommendations based on conversation.	02:07	[140, 148]
	16	**Non-Curious:** Standard ticket machine provides options for destination, train type, and payment.	00:51	

5 Item Testing

The next step in our scale development process, suggested by Boateng et al. [16], is to test the validated item set and collect data for the item reduction in a sample size of at least 200 to 300 participants. For this, we conducted an online survey with 16 audio scenarios, c.f., Table 5. We recruited 400 participants via Prolific, who each rated one scenario with our item set. For each scenario domain, we created one version in which the system acts curious and another with a non-curious system.

5.1 Procedure

After a short introduction, we asked the participants for their informed consent. After giving their consent, participants had to answer demographic questions. Next, we assigned each participant one scenario, balancing the number of responses we got per scenario. Participants could listen to the described scenario as many times as they wanted. We ensured that they listened to it at least once by hiding the continue button until the scenario had been played once fully. We then asked participants to summarize the scenario shortly to ensure they correctly understood it. We then asked participants to rate their perceived curiosity about the system using our 64 items. We displayed the questions in randomized order. Each question could be answered by a 101-point slider without ticks ranging from *strongly disagree* on the left to *strongly agree* on the right, which has been shown to lead to more precise results and allows for more statistical tests [97, 124]. Lastly, participants had to fill out the Affinity for Technology (ATI) [145] questionnaire.

5.2 Scenario Design

We systematically designed 2×8 scenarios, i.e., non-curious vs. curious systems, each in 8 scenarios. We took inspiration from related work on what kind of systems to use. We created scenarios with systems ranging from smart homes with no physical interaction capabilities to humanoid robots based on related work, c.f. Table 5. We modulated the scenarios based on the following factors: interaction (agent-interaction, human-agent-interaction, human-human-agent interaction), immersion (dialogue, vignette), and embodiment.

We designed the curious systems based on related work on human curiosity (i.e., epistemic curiosity [87], perceptual curiosity [32], and social curiosity [126]) and five factors to design curious

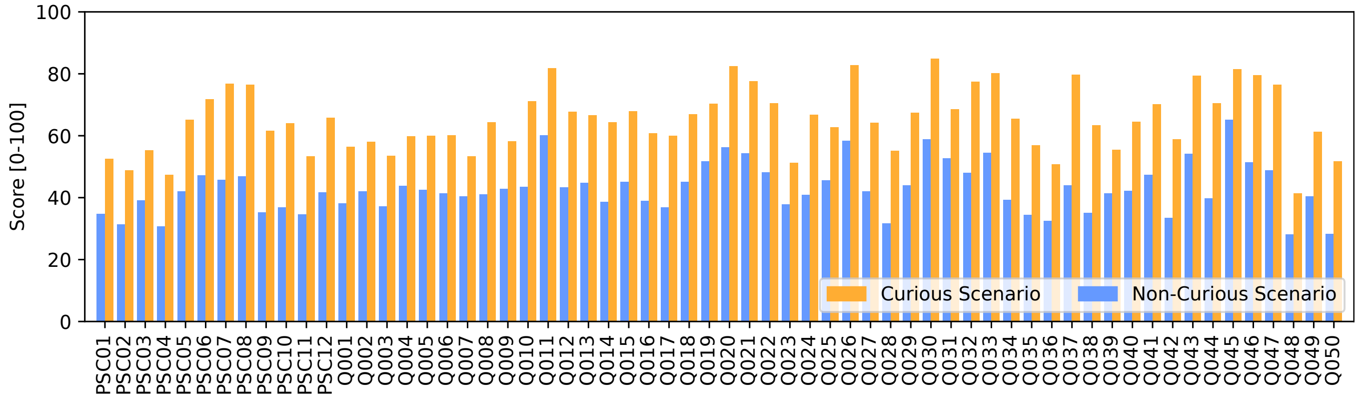

Figure 2: The mean ratings from our 400 participants across all 62 items.

systems: (1) Elements That Make Robots Appear Curious, (2) Ensuring Learning Progress, (3) Acceptance of Feedback, (4) Generating New Ideas, and (5) Exploration. We ensured that all our curious systems employed all of these behaviors. Furthermore, we ensured a clear difference between curious and non-curious systems and also varied how these systems expressed their curiosity, while the result of each scenario had to be the same for both the curious and non-curious scenarios.

5.3 Participants

We recruited all participants via Prolific and reimbursed them with £9 per hour. We rejected 48 participants for failed attention or comprehension checks. This resulted in a final valid set of 400 participants (182 female and 218 male). Participants took 14.53 min (SD = 6.91) on average to complete the survey. All participants spoke fluent English. Participation was voluntary, and we made participants aware that they could stop the survey at any point. On average, they were 31.8 years old (SD = 3.4). Nine participants had a doctoral degree, 85 had a master's degree, 184 had a bachelor's degree, 18 had vocational education, 97 had high school graduation, and 6 had another form of education. Participants were born in 76 different countries and lived in 42 different countries on six continents (261 in Europe, 52 in North America, 33 in Africa, 22 in South America, 21 in Asia, and 11 in Oceania). Their affinity for technology score using the ATI questionnaire [145] was 4.3 (SD = 0.8).

5.4 Results

The results for the 62 items are depicted in Figure 2. The mean score for the *curious scenarios* was 65.1 (SD = 28.4), and for the *non-curious scenarios*, the score was 42.9 on average (SD = 30.1). We conducted a MANOVA using IsCurious and Scenario as independent variables. We found significant main effects of both IsCurious ($F(62, 323) = 6.514$, $p < .001$, $Pillai's\ trace = 0.556$) and Scenario ($F(868, 4704) = 1.936$, $p < .001$, $Pillai's\ trace = 3.685$). This indicates that our scenarios were different for the overall perceived curiosity on the 62 items, and also, the scenarios produced a high diversity, ensuring the generalizability of the results.

6 Item Reduction: Exploratory Factor Analysis

In the following, we used Exploratory Factor Analysis (EFA) to determine the optimal number of factors and reduce the number of items.

6.1 Data Preparation

We visually inspected the items using histograms to identify items with high skewness and kurtosis, based on which three authors agreed to remove 12 items. Next, we mathematically checked that the absolute skewness was below 1. This resulted in five additional items being removed. Finally, we ensured that the item-total correlation was over 0.5; this removed two items. While we checked for kurtosis and item correlation, we found no conspicuous items. Thus, we removed 19 items in total, reducing the set to 43 items.

6.2 Exploratory Factor Analysis

We confirmed that Bartlett's sphericity test was significant ($\chi^2(903)$ = 15036.485, $p < .001$). This indicates that the remaining 43 items are sufficiently correlated and suitable for an EFA. Moreover, the Kaiser-Meyer-Olkin criterion [62] yields a score of 0.97, indicating a *marvelous* outcome. All individual question scores are > .96. Scree plots (using the method "principal axis factoring" and "maximum likelihood") suggest that three factors are best to represent the data. Finally, we performed EFA using *psych* [127] with Promax Rotation and Principal Axis Factoring, as we expected the three factors to be correlated based on our previous theoretical considerations and analysis of the qualitative data and oblique rotation is preferred for covarying factors [98].

We removed all factor loadings below 0.6 and all cross-loadings; removing 9 items. We show the resulting factor loadings in Table 6. Thus, we have 3 factors with a total of 22 items. As the next step, we looked at the items that are valuable to the construct. Here, we found that 8 items are conceptually very similar to other remaining items. See Table 6 for the 8 strongly similar items. As they are conceptually the same and have a high correlation, we decided to remove the duplicates. This step removed the items in the scale to 14, with 6 items for the first factor, 4 items for the second factor, and 4 items for the third factor.

Table 6: The EFA factor loading for the items without cross loading and ladings over .4. * items are marked in the removed column, are too similar with another question, and we removed them for this reason.

	Item	PA1	PA2	PA3	Removed*
PSC_3	The system is interested in interacting with novel objects.	0.886			
$Q1$	The system likes to interact with objects to find out their internal workings.	0.818			PSC_5
PSC_1	The system enjoys novel experiences.	0.801			
$Q2$	The system enjoys learning new skills.	0.801			PSC_1
$Q3$	The system is interested in novel activities.	0.759			PSC_1
PSC_2	The system is driven towards novel sensations.	0.753			
$Q4$	The system tries to understand novel objects.	0.729			PSC_3
$Q5$	The system is interested in performing tasks that it has not done before.	0.719			PSC_1
PSC_5	The system likes to find out how things work.	0.702			
$Q6$	The system is interested in objects that it has not encountered before.	0.686			PSC_3
$Q7$	The system likes interacting with new systems.	0.664			PSC_3
$Q8$	The system enjoys finding out how things work.	0.635			PSC_5
PSC_4	The system is excited by unpredictable tasks.	0.624			
$Q9$	The system likes exploring new environments.	0.601			
PSC_8	The system is inquisitive.		0.734		
PSC_7	The system likes to investigate.		0.649		
$Q10$	The system enjoys deepening its knowledge.		0.627		
PSC_6	The system invests substantial time to collect new knowledge.		0.614		
PSC_{10}	The system wants to know what others are doing.			0.803	
PSC_{11}	The system likes listening to others' conversations.			0.724	
PSC_9	The system likes finding out why others behave the way they do.			0.674	
PSC_{12}	The system is interested in learning more about people.			0.669	

The results of the 14-item scale show an overall Cronbach's alpha of $\alpha = .942$. The EFA with 14 items yields a Tucker-Lewis Index (TLI) of .992 and a Root Mean Square Error of Approximation (RSMEA) of .028. Thus, we concluded the EFA resulted in a good model fit.

7 Scale Evaluation

Next, we verified the factor structure found by the EFA and ensure our scale's measured construct was distinct from related constructs. For this, we conducted a confirmatory factor analysis (CFA), checked for discriminant and convergent validity, and checked test-retest reliability [16]. We used the same 16 scenarios we created for our EFA.

7.1 Test of Dimensionality

We used CFA to test our factor structure. Thus, we conducted another online survey with 340 new participants.

7.1.1 Procedure. We followed the same procedure as in Section 5.1. After introducing participants to the survey, we asked for their informed consent and demographics. Next, we randomly assigned each participant one scenario they had to listen to at least once. Participants also had the option to read the scenario and listen to it as many times as they wanted. We then asked participants to summarize the scenario briefly and then to rate the 14 items from our *PSC* scale using sliders ranging from 0 (strongly disagree) to 100 (strongly agree). Finally, we asked participants to fill out the ATI questionnaire [145].

7.1.2 Participants. We recruited 391 participants to participate in our study using Prolific and reimbursed them with £9 per hour. We rejected 51 participants due to failed attention or comprehension checks, leaving us with a final valid set of 340 participants (173 female, 163 male, 4 non-binary). Participants took 14.37 min (SD = 6.59) on average to complete the survey. On average, participants were 34.8 years old (SD = 11.0). Six participants had a doctoral degree, 78 had a master's degree, 156 had a bachelor's degree, 16 had vocational education, 78 had high school graduation, and six had another form of education. Participants were born in 43 countries and lived in 35 countries (206 in Europe, 26 in North America, 26 in South America, 13 in Oceania, 41 in Africa, and 3 in Asia). Their affinity for technology score using the ATI questionnaire [145] was 4.2 (SD = 0.8).

7.1.3 Confirmatory Factor Analysis. We used the lavaan package in R [129] to conduct the CFA. Boateng et al. [16] summarize acceptable fit indices and optimal thresholds being $RMSEA < 0.08$, $TLI \geq 0.95$, $CFI \geq 0.95$, and $SRMR \geq 0.95$. The CFA revealed that our 14-item model had some values outside of the acceptable thresholds ($RMSEA = 0.084$, $TLI = 0.926$, $CFI = 0.94$, $SRMR = 0.047$) [9, 16, 100, 147]. Thus, we looked at the item loadings, cross-loadings, and item covariances to perform an item purification process [158, 159]. We found that item $Q10$ is covarying with four items (PSC_5, PSC_4, PSC_{10}, $Q9$). Item $Q9$ has the lowest loadings of our resulting model and loads equally towards 2 factors. Removing these two items results in a model with 12 items and adequate model fit indices ($RMSEA = 0.06$, $TLI = 0.965$, $CFI = 0.973$, $SRMR = 0.038$) [9, 16, 100, 147], see Table 7 for the final items. All

Table 7: The resulting scale with their loadings after CFA.

	Item	PA1	PA2	PA3	Factor
PSC_1	The system enjoys novel experiences.	0.991			Explorative (PSC_E)
PSC_2	The system is driven towards novel sensations.	0.846			Explorative (PSC_E)
PSC_3	The system is interested in interacting with novel objects.	0.746			Explorative (PSC_E)
PSC_4	The system is excited by unpredictable tasks.	0.579			Explorative (PSC_E)
PSC_5	The system likes to find out how things work.	0.554			Explorative (PSC_E)
PSC_6	The system invests substantial time to collect new knowledge.		0.844		Investigative (PSC_I)
PSC_7	The system likes to investigate.		0.711		Investigative (PSC_I)
PSC_8	The system is inquisitive.		0.635		Investigative (PSC_I)
PSC_9	The system likes finding out why others behave the way they do.			0.785	Social (PSC_S)
PSC_{10}	The system wants to know what others are doing.			0.769	Social (PSC_S)
PSC_{11}	The system likes listening to others' conversations.			0.683	Social (PSC_S)
PSC_{12}	The system is interested in learning more about people.			0.587	Social (PSC_S)

remaining items have loadings ≥ 0.55, and seven have loadings > 0.7 to their factors. We name the three factors: *Perceived Explorative Curiosity* (PSC_E), *Perceived Investigative Curiosity* (PSC_I), and *Perceived Social Curiosity* (PSC_S).

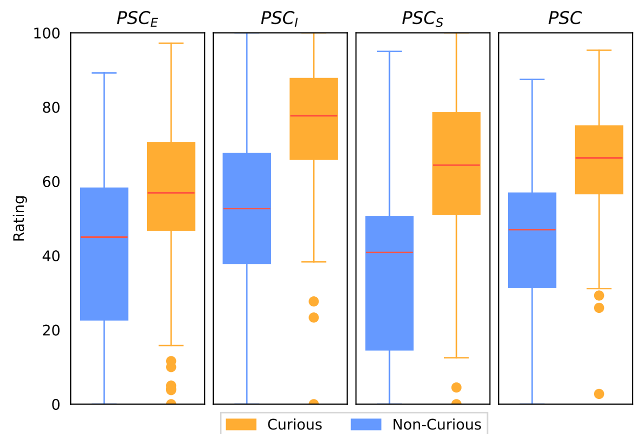

Figure 3: Boxplots of the CFA-dataset, split by the curiosity.

7.1.4 Internal Consistency. We measured the internal consistency of the *PSC* by calculating Cronbach's alpha for both the entire scale ($\alpha = 0.92$) and each subscale, PSC_E ($\alpha = 0.89$), PSC_I ($\alpha = 0.82$), and PSC_S ($\alpha = 0.85$), which are all over the preferred threshold of 0.8 [16, 30]. Furthermore, we found $\omega_h = 0.84$, and $\omega_{total} = 0.94$ as reliability coefficients [128], which show the total second-order factor variance-covariance and the reliability of combined first-order factors respectively [30]. Cheung et al. [30] state that while Cronbach's alpha is the most commonly reported metric to show reliability, it is not based on structural equation modeling to better measure and report the construct reliability. We calculated the split-half reliability using the psych R package [127] with a sample of 10000. We found that $\beta = 0.84$, representing the smallest split-half reliability [128]. This is above the highest threshold of 0.8 and, thus, shows adequate reliability [30].

Furthermore, we conducted t-tests for each subscale on the final CFA dataset. The data for PSC_E ($W = .963$, $p < .001$), PSC_I ($W = .947$, $p < .001$), and PSC_S ($W = .970$, $p < .001$) all is not normally distributed. Thus, we conducted kruskal wallis tests to determine whether curiosity in the scenarios affected the three factors. We found a significant effect for all subscales ($p < .001$ for all). We then performed Mann-Whitney U tests as post hoc tests and found that for all subscales, the curious scenarios were rated significantly more curious, see Figure 3. We furthermore correlated participants' age with perceived curiosity. We computed Spearman's rank correlation to assess the relationship between perceived system curiosity and age. There is no significant correlation between age and *PSC* ($r = -.094$, $p = .084$) and PSC_I ($r = -.016$, $p = .776$). There is a significant, weak negative correlation between age and PSC_E ($r = -.115$, $p = .035$) and PSC_S ($r = -.110$, $p = .042$). Thus, we see a slight decrease in perceived explorative and social curiosity for older participants.

7.2 Test-Retest Reliability

We re-recruited a subset of participants who had previously taken part in our CFA study to evaluate the test-retest reliability and, thus, the stability of our scale. We again provided every participant with a scenario and our final 12-item *PSC* scale. The procedure was the same as the previous procedures. Every participant had at least 14 days between the two questionnaires.

7.2.1 Participants. We invited 105 participants from the 340 valid participants from our CFA study via Prolific and reimbursed them with £9 per hour. We had to exclude five participants due to failed comprehension or attention checks. This left us with 100 valid participants (50 female, 49 male, and 1 non-binary), with 50 each for the curious and non-curious scenarios. Participants took, on average, 7.4 min ($SD = 3.5$) to complete the survey. The average age was 35.4 years ($SD = 10.6$), and participants were born in 30 countries.

7.2.2 Results. We calculated the intra-class coefficient (ICC) and Pearson product-moment correlation [130], as suggested by Boateng et al. [16], to measure the reliability of our scale. We found high scores for both ICC and Pearson correlation for all subscales and the total *PSC* scale, indicating great reliability of the *PSC* scale, see Table 8.

7.3 Construct Validity

We selected measures that the literature links to curiosity to assess whether our construct of perceived system curiosity correlates with these measurements. We used the same participants as for Section 7.1.2.

7.3.1 Measurements. Literature has found that many human characteristics in the system lead to higher levels of perceived human likeness [1, 47, 112]. Curiosity is generally associated with human likeness. Thus, we expect a curious system to score high on the Godspeed anthropomorphism subscale [8] and test for convergent validity.

Intelligence is associated with fast learning and collecting information, and curiosity is a mechanism that correlates highly to information acquisition. Grand [46] highlights that intelligence is linked with curiosity, which leads to learning and giving one more knowledge. However, previous work has also shown that intelligence and curiosity do not correlate as much as one would expect [125]. We still expect a curious system to score high on the Godspeed perceived intelligence subscale [8] and test for convergent validity.

Additionally, Leonard and Harvey [79] linked curiosity to emotional intelligence. Thus, we expect a curious system to score high on perceived social intelligence subscales from the Perceived Social Intelligence (PSI) scale [7] and test four fitting subscales for convergent validity.

The expression of curiosity makes systems use more expressions and gestures. Curiosity is a precedes of actions. Curiosity leads to taking the initiative as that entity oftentimes needs to steer the interaction. Thus, we expect the Godspeed animacy subscale [8] to correlate with our *PSC* and test for convergent validity.

Likeability is a concept that might be associated with anthropomorphism as it contains concepts that are human-related; however, in the context of curiosity, it is highly dependent, as some individuals would prefer straightforward interactions while others would prefer fine-grained decision-making. For example, when being curious, this means that at this moment, the interaction is drawn out to be longer, as the system takes time actually to do curious expressions. Thus, we test the likeability subscale from the Godspeed questionnaire [8] for discriminant validity.

In general, a curious system seems correlated to usability, as systems that do not work are unlikely to rank high in usability. Personalization has been previously linked to usability [49, 54]. However, in the context of curious systems, this personalization comes at an expense in the initial interactions, which naturally become longer, conflicting with another usability concept, which is efficiency; therefore, in the first interactions with a curious system, the usability of the system does not clearly correlate with system curiosity. Thus, we expect the system usability scale [20] to be uncorrelated to the *PSC* and test for discriminant validity.

Discriminant Validity. We used semTools [59] discriminant validity method to assess discriminate validity. Here, we build a model with our scales and all the scales we want to test for discriminant validity. We then perform CFA on this model and compare the CI_{upper} with thresholds proposed by Rönkkö and Cho [132]. For PSC_E and likeability we find $CI_{upper} = 0.471$, For PSC_I and likeability we find $CI_{upper} = 0.399$, For PSC_S and likeability we find $CI_{upper} = 0.244$, For PSC_E and SUS we find $CI_{upper} = 0.137$, For PSC_I and SUS we find $CI_{upper} = 0.072$, For PSC_S and SUS we find $CI_{upper} = -0.071$. These are all below the threshold of 0.8; thus, we found discriminant validity for all.

Table 8: Intraclass correlation coefficients (ICC) (absolute rater average) and Pearson Correlation for all subscales PSC_E, PSC_I, and PSC_S and the general scale *PSC* for the test-retest study.

Scale	Pearson Corr.		Intra-Class Coefficient			
	$r(98)$	p	ICC	CI_{lower}	CI_{upper}	p
PSC_E	.73	< .001	.85	.77	.90	<.001
PSC_I	.79	< .001	.87	.81	.91	<.001
PSC_S	.71	< .001	.83	.74	.88	<.001
PSC	.81	< .001	.90	.84	.93	<.001

Table 9: The values for the three conditions for construct validity. For convergent validity between two subscales, $\omega > .7$, $AVE > .5$, and $SFL > .5$ [30].

Subscale	PSC_E			PSC_I			PSC_S		
	ω	AVE	SFL	ω	AVE	SFL	ω	AVE	SFL
Intelligence	.899	.626	.438	.829	.613	.398	.840	.576	.218
Anthropo.	.897	.626	.600	.825	.612	.463	.842	.577	.431
Animacy	.897	.626	.537	.825	.611	.502	.842	.577	.371
PSI-FRD	.897	.626	.655	.825	.612	.598	.841	.576	.674
PSI-HLP	.897	.626	.555	.824	.611	.568	.842	.577	.430
PSI-SOC	.896	.626	.580	.827	.612	.603	.840	.576	.549
PSI-RC	.898	.626	.675	.825	.612	.529	.842	.577	.589

Convergent Validity. For testing for convergent validity, we follow the proposed approach by Cheung et al. [30]. They propose that two constructs are convergent valid if the three following conditions are fulfilled: (1) The ω-values > 0.7, (2) the standardized factor loadings > 0.5, and (3) the AVE values are > 0.5. We used the semTools reliability function [59] to calculate ω[2], and constructed a CFA to gather the standardized factor loadings. We depict all values in Table 9. We found no convergent validity between our three subscales and Godspeed intelligence, we found convergent validity for Godspeed anthropomorphism and PSC_E, but not for PSC_I and PSC_S. We found convergent validity between Godspeed animacy and PSC_E and PSC_I, but not for PSC_S. We found convergent validity between PSI-FRD and all of our subscales. We found convergent validity between PSI-HLP and PSC_E and PSC_I, but not PSC_S, We found convergent validity between PSI-SOC and all our subscales. We found convergent validity between PSI-RC and all our subscales.

[2]We use the most conservative ω_3

8 Discussion

Our *PSC* scale measures the perceived system curiosity from the perspective of a user who has interacted with that system. We suggest that the *PSC* scale can be used to evaluate the perceived curiosity for any technical system, ranging from voice assistants to robots. In the following, we provide guidelines on how to use our scale, and discuss its dimensions, limitations, and insights.

8.1 Usage Guidelines

When using the *PSC* scale, we argue that the 12 items should be shown in randomized order. This helps to prevent order effects and biases [134, 146]. As we also randomized the order of items in all of our validation steps, we argue that our construct is reliable with randomized item order. Furthermore, we evaluated the *PSC* scale with VAS sliders, ranging from 0 (strongly disagree) to 100 (strongly disagree). We recommend using continuous VAS sliders without ticks as this has been shown to lead to higher data quality and allows for more statistical tests [97, 124]. Our scale does not have reversed items. All items in our scale should have an equal weighting for both each subscale and the total score. Thus, these are the resulting formulas to calculate the scores for the subscales PSC_E, PSC_I, and PSC_S and the generally perceived curiosity *PSC*:

$$PSC_E = (PSC_1 + PSC_2 + PSC_3 + PSC_4 + PSC_5) \;/\; 5$$
$$PSC_I = (PSC_6 + PSC_7 + PSC_8) \;/\; 3$$
$$PSC_S = (PSC_9 + PSC_{10} + PSC_{11} + PSC_{12}) \;/\; 4$$
$$PSC = (PSC_E + PSC_I + PSC_S) \;/\; 3$$

For now, we cannot give interpretations of what scores mean. However, we generally recommend using appropriate statistical tests to compare subscale and total scale values between the tested systems and find whether the systems are statistically different[3]. When combining all subscales, the total score perceived system curiosity value (*PSC*) is calculated. We recommend reporting both subscale and total scale results. We provide details usage instructions in Appendix A.

8.2 Dimensions

We generated our items using a bottom-up approach, drawing from four sources, primarily literature on human curiosity. However, the three resulting factors of our scale still can correspond to three distinct types of curiosity: perceptual, epistemic, and social.

While curiosity is defined in various ways, most psychological literature focuses on intrinsic human reasons for curiosity. However, to assess curiosity in another entity, observable behaviors must be identified. We found that perceptual curiosity involves interest in sensory stimuli [32, 58], epistemic curiosity relates to seeking knowledge [91], often through information-seeking behavior, and social curiosity pertains to interest in people [71]. Reio Jr. et al. [123] similarly identified three curiosity types: cognitive, physical thrill-seeking, and social thrill-seeking.

Our subscales align with these types: PSC_S measures social curiosity, PSC_I reflects epistemic curiosity through inquisitiveness, and PSC_E captures both perceptual and epistemic curiosity, focusing on interest in sensations and objects and the search for information. Literature suggests both epistemic and perceptual curiosity are driven by a desire for new information that can lead to lasting knowledge [85, 102]. Litman and Spielberger [91] found through factor analysis that perceptual and epistemic curiosity are related but distinct dimensions, sharing an underlying structure. However, one more differentiation between perceptual and epistemic curiosity is how it is triggered. Sakaki et al. [135] propose that perceptual curiosity is triggered by novelty, and epistemic curiosity is triggered by prior knowledge. As three questions in PSC_E are regarding novel experiences (PSC_1), novel sensations (PSC_2), and novel objects (PSC_3), this supports that PSC_E indeed matches perceptual curiosity from human literature and PSC_I matches epistemic curiosity.

This structure is furthermore supported by the correlations of *PSC* with other measurements shown in Section 7.3. We found convergent validity between most subscales of the PSI questionnaire measuring perceived social intelligence and PSC_S. Social curiosity is also one of Kashdan et al. [65]'s five-dimensional curiosity model. Additionally, we found convergent validity between PSC_E and PSC_I and animacy. Questions in both of these subscales describe the robot actively engaging with the environment. Furthermore, PSC_E includes questions that can also be classified as joyous exploration by one of the five dimensions from Kashdan et al. [65], e.g. PSC_1. *The system enjoys novel experiences.* or PSC_6 *The system likes to find out how things work.* Thus, while this is the first scale that measures perceived curiosity instead of self-reported curiosity, we still find that our scale matches concepts from related work on human curiosity.

8.3 Curiosity as a System Attribute

Curiosity, as conceptualized in our scale, draws inspiration from human traits and various forms of expressing curiosity. While intrinsic motivation is a central concept in human curiosity [33, 131], systems lack this intrinsic drive; instead, their curiosity can be modeled as a learning function, such as in reinforcement learning [14, 104]. However, when humans are curious, they exhibit observable behaviors that others can recognize as them being curious. Since our scale evaluates a system based on user interactions, it focuses on the user's perception of observable curiosity. This approach led to the development of three subscales: perceived explorative curiosity, perceived investigative curiosity, and perceived social curiosity. By formalizing curiosity as a system attribute, this work provides a structured framework for studying its impact on human-system interactions, laying the groundwork for understanding when and how curious behaviors can enhance usability and engagement.

Curious systems can offer several advantages, particularly in environments requiring adaptability and engagement. By continuously learning from their surroundings, these systems can personalize interactions and enhance problem-solving. Additionally, the anthropomorphic traits associated with curiosity, like asking questions or exploring new possibilities, can enhance user trust, promote transparency, and create a more engaging interaction experience. However, high levels of curiosity may not always be desired and beneficial. In task-oriented scenarios, excessive exploration

[3] As we used promax rotation to derive our factor structure, each subscale can be used independently [28].

may hinder efficiency or frustrate users who prefer straightforward interactions. Overly inquisitive systems can raise privacy concerns, as users might question why the system requires such detailed information. Balancing curiosity with context is crucial to ensure systems support user goals without overwhelming them.

To determine the appropriate level of curiosity for a given situation, it is essential to measure and evaluate curiosity across diverse contexts and systems. Different scenarios demand varying degrees of curious behavior, and understanding these nuances requires systematic experimentation. By applying our PSC scale, researchers can assess how users perceive and respond to curiosity in different environments, ranging from task-oriented systems like chat-bots to adaptive and exploratory robots in education or caregiving. These measurements enable us to identify where curiosity enhances usability and engagement and where it becomes counterproductive. Over time, such studies can provide valuable insights into optimizing the balance of curiosity, tailoring it to align with specific user needs and goals across varied applications.

8.4 Limitations

While we made a huge effort to develop and validate with as much rigor as possible and thus propose that the *PSC* scale brings value to assess perceived system curiosity, we want to acknowledge some limitations.

First, we recruited participants for the relevance rating and workshop via convenience sampling. While we put strong emphasize on inviting participants with wide but fitting background this might not be a full sample of users who might use this scale. Furthermore, the testing and evaluation of the scale were done with online participants from Prolific. Here, we balanced participants across genders and continents. However, this might still not be a fully representative sample of the population who might use this scale [37]. Furthermore, online studies have been shown to induce selection bias [6, 12]. However, as a large sample is needed for the scale validation process is needed, online studies are still the only way to acquire a large sample. Still, future work should validate whether this scale is valid for other samples. We intend to apply the *PSC* scale to interactive experiments and, thus, evaluate it further.

Second, the initials reverse-coded items were all eliminated in the process of creating a valid and unique item set. While only using positive items can lead to acquiescence bias [144], the majority of the literature suggests that mixing negatively and positively coded items can lead to measurement problems [136, 155], especially in the process of scale validation [17].

Lastly, as there are no validated scales and models on perceived curiosity, neither for humans nor systems, we had to create our items in a bottom-up process based mostly on the literature on human curiosity. Jung [60] states that to measure the affective dimensions of robots, there needs to be affective grounding, meaning that the user needs to first understand what the robot wants to convey before they can rate it. We tried to encompass this by only wording items in a way that the measured concept can be observable. Still, it is unclear how the construct of human curiosity can be mapped to perceived system curiosity, and this needs to be evaluated over time by the usage of this scale.

9 Conclusion

In this work, we developed and validated the Perceived System Curiosity (*PSC*) scale, which has 12 items (see Table 10), and the three factors perceived explorative curiosity (PSC_E), perceived investigative curiosity (PSC_I), and perceived social curiosity (PSC_S). We followed a standardized scale development process [16], which is based on domain identification and item generation, item validation through expert studies, and item testing and scale validation through large-scale user studies. We used 16 scenarios in which we described curious and non-curious systems for this process. Our scale shows high internal consistency, great reliability, divergent validity, and convergent validity. We envision that this scale will support the development of curious systems, ultimately leading to more natural learning systems.

References

[1] Hooman Aghaebrahimi Samani, Adrian David Cheok, Mili John Tharakan, Jeffrey Koh, and Newton Fernando. 2011. A Design Process for Lovotics. In *Human-Robot Personal Relationships*. Springer Berlin Heidelberg, Berlin, Heidelberg, 118–125. doi:10.1007/978-3-642-19385-9_15

[2] Jamilah Ahmad and Nyet Moi Siew. 2021. Curiosity Towards STEM Education: A Questionnaire for Primary School Students. *Journal of Baltic Science Education* 20, 2 (2021), 289–304. doi:10.33225/jbse/21.20.289

[3] Aristotle. 2006. *Metaphysics*. Digireads, Stilwell, KS. 3 pages.

[4] Filippo Aschieri and Ilaria Durosini. 2015. Development of the Self-Curiosity Attitude-Interest Scale. *Testing, Psychometrics, Methodology in Applied Psychology* 22 (2015), 327–347. doi:10.4473/tpm22.3.2

[5] Filippo Aschieri, Angélica Quiroga, Semira Tagliabue, and Ilaria Durosini. 2021. Psychometric properties of the Self-Curiosity Attitude-Interest Scale in two paired samples from Italy and Mexico. *Testing, Psychometrics, Methodology in Applied Psychology* 28 (2021), 415–426. doi:10.4473/tpm28.4.2

[6] Helen L. Ball. 2019. Conducting Online Surveys. 35, 3 (2019), 413–417. doi:10.1177/0890334419848734

[7] Kimberly A. Barchard, Leiszle Lapping-Carr, R. Shane Westfall, Andrea Fink-Armold, Santosh Balajee Banisetty, and David Feil-Seifer. 2020. Measuring the Perceived Social Intelligence of Robots. *J. Hum.-Robot Interact.* 9, 4, Article 24 (sep 2020), 29 pages. doi:10.1145/3415139

[8] Christoph Bartneck, Dana Kulić, Elizabeth Croft, and Susana Zoghbi. 2009. Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International journal of social robotics* 1 (2009), 71–81. doi:10.1007/s12369-008-0001-3

[9] Peter M. Bentler. 1990. Comparative Fit Indexes in Structural Models. 107, 2 (1990), 238–246. doi:10.1037/0033-2909.107.2.238

[10] Daniel Ellis Berlyne. 1954. A Theory of Human Curiosity. *British Journal of Psychology. General Section* 45, 3 (1954), 180–191. doi:10.1111/j.2044-8295.1954.tb01243.x

[11] Daniel Ellis Berlyne. 1960. *Conflict, arousal, and curiosity*. McGraw-Hill Book Company, New York. 79–265 pages. doi:10.1037/11164-000

[12] Jelke Bethlehem. 2010. Selection Bias in Web Surveys. 78, 2 (2010), 161–188. doi:10.1111/j.1751-5823.2010.00112.x

[13] Laura Bishop, Anouk Van Maris, Sanja Dogramadzi, and Nancy Zook. 2019. Social Robots: The Influence of Human and Robot Characteristics on Acceptance. 10, 1 (2019), 346–358. doi:10.1515/pjbr-2019-0028

[14] Erdem Biyik, Malayandi Palan, Nicholas C. Landolfi, Dylan P. Losey, and Dorsa Sadigh. 2019. Asking Easy Questions: A User-Friendly Approach to Active Reward Learning. *ArXiv* abs/1910.04365 (2019). https://api.semanticscholar.org/CorpusID:204008187

[15] Matthias Bluemke, Lukas Engel, David J. Grüning, and Clemens M. Lechner. 2024. Measuring Intellectual Curiosity across Cultures: Validity and Comparability of a New Scale in Six Languages. *Journal of personality assessment* 106, 2 (2024), 156–173. doi:10.1080/00223891.2023.2199863

[16] Godfred O. Boateng, Torsten B. Neilands, Edward A. Frongillo, Hugo R. Melgar-Quiñonez, and Sera L. Young. 2018. Best Practices for Developing and Validating Scales for Health, Social, and Behavioral Research: A Primer. *Frontiers in public health* 6 (2018), 149. doi:10.3389/fpubh.2018.00149

[17] B. Bynum Boley, Evan Jordan, and Kyle Maurice Woosnam. 2021. Reversed Polarity Items in Tourism Scales: Best Practice or Dimensional Pitfall? 24, 4 (2021), 466–478. doi:10.1080/13683500.2020.1774517

[18] Cynthia Breazeal. 2002. *Designing Sociable Robots*. The MIT Press. doi:10.7551/mitpress/2376.001.0001

[19] C. Breazeal. 2004. Function meets style: insights from emotion theory applied to HRI. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 34, 2 (2004), 187–194. doi:10.1109/tsmcc.2004.826270

[20] John Brooke et al. 1996. SUS-A quick and dirty usability scale. *Usability evaluation in industry* 189, 194 (1996), 4–7.

[21] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D. Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel Ziegler, Jeffrey Wu, Clemens Winter, Chris Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. Language Models are Few-Shot Learners. In *Advances in Neural Information Processing Systems*, Vol. 33. Curran Associates, Inc, 1877–1901. https://proceedings.neurips.cc/paper_files/paper/2020/file/1457c0d6bfcb4967418bfb8ac142f64a-Paper.pdf

[22] Till Johannes Bugaj, Tim Alexander Schwarz, Valentin Terhoeven, Ede Nagy, Anna Cranz, Hans-Christoph Friederich, and Christoph Nikendei. 2023. Measuring an understudied factor in medical education - development and validation of the medical curiosity scale. *Medical education online* 28, 1 (2023), 2198117. doi:10.1080/10872981.2023.2198117

[23] Yuri Burda, Harri Edwards, Deepak Pathak, Amos Storkey, Trevor Darrell, and Alexei A. Efros. 2018. Large-Scale Study of Curiosity-Driven Learning. arXiv:1808.04355 [cs, stat]

[24] Yuri Burda, Harrison Edwards, Deepak Pathak, Amos J. Storkey, Trevor Darrell, and Alexei A. Efros. 2018. Large-Scale Study of Curiosity-Driven Learning. *ArXiv* abs/1808.04355 (2018). https://api.semanticscholar.org/CorpusID:51979536

[25] Reijo Byman. 2016. The development of a gender-free curiosity inventory. *Personality and Individual Differences* 101 (2016), 177–184. doi:10.1016/j.paid.2016.05.039

[26] Richard W. Byrne. 2013. Animal Curiosity. 23, 11 (2013), R469–r470. doi:10.1016/j.cub.2013.02.058

[27] Elizabeth J. Carter, Laura M. Hiatt, and Stephanie Rosenthal. 2022. You're delaying my task?! The Impact of Task Order and Motive on Perceptions of a Robot. In *2022 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI '22)*. IEEE, 304–312. doi:10.1109/hri53351.2022.9889307

[28] Michelle Alessandra De Castro, Valéria Troncoso Baltar, Soraya Sant'Ana De Castro Selem, Dirce Maria Lobo Marchioni, and Regina Mara Fisberg. 2015. Empirically Derived Dietary Patterns: Interpretability and Construct Validity According to Different Factor Rotation Methods. 31, 2 (2015), 298–310. doi:10.1590/0102-311x00070814

[29] Jessy Ceha, Nalin Chhibber, Joslin Goh, Corina McDonald, Pierre-Yves Oudeyer, Dana Kulić, and Edith Law. 2019. Expression of Curiosity in Social Robots: Design, Perception, and Effects on Behaviour. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–12. doi:10.1145/3290605.3300636

[30] Gordon W. Cheung, Helena D. Cooper-Thomas, Rebecca S. Lau, and Linda C. Wang. 2024. Reporting Reliability, Convergent and Discriminant Validity with Structural Equation Modeling: A Review and Best-Practice Recommendations. 41, 2 (2024), 745–783. doi:10.1007/s10490-023-09871-y

[31] Cicero and Tullius, Markus. 1931. *De Finibus* (repr ed.). Harvard University Press, Cambridge, Mass. 449 pages.

[32] Robert P. Collins, Jordan A. Litman, and Charles D. Spielberger. 2004. The measurement of perceptual curiosity. *Personality and Individual Differences* 36, 5 (2004), 1127–1141. doi:10.1016/s0191-8869(03)00205-8

[33] Stefano I. Di Domenico and Richard M. Ryan. 2017. The Emerging Neuroscience of Intrinsic Motivation: A New Frontier in Self-Determination Research. *Frontiers in Human Neuroscience* 11 (March 2017). doi:10.3389/fnhum.2017.00145

[34] Malcolm Doering, Phoebe Liu, Dylan F. Glas, Takayuki Kanda, Dana Kulić, and Hiroshi Ishiguro. 2019. Curiosity Did Not Kill the Robot: A Curiosity-based Learning System for a Shopkeeper Robot. *ACM Transactions on Human-Robot Interaction* 8, 3 (2019), 1–24. doi:10.1145/3326462

[35] Ed Donnellan, Sumeyye Aslan, Greta M. Fastrich, and Kou Murayama. 2022. How Are Curiosity and Interest Different? Naïve Bayes Classification of People's Beliefs. 34, 1 (2022), 73–105. doi:10.1007/s10648-021-09622-9

[36] Rachit Dubey and Thomas L Griffiths. 2020. Understanding exploration in humans and machines by formalizing the function of curiosity. *Current Opinion in Behavioral Sciences* 35 (2020), 118–124. doi:10.1016/j.cobeha.2020.07.008 Curiosity (Explore vs Exploit).

[37] Mark Damian Duda and Joanne L. Nobile. 2010. The Fallacy of Online Surveys: No Data Are Better Than Bad Data. 15, 1 (2010), 55–64. doi:10.1080/10871200903244250

[38] Thomas Erickson and Wendy A. Kellogg. 2000. Social Translucence: An Approach to Designing Systems That Support Social Processes. 7, 1 (2000), 59–83. doi:10.1145/344949.345004

[39] Julia Fink. 2012. Anthropomorphism and Human Likeness in the Design of Robots and Human-Robot Interaction. In *Social Robotics*. Vol. 7621. Springer Berlin Heidelberg, 199–208. doi:10.1007/978-3-642-34103-8_20

[40] Sigmund Freud. 2017. *Three essays on the theory of sexuality: The 1905 edition*. Verso Books.

[41] Mingcai Hu Fu Guo, Xiaohui Tian and Zeyu Zhang. 2024. Affective Design of Smart TV Navigation Interface Considering the Diversity of User Needs. *International Journal of Human–Computer Interaction* 40, 12 (2024), 3092–3111. doi:10.1080/10447318.2023.2179217

[42] Pascale Fung, Dario Bertero, Yan Wan, Anik Dey, Ricky Ho Yin Chan, Farhad Bin Siddique, Yang Yang, Chien-Sheng Wu, and Ruixi Lin. 2018. Towards Empathetic Human-Robot Interactions. In *Computational Linguistics and Intelligent Text Processing*. Springer International Publishing, Cham, 173–193. doi:10.1007/978-3-319-75487-1_14

[43] Aimi S. Ghazali, Jaap Ham, Emilia I. Barakova, and Panos Markopoulos. 2018. Effects of Robot Facial Characteristics and Gender in Persuasive Human-Robot Interaction. 5 (2018), 73. doi:10.3389/frobt.2018.00073

[44] Stephen E. Glickman and Richard W. Sroges. 1966. Curiosity in Zoo Animals. 26, 1-2 (1966), 151–187. doi:10.1163/156853966x00074

[45] Goren Gordon, Cynthia Breazeal, and Susan Engel. 2015. Can Children Catch Curiosity from a Social Robot?. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction* (Portland, Oregon, USA) *(HRI '15)*. Association for Computing Machinery, New York, NY, USA, 91–98. doi:10.1145/2696454.2696469

[46] S. Grand. 1998. Curiosity created the cat [the relationship between curiosity and intelligence]. *IEEE Intelligent Systems and their Applications* 13, 3 (1998), 2–4. doi:10.1109/5254.683171

[47] Alla Gubenko, Christiane Kirsch, Jan Nicola Smilek, Todd Lubart, and Claude Houssemand. 2021. Educational Robotics and Robot Creativity: An Interdisciplinary Dialogue. *Frontiers in robotics and AI* 8 (2021), 662030. doi:10.3389/frobt.2021.662030

[48] Belinda A. Hall, Vicky Melfi, Alicia Burns, David M. McGill, and Rebecca E. Doyle. 2018. Curious Creatures: A Multi-Taxa Investigation of Responses to Novelty in a Zoo Environment. 6 (2018), e4454. doi:10.7717/peerj.4454

[49] V.L. Hanson, A. Snow-Weaver, and S. Trewin. 2006. Software Personalization to Meet the Needs of Older Adults. 5, 3 (2006), 160–169. doi:10.4017/gt.2006.05.03.005.00

[50] Harry F. Harlow, Margaret Kuenne Harlow, and Donald R. Meyer. 1950. Learning motivated by a manipulation drive. *Journal of experimental psychology* 40, 2 (1950), 228–234. doi:10.1037/h0056906

[51] Paul L. Harris, Deborah T. Bartz, and Meredith L. Rowe. 2017. Young Children Communicate Their Ignorance and Ask Questions. 114, 30 (2017), 7884–7891. doi:10.1073/pnas.1620745114

[52] Freda-Marie Hartung and Britta Renner. 2011. Social Curiosity and Interpersonal Perception: A Judge × Trait Interaction. 37, 6 (2011), 796–814. doi:10.1177/0146167211400618

[53] Harold Harty and Dwight Beall. 1984. Toward the development of a children's science curiosity measure. *Journal of Research in Science Teaching* 21, 4 (1984), 425–436. doi:10.1002/tea.3660210410

[54] Martina Hasenjager and Heiko Wersing. 2017. Personalization in Advanced Driver Assistance Systems and Autonomous Vehicles: A Review. In *2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)* (Yokohama, 2017-10). IEEE, 1–7. doi:10.1109/itsc.2017.8317803

[55] Clark L. Hull. 1943. Principles of Behavior. An Introduction to Behavior Theory. *The Journal of Philosophy* 40, 20 (1943). doi:10.2307/2019960

[56] Ruchika Jain, Naval Garg, and Shikha N. Khera. 2023. Effective human–AI work design for collaborative decision-making. *Kybernetes* 52, 11 (01 Jan 2023), 5017–5040. doi:10.1108/k-04-2022-0548

[57] William James. 1899. *Talks to teachers on psychology and to students on some of life's ideals*. Metropolitan Books/Henry Holt and Company, New York. 45 pages. doi:10.1037/10814-000

[58] Marieke Jepma, Rinus G. Verdonschot, Henk Van Steenbergen, Serge A. R. B. Rombouts, and Sander Nieuwenhuis. 2012. Neural Mechanisms Underlying the Induction and Relief of Perceptual Curiosity. 6 (2012). doi:10.3389/fnbeh.2012.00005

[59] Terrence D. Jorgensen, Sunthud Pornprasertmanit, Alexander M. Schoemann, and Yves Rosseel. 2012. *semTools: Useful Tools for Structural Equation Modeling*. doi:10.32614/CRAN.package.semTools

[60] Malte F. Jung. 2017. Affective Grounding in Human-Robot Interaction. In *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction* (New York, NY, USA, 2017-03-06) *(HRI '17)*. Association for Computing Machinery, 263–273. doi:10.1145/2909824.3020224

[61] Dan M. Kahan, Asheley Landrum, Katie Carpenter, Laura Helft, and Kathleen Hall Jamieson. 2017. Science Curiosity and Political Information Processing. *Political Psychology* 38, S1 (2017), 179–199. doi:10.1111/pops.12396

[62] Henry F. Kaiser. 1974. An Index of Factorial Simplicity. *Psychometrika* 39, 1 (1974), 31–36. doi:10.1007/bf02291575

[63] Max Kanderske. 2023. Domesticating Motile Media: The Routes and Routines of Vacuum Robots. *Digital Culture & Society* 9, 1 (2023), 71–98. doi:10.14361/dcs-2023-0105

[64] Todd B. Kashdan, David J. Disabato, Fallon R. Goodman, and Patrick E. McKnight. 2020. The Five-Dimensional Curiosity Scale Revised (5DCR): Briefer subscales while separating overt and covert social curiosity. *Personality and Individual Differences* 157 (2020), 109836. doi:10.1016/j.paid.2020.109836

[65] Todd B. Kashdan, Matthew W. Gallagher, Paul J. Silvia, Beate P. Winterstein, William E. Breen, Daniel Terhar, and Michael F. Steger. 2009. The Curiosity and Exploration Inventory-II: Development, Factor Structure, and Psychometrics. *Journal of research in personality* 43, 6 (2009), 987–998. doi:10.1016/j.jrp.2009.04.011

[66] Todd B. Kashdan, Fallon R. Goodman, David J. Disabato, Patrick E. McKnight, Kerry Kelso, and Carl Naughton. 2020. Curiosity has comprehensive benefits in the workplace: Developing and validating a multidimensional workplace curiosity scale in United States and German employees. *Personality and Individual Differences* 155 (2020), 109717. doi:10.1016/j.paid.2019.109717

[67] Todd B. Kashdan, Paul Rose, and Frank D. Fincham. 2004. Curiosity and Exploration: Facilitating Positive Subjective Experiences and Personal Growth Opportunities. *Journal of Personality Assessment* 82, 3 (2004), 291–305. doi:10.1207/s15327752jpa8203_05

[68] Todd B. Kashdan, Ryne A. Sherman, Jessica Yarbro, and David C. Funder. 2013. How Are Curious People Viewed and How Do They Behave in Social Situations? From the Perspectives of Self, Friends, Parents, and Unacquainted Observers. 81, 2 (2013), 142–154. doi:10.1111/j.1467-6494.2012.00796.x

[69] Todd B Kashdan and Paul J Silvia. 2009. *Curiosity and interest: The benefits of thriving on novelty and challenge*. Vol. 2. 367–374.

[70] Todd B. Kashdan and Paul J. Silvia. 2009. Curiosity and Interest: The Benefits of Thriving on Novelty and Challenge. In *The Oxford Handbook of Positive Psychology*. Oxford University Press, 366–374. doi:10.1093/oxfordhb/9780195187243.013.0034

[71] Todd B. Kashdan, Melissa C. Stiksma, David J. Disabato, Patrick E. McKnight, John Bekier, Joel Kaji, and Rachel Lazarus. 2018. The five-dimensional curiosity scale: Capturing the bandwidth of curiosity and identifying four unique subgroups of curious people. *Journal of Research in Personality* 73 (2018), 130–149. doi:10.1016/j.jrp.2017.11.011

[72] Alex Ketchum. 2023. Kitchen Bytes: The Retrofuturism of Kitchen Computers and Robots. *Gastronomica* 23, 1 (02 2023), 83–93. doi:10.1525/gfc.2023.23.1.83

[73] Georges E. Khalil, Karen S. Calabro, and Alexander V. Prokhorov. 2018. Development and initial testing of the brief adolescent smoking curiosity scale (ASCOS). *Addictive behaviors* 78 (2018), 67–73. doi:10.1016/j.addbeh.2017.11.008

[74] Mumtaz Khan, Shah Khusro, Iftikhar Alam, Shaukat Ali, and Inayat Khan. 2022. Perspectives on the Design, Challenges, and Evaluation of Smart TV User Interfaces. *Scientific Programming* 2022, 1 (2022), 2775959. doi:10.1155/2022/2775959

[75] Celeste Kidd and Benjamin Y. Hayden. 2015. The Psychology and Neuroscience of Curiosity. *Neuron* 88, 3 (2015), 449–460. doi:10.1016/j.neuron.2015.09.010

[76] Todd Kulesza, Margaret Burnett, Weng-Keen Wong, and Simone Stumpf. 2015. Principles of Explanatory Debugging to Personalize Interactive Machine Learning. In *Proceedings of the 20th International Conference on Intelligent User Interfaces* (Atlanta Georgia USA, 2015-03-18). Acm, 126–137. doi:10.1145/2678025.2701399

[77] Nancy Lankton, D. McKnight, and John Tripp. 2015. Technology, Humanness, and Trust: Rethinking Trust in Technology. *Journal of the Association for Information Systems* 16 (10 2015), 880–918. doi:10.17705/1jais.00411

[78] Nayen Lee, Vanessa Lazaro, Jinjing Jenny Wang, Hilal H. Şen, and Kelsey Lucca. 2022. Exploring individual differences in infants' looking preferences for impossible events: The Early Multidimensional Curiosity Scale. *Frontiers in psychology* 13 (2022), 1015649. doi:10.3389/fpsyg.2022.1015649

[79] Nancy H. Leonard and Michael Harvey. 2007. The Trait of Curiosity as a Predictor of Emotional Intelligence. 37, 7 (2007), 1545–1561. doi:10.1111/j.1559-1816.2007.00226.x

[80] Jan Leusmann, Jannik Wiese, Moritz Ziarko, and Sven Mayer. 2023. Investigating Opportunities for Active Smart Assistants to Initiate Interactions With Users. In *Proceedings of the 22nd International Conference on Mobile and Ubiquitous Multimedia* (Vienna, Austria) *(MUM '23)*. Association for Computing Machinery, New York, NY, USA, 495–498. doi:10.1145/3626705.3631787

[81] Malin K. Lilley, Amber J. de Vere, Deirdre B. Yeater, and Stan Abraham Kuczaj. 2018. Characterizing Curiosity-Related Behavior in Bottlenose (Tursiops truncatus) and Roughtoothed (Steno bredanensis) Dolphins. *International Journal of Comparative Psychology* 31 (2018). https://api.semanticscholar.org/CorpusID:54786804

[82] Emily G Liquin and Tania Lombrozo. 2020. Explanation-Seeking Curiosity in Childhood. 35 (2020), 14–20. doi:10.1016/j.cobeha.2020.05.012

[83] Jordan Litman. 2005. Curiosity and the pleasures of learning: Wanting and liking new information. *Cognition and Emotion* 19, 6 (2005), 793–814. doi:10.1080/02699930541000101

[84] Jordan Litman. 2019. Curiosity: Nature, Dimensionality, and Determinants. In *The Cambridge Handbook of Motivation and Learning*. Cambridge University Press, 418–442. doi:10.1017/9781316823279.019

[85] Jordan Litman, Tiffany Hutchins, and Ryan Russon. 2005. Epistemic Curiosity, Feeling-of-Knowing, and Exploratory Behaviour. *Cognition & Emotion* 19, 4 (June 2005), 559–582. doi:10.1080/02699930441000427

[86] Jordan A. Litman. 2008. Interest and deprivation factors of epistemic curiosity. *Personality and Individual Differences* 44, 7 (2008), 1585–1595. doi:10.1016/j.paid.2008.01.014

[87] Jordan A. Litman. 2012. *Epistemic Curiosity*. Springer US, Boston, MA, 1162–1165. doi:10.1007/978-1-4419-1428-6_1645

[88] Jordan A. Litman, Robert P. Collins, and Charles D. Spielberger. 2005. The nature and measurement of sensory curiosity. *Personality and Individual Differences* 39, 6 (2005), 1123–1133. doi:10.1016/j.paid.2005.05.001

[89] Jordan A. Litman and Tiffany L. Jimerson. 2004. The Measurement of Curiosity As a Feeling of Deprivation. *Journal of Personality Assessment* 82, 2 (2004), 147–157. doi:10.1207/s15327752jpa8202_3

[90] Jordan A. Litman and Paul J. Silvia. 2006. The latent structure of trait curiosity: evidence for interest and deprivation curiosity dimensions. *Journal of personality assessment* 86, 3 (2006), 318–328. doi:10.1207/s15327752jpa8603{_}07

[91] Jordan A. Litman and Charles D. Spielberger. 2003. Measuring Epistemic Curiosity and Its Diversive and Specific Components. *Journal of Personality Assessment* 80, 1 (2003), 75–86. doi:10.1207/s15327752jpa8001_16

[92] George Loewenstein. 1994. The psychology of curiosity: A review and reinterpretation. *Psychological bulletin* 116, 1 (1994), 75. doi:10.1037/0033-2909.116.1.75

[93] Tania Lombrozo and Emily G. Liquin. 2023. Explanation Is Effective Because It Is Selective. 32, 3 (2023), 212–219. doi:10.1177/09637214231156106

[94] Jonas Lowgren and Erik Stolterman. 2019. *Thoughtful Interaction Design - A Design Perspective on Information Technology*. The MIT Press, Cambridge.

[95] Rosa E.S. Maessen, J. Micah Prendergast, and Luka Peternel. 2023. Robotic Skill Mutation in Robot-to-Robot Propagation During a Physically Collaborative Sawing Task. *IEEE Robotics and Automation Letters* 8, 10 (2023), 6483–6490. doi:10.1109/lra.2023.3307289

[96] Amanda Markey and George Loewenstein. 2014. Curiosity. In *International handbook of emotions in education*. Routledge/Taylor & Francis Group, New York, NY, US, 228–245. doi:10.4324/9780203148211

[97] Justin Matejka, Michael Glueck, Tovi Grossman, and George Fitzmaurice. 2016. The Effect of Visual Appearance on the Performance of Continuous Sliders and Visual Analogue Scales. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2016-05-07) *(CHI '16)*. Association for Computing Machinery, 5421–5432. doi:10.1145/2858036.2858063

[98] D. Betsy McCoach, Robert K. Gable, and John P. Madura. 2013. *Instrument Development in the Affective Domain: School and Corporate Applications*. Springer New York, New York, NY. doi:10.1007/978-1-4614-7135-6

[99] Lisa McNary. 2024. Curiosity: A conceptual re-analysis for improved measurement. *Current Psychology* 43, 1 (2024), 575–586. doi:10.1007/s12144-022-04170-z

[100] Robert Cudeck Michael W. Browne. 1993. *Alternative ways of assessing model fit*. Sage, 136–162.

[101] Nature. 2022. Growth in AI and robotics research accelerates. *Nature* 610, 7931 (2022), S9–s9. doi:10.1038/d41586-022-03210-9

[102] Elizabeth Neal. 1970. A Review of Epistemic Curiosity and Behavior. *Educ Leadership* (1970).

[103] Heramb Nemlekar, Neel Dhanaraj, Angelos Guan, Satyandra K. Gupta, and Stefanos Nikolaidis. 2023. Transfer Learning of Human Preferences for Proactive Robot Assistance in Assembly Tasks. In *Proceedings of the 2023 ACM/IEEE International Conference on Human-Robot Interaction* (Stockholm, Sweden) *(HRI '23)*. Association for Computing Machinery, New York, NY, USA, 575–583. doi:10.1145/3568162.3576965

[104] Sao Mai Nguyen, Serena Ivaldi, Natalia Lyubova, Alain Droniou, Damien Gérardeaux-Viret, David Filliat, Vincent Padois, Olivier Sigaud, and Pierre-Yves Oudeyer. 2013. Learning to recognize objects through curiosity-driven manipulation with the iCub humanoid robot. In *2013 IEEE Third Joint International Conference on Development and Learning and Epigenetic Robotics (ICDL)*. 1–8. doi:10.1109/DevLrn.2013.6652525

[105] Jeesun Oh, Wooseok Kim, Sungbae Kim, Hyeonjeong Im, and Sangsu Lee. 2024. Better to Ask Than Assume: Proactive Voice Assistants' Communication Strategies That Respect User Agency in a Smart Home Environment. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 846, 17 pages. doi:10.1145/3613904.3642193

[106] Raquel Oliveira, Patrícia Arriaga, Minja Axelsson, and Ana Paiva. 2021. Humor–Robot Interaction: A Scoping Review of the Literature and Future Directions. *International Journal of Social Robotics* 13, 6 (2021), 1369–1383. doi:10.1007/s12369-020-00727-9

[107] Pierre-Yves Oudeyer, Frdric Kaplan, and Verena V Hafner. 2007. Intrinsic motivation systems for autonomous mental development. *IEEE Transactions on Evolutionary Computation* 11, 2 (2007), 265–286. doi:10.1109/tevc.2006.890271

[108] Seong-Hee Park, Jae-Pil Ha, and Daniel Mahony. 2014. Development and Validation of a Measure of Sport Fans' Specific Curiosity. *Journal of Sport Management* 28, 6 (2014), 621–632. doi:10.1123/jsm.2013-0198

[109] Seong Hee Park, Daniel F. Mahony, and T. Christopher Greenwell. 2010. The Measurement of Sport Fan Exploratory Curiosity. *Journal of Sport Management* 24, 4 (2010), 434–455. doi:10.1123/jsm.24.4.434

[110] P. Ivan Pavlov. 2010. Conditioned reflexes: An investigation of the physiological activity of the cerebral cortex. *Annals of neurosciences* 17, 3 (2010), 136–141. doi:10.5214/ans.0972-7531.1017309

[111] Pamela H Pearson. 1970. Relationships between global and specified measures of novelty seeking. *Journal of Consulting and Clinical Psychology* 34, 2 (1970), 199. doi:10.1037/h0029010

[112] André Pereira, Iolanda Leite, Samuel Mascarenhas, Carlos Martinho, and Ana Paiva. 2011. Using Empathy to Improve Human-Robot Relationships. In *Human-Robot Personal Relationships*. Springer Berlin Heidelberg, Berlin, Heidelberg, 130–138. doi:10.1007/978-3-642-19385-9_17

[113] T. X. N. Pham, Kotaro Hayashi, Christian Becker-Asano, S. Lacher, and Ikuo Mizuuchi. 2017. Evaluating the usability and users' acceptance of a kitchen assistant robot in household environment. In *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (Ro-man'17)*. 987–992. doi:10.1109/roman.2017.8172423

[114] Richard Phillips. 2016. Curious about Others: Relational and Empathetic Curiosity for Diverse Societies. 88, 88 (2016), 123–142. doi:10.3898/newf.88.02.2016

[115] Jessica Taylor Piotrowski, Jordan A. Litman, and Patti Valkenburg. 2014. Measuring Epistemic Curiosity in Young Children. *Infant and Child Development* 23, 5 (2014), 542–553. doi:10.1002/icd.1847

[116] Plutarch. 1939. *Moralia VI*. Harvard University Press, Cambridge, Mass. 489 pages. Translated by Helmbold, W. C..

[117] Denise F. Polit and Cheryl Tatano Beck. 2006. The content validity index: Are you sure you know what's being reported? critique and recommendations. *Research in Nursing & Health* 29, 5 (2006), 489–497. doi:10.1002/nur.20147 arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/nur.20147

[118] Tim Post and Juliette H. Walma van der Molen. 2019. Development and validation of a questionnaire to measure primary school children's images of and attitudes towards curiosity (the CIAC questionnaire). *Motivation and Emotion* 43, 1 (01 Feb 2019), 159–178. doi:10.1007/s11031-018-9728-9

[119] Erwin Prassler, Arno Ritter, Christoph Schaeffer, and Paolo Fiorini. 2000. A Short History of Cleaning Robots. *Autonomous Robots* 9, 3 (2000), 211–226. doi:10.1023/a:1008974515925

[120] Viva Sarah Press and Hadas Erel. 2023. Humorous Robotic Behavior as a New Approach to Mitigating Social Awkwardness. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 427, 16 pages. doi:10.1145/3544548.3580821

[121] Piotr Próchniak. 2017. Development and testing of the elements of the Nature Curiosity Scale. *Social Behavior and Personality: an international journal* 45, 8 (2017), 1245–1254. doi:10.2224/sbp.6130

[122] Piotr Próchniak and Adam Ossowski. 2023. Development and Validation of the Curiosity of Climate Changes Scale. *Psychology research and behavior management* 16 (2023), 4829–4838. doi:10.2147/prbm.s425867

[123] Thomas Reio Jr., Joseph Petrosko, Albert Wiswell, and Juthamas Thongsukmag. 2006. The Measurement and Conceptualization of Curiosity. *The Journal of Genetic Psychology* 167, 2 (June 2006), 117–135. doi:10.3200/gntp.167.2.117-135

[124] Ulf-Dietrich Reips and Frederik Funke. 2008. Interval-Level Measurement with Visual Analogue Scales in Internet-based Research: VAS Generator. 40, 3 (2008), 699–704. doi:10.3758/brm.40.3.699

[125] Steven Reiss and Maggi M. Reiss. 2004. Curiosity and Mental Retardation: Beyond IQ. 42, 1 (2004), 77–81. doi:10.1352/0047-6765(2004)42<77:camrbi>2.0.co;2

[126] Britta Renner. 2006. Curiosity about people: the development of a social curiosity measure in adults. *Journal of personality assessment* 87, 3 (2006), 305–316. doi:10.1207/s15327752jpa8703_11

[127] William Revelle. 2024. Psych: Procedures for Psychological, Psychometric, and Personality Research. https://cran.r-project.org/web/packages/psych/index.html

[128] William Revelle and David M. Condon. 2019. Reliability from α to ω: A Tutorial. 31, 12 (2019), 1395–1411. doi:10.1037/pas0000754

[129] Yves Rosseel. 2012. lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software* 48, 2 (2012), 1–36. doi:10.18637/jss.v048.i02

[130] Valentin Rousson, Theo Gasser, and Burkhardt Seifert. 2002. Assessing Intrarater, Interrater and Test–Retest Reliability of Continuous Measurements. *Statistics in Medicine* 21, 22 (Nov. 2002), 3431–3446. doi:10.1002/sim.1253

[131] Richard M. Ryan and Edward L. Deci. 2000. Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions. *Contemporary Educational Psychology* 25, 1 (Jan. 2000), 54–67. doi:10.1006/ceps.1999.1020

[132] Mikko Rönkkö and Eunseong Cho. 2022. An Updated Guideline for Assessing Discriminant Validity. 25, 1 (2022), 6–14. doi:10.1177/1094428120968614

[133] Dorsa Sadigh, S. Shankar Sastry, Sanjit A. Seshia, and Anca Dragan. 2016. Information Gathering Actions over Human Internal State. In *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (Daejeon, South Korea, 2016-10). IEEE, 66–73. doi:10.1109/iros.2016.7759036

[134] Murat Doğan Şahin. 2021. Effect of Item Order on Certain Psychometric Properties: A Demonstration on a Cyberloafing Scale. *Frontiers in Psychology* 12 (Jan. 2021), 590545. doi:10.3389/fpsyg.2021.590545

[135] Michiko Sakaki, Alexandr Ten, Hannah Stone, and Kou Murayama. 2024. Role of Metacognitive Confidence Judgments in Curiosity: Different Effects of Confidence on Curiosity Across Epistemic and Perceptual Domains. *Cognitive Science* 48, 6 (June 2024), e13474. doi:10.1111/cogs.13474

[136] Jeff Sauro and James R. Lewis. 2011. When designing usability questionnaires, does it hurt to be positive?. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vancouver, BC, Canada) *(CHI '11)*. Association for Computing Machinery, New York, NY, USA, 2215–2224. doi:10.1145/1978942.1979266

[137] Juergen Schmidhuber. 1991. Curious model-building control systems. In *Proceedings 1991 IEEE International Joint Conference on Neural Networks*. IEEE, 1458–1463 vol.2. doi:10.1109/ijcnn.1991.170605

[138] Matthias Schmidmaier, Jonathan Rupp, Darina Cvetanova, and Sven Mayer. 2024. Perceived Empathy of Technology Scale (PETS): Measuring Empathy of Systems Toward the User. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (2024-05-11) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA. doi:10.1145/3613904.3642035

[139] Coltan Scrivner. 2021. The psychology of morbid curiosity: Development and initial validation of the morbid curiosity scale. *Personality and Individual Differences* 183 (2021), 111139. doi:10.1016/j.paid.2021.111139

[140] Ruyu Sheng, Yanqing Wang, and Longfei Huang. 2021. Intelligent Service Robot for High-Speed Railway Passengers. In *Data Science*. Springer Singapore, Singapore, 263–271.

[141] Dajung Diane Shin and Sung-il Kim. 2019. Homo Curious: Curious or Interested? 31, 4 (2019), 853–874. doi:10.1007/s10648-019-09497-x

[142] Dajung Diane Shin, Yoonah Park, Minhye Lee, Sung-il Kim, and Mimi Bong. 2023. Are curiosity and situational interest different? Exploring distinct antecedents and consequences. *British Journal of Educational Psychology* 93, 4 (2023), 1207–1223. doi:10.1111/bjep.12627

[143] Masahiro Shiomi, Takayuki Kanda, Iris Howley, Kotaro Hayashi, and Norihiro Hagita. 2015. Can a Social Robot Stimulate Science Curiosity in Classrooms? *International Journal of Social Robotics* 7, 5 (01 Nov 2015), 641–652. doi:10.1007/s12369-015-0303-1

[144] Martín Solís. 2015. The Dilemma of Combining Positive and Negative Items in Scales. 27.2 (2015), 192–199. doi:10.7334/psicothema2014.266

[145] Christiane Attig Thomas Franke and Daniel Wessel. 2019. A Personal Resource for Technology Interaction: Development and Validation of the Affinity for Technology Interaction (ATI) Scale. *International Journal of Human–Computer Interaction* 35, 6 (2019), 456–467. doi:10.1080/10447318.2018.1456150

[146] Roger Tourangeau, Eleanor Singer, and Stanley Presser. 2003. Context Effects in Attitude Surveys: Effects on Remote Items and Impact on Predictive Validity. *Sociological Methods & Research* 31, 4 (May 2003), 486–513. doi:10.1177/0049124103251950

[147] Li tze Hu and Peter M. Bentler. 1999. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal* 6, 1 (1999), 1–55. doi:10.1080/10705519909540118

[148] Daniel Ullrich. 2017. Robot Personality Insights. Designing Suitable Robot Personalities for Different Domains. *i-com* 16, 1 (2017), 57–67. doi:10.1515/icom-2017-0003

[149] Jo Vermeulen, Jonathan Slenders, Kris Luyten, and Karin Coninx. 2009. I Bet You Look Good on the Wall: Making the Invisible Computer Visible. In *Ambient Intelligence*. Vol. 5859. Springer Berlin Heidelberg, 196–205. doi:10.1007/978-3-642-05408-2_24

[150] D. C. Vidler and R. A. Hansen. 1978. Factor analysis of a scale of academic curiosity. *Journal of Clinical Psychology* 34 (1978), 386–387. doi:10.1002/1097-4679(197804)34:2<386::aid-jclp2270340225>3.0.co;2-y

[151] Sarah Theres Völkel, Ramona Schoedel, Lale Kaya, and Sven Mayer. 2022. User Perceptions of Extraversion in Chatbots after Repeated Use. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 253, 18 pages. doi:10.1145/3491102.3502058

[152] Nick Walker, Kevin Weatherwax, Julian Allchin, Leila Takayama, and Maya Cakmak. 2020. Human Perceptions of a Curious Robot That Performs Off-Task Actions. In *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*. Acm, Cambridge United Kingdom, 529–538. doi:10.1145/3319502.3374821

[153] Dakuo Wang, Elizabeth Churchill, Pattie Maes, Xiangmin Fan, Ben Shneiderman, Yuanchun Shi, and Qianying Wang. 2020. From Human-Human Collaboration to Human-AI Collaboration: Designing AI Systems That Can Work Together with People. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI EA '20)*. Association for Computing Machinery, New York, NY, USA, 1–6. doi:10.1145/3334480.3381069

[154] Jennifer L. Weible and Heather Toomey Zimmerman. 2016. Science curiosity in learning environments: Developing an attitudinal scale for research in schools,

homes, museums, and the community. *International Journal of Science Education* 38, 8 (2016), 1235–1255. doi:10.1080/09500693.2016.1186853

[155] Bert Weijters and Hans Baumgartner. 2012. Misresponse to Reversed and Negated Items in Surveys: A Review. 49, 5 (2012), 737–747. doi:10.1509/jmr.11.0368

[156] Kassandra Wellendorf, Karen Louise Grova Søilen, and Kristin Veel. 2022. Calm Surveillance in the Leaky Home: Living with a Robot Vacuum Cleaner. *The Journal of Media Art Study and Theory* 3, 1 (2022), 41–62.

[157] William M. Whitecross and Michael Smithson. 2023. Curiously Different: Interest-curiosity and Deprivation-Curiosity May Have Distinct Benefits and Drawbacks. 213 (2023), 112310. doi:10.1016/j.paid.2023.112310

[158] Andreas Wieland, Christian Durach, Joakim Kembro, and Horst Treiblmaier. 2017. Statistical and judgmental criteria for scale purification. *Supply Chain Management* 22 (08 2017). doi:10.1108/scm-07-2016-0230

[159] Andreas Wieland, Florian Kock, and Alexander Josiassen. 2018. Scale Purification: State-of-the-art Review and Guidelines. *International Journal of Contemporary Hospitality Management* 30, 11 (2018), 3346–3362. doi:10.1108/ijchm-11-2017-0740

[160] Jason Wu, Chris Harrison, Jeffrey P. Bigham, and Gierad Laput. 2020. Automated Class Discovery and One-Shot Interactions for Acoustic Activity Recognition. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–14. doi:10.1145/3313831.3376875

[161] Muhamad Saiful Bahri Yusoff. 2019. ABC of Content Validation and Content Validity Index Calculation. *Education in medicine journal* 11, 2 (2019), 49–54. doi:10.21315/eimj2019.11.2.6

[162] Jiang Zhang, Yufeng Wang, Zhiyuan Yuan, and Qun Jin. 2020. Personalized real-time movie recommendation system: Practical prototype and evaluation. *Tsinghua Science and Technology* 25, 2 (2020), 180–191. doi:10.26599/tst.2018.9010118

[163] Michal Zion and Irit Sadeh. 2007. Curiosity and Open Inquiry Learning. 41, 4 (2007), 162–169. doi:10.1080/00219266.2007.9656092

[164] Perry Zurn and Danielle S. Bassett. 2018. On Curiosity: A Fundamental Aspect of Personality, a Practice of Network Growth. *Personality Neuroscience* 1 (2018), e13. doi:10.1017/pen.2018.3

A The Perceived System Curiosity (PSC) Scale Usage Instructions

Table 10: The final 3-factor, 12-item PSC scale, which is used to measure perceived system curiosity. Each item is answered on a 101-point VAS slider ranging from (0) *strongly disagree* to (100) *strongly agree.*

ID	Item
Perceived Explorative Curiosity (PSC_E) $\alpha = .89$	
PSC_1	The system enjoys novel experiences.
PSC_2	The system is driven towards novel sensations.
PSC_3	The system is interested in interacting with novel objects.
PSC_4	The system is excited by unpredictable tasks.
PSC_5	The system likes to find out how things work.
Perceived Investigative Curiosity (PSC_I) $\alpha = .82$	
PSC_6	The system invests substantial time to collect new knowledge.
PSC_7	The system likes to investigate.
PSC_8	The system is inquisitive.
Perceived Social Curiosity (PSC_S) $\alpha = .85$	
PSC_9	The system likes finding out why others behave the way they do.
PSC_{10}	The system wants to know what others are doing.
PSC_{11}	The system likes listening to others' conversations.
PSC_{12}	The system is interested in learning more about people.

(1) Use 101-point Visual Analog Sliders (VAS) ranging from 0 to 100, with the anchors *strongly disagree* (0) and *strongly agree* (100).
(2) We recommend using and reporting all three subscales (PSC_E, PSC_I, PSC_S) and the total scale value (PSC) when measuring perceived system curiosity with the PSC to measure the construct of perceived curiosity; however, each subscale can also be used independently to then only measure perceived explorative curiosity, perceived investigative curiosity, or perceived social curiosity.
(3) When evaluating the questionnaire, we always displayed the questions in random order. Thus, we suggest also randomizing the order of questions across all subscales when using the PSC.
(4) Each subscale score is derived by calculating the arithmetic mean of all items corresponding to that subscale. The total scale score is derived by calculating the arithmetic mean of all items. There are no reverse-scored items. Higher scores indicate higher perceived curiosity.

$$
\begin{aligned}
PSC_E &= (PSC_1 + PSC_2 + PSC_3 + PSC_4 + PSC_5) \;/\; 5 \\
PSC_I &= (PSC_6 + PSC_7 + PSC_8) \;/\; 3 \\
PSC_S &= (PSC_9 + PSC_{10} + PSC_{11} + PSC_{12}) \;/\; 4 \\
PSC &= (PSC_E + PSC_I + PSC_S) \;/\; 3
\end{aligned}
$$