International Journal of Human–Computer Interaction

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/hihc20

Investigating the Crowd's Creativity for Creating On-Demand IoT Scenarios

Tahir Abbas, Vassilis-Javed Khan & Panos Markopoulos

INTERNATIONAL JOURNAL OF HUMAN–COMPUTER INTERACTION
2020, VOL. 36, NO. 11, 1022–1049
https://doi.org/10.1080/10447318.2019.1709331

Taylor & Francis
Taylor & Francis Group

Investigating the Crowd's Creativity for Creating On-Demand IoT Scenarios

Tahir Abbas[a,b], Vassilis-Javed Khan[a], and Panos Markopoulos[a]

[a]Department of Industrial Design, Eindhoven University of Technology, Eindhoven, The Netherlands; [b]Department of Software Engineering, Mirpur University of Science & Technology (MUST), Mirpur, Pakistan

ABSTRACT

The IoT industry supplies a plethora of Internet connected devices and services supporting smart home automation. However, end-users having little knowledge of the features and possibilities of such technologies, face difficulties in conjuring up useful application scenarios combining such devices and services, thus missing out on potential applications outside those provided by vendors. A remedy for such end-users can potentially be found in crowdsourcing IoT scenario creation. For such an enterprise to be viable it is essential to assess whether crowdsourcing can result in practical and original scenarios. This article reports two studies aiming to establish the practicality and originality of crowdsourced IoT scenarios for smart homes. In the first study, we recruited 102 crowd workers who created 306 scenarios in various categories. We then recruited a second cohort of 620 crowd workers to rate the scenarios' creativity. In the second study, we evaluated the corpus of IoT scenarios by 20-experienced smart home users recruited through a screening survey. Our results show that the crowd evaluations of originality and creativity are strongly correlated with those of smart home users. Our major IoT-specific findings in relation to creativity are: a) The number of IoT devices and the number of combination of devices impact how creative the scenarios are perceived; b) Workers with self-reported intermediate programming knowledge wrote more creative scenarios when compared to workers having expert knowledge; c) Computational metrics such as text metrics can provide the basis for automated assessment of the scenarios' creativity. Finally, an inductive thematic analysis of the scenarios revealed interesting themes (e.g., types of rules, automation styles and novel operators) which can serve as a guide for designing more expressive and intuitive end-user development solutions, in the context of IoT.

1. Introduction

Developments in IoT are increasingly enabling the realization of "smart home" scenarios for different purposes including health and wellness, cooking, elderly care, and communication (Mennicken & Huang, 2012). Efforts to ensure the interoperability of different devices and services, have resulted in end-user development services such as IFTTT, Zapier, Apiant – to name a few, which allow end-users to create configurations and applications fitting their own personal needs. For example, in IFTTT, end-users make "recipes" -as they are known- to program devices and applications by defining triggers to launch certain actions. These technological developments have opened up a design space that is constantly growing by programming various combinations of trigger and action devices (Ur et al., 2016).

To empower end-users to compose Internet of Things (IoT) applications, researchers have developed End-User Development (EUD) environments where less tech-savvy users can develop IoT applications by combining the functionalities of devices and services in a lower level of abstraction. The most commonly used EUD techniques are object composition (Desolda, Ardito, & Matera, 2017), process-oriented or rule-based (Brich, Walch, Rietzler, Weber, & Schaub, 2017), jigsaw puzzle (Danado & Paterno, 2015), interaction by demonstration (Chen, He, Anderson, Keller, & Skubic, 2006), mark-up languages and ontologies to formulate and comprehend complex logical expressions (Metaxas & Markopoulos, 2017). These lower-level abstractions prevent users from thinking of valuable applications to exploit the full range of capabilities of smart home devices and thereby create gaps between user expectations and reality; consequently halting acceptance (Clark, Newman, & Dutta, 2017).

More recently, EUD techniques targeting "higher level" abstractions have emerged which allow users to define rules without knowing vendor-specific behaviors of devices (e.g., "turn the light on" is offered by all lamps irrespective of their vendors) (Corno, De Russis, & Monge Roffarello, 2019). These higher-level approaches have focused more on making EUD accessible through natural language. Nonetheless, prior works have shown that end-users face significant challenges much before defining rules even when planning for getting IoT infrastructure. For instance: (1) end-users in IoT face significant barriers before even getting to the task of programming triggers and actions (Jakobi, Ogonowski, Castelli, Stevens, & Wulf, 2017); (2) smart home users have very little

CONTACT Tahir Abbas t.abbas@tue.nl Industrial Design Dept. (Future Everyday group) Eindhoven University of Technology, Eindhoven, the Netherlands Atlas, PO Box 513, 5600 MB Eindhoven, The Netherlands.

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/hihc.

knowledge of the various features of smart home devices and their potential for combination (Jakobi et al., 2017); (3) information provided by vendors on their websites does not help end-users in translating their high-level goals to useful scenarios (Mennicken & Huang, 2012) who consequently find it difficult to conceive combinations of devices and services that will suite their context and needs. Finally, as the number of devices and services increases, the potential combinations of those in a smart home context increases exponentially, which suggests that this challenge is bound to increase further and hamper the uptake and utilization of end-user development and customization of IoT technologies (Mennicken, Vermeulen, & Huang, 2014).

To support end-users in planning and preparation phase and to exploit useful combination between devices and services, researchers have developed participatory design techniques to involve non-experts in idea generation. These techniques usually use card-based tools to foster out-of-box thinking for generating novel solutions combining several devices and services. For example, Tiles card (Mora, Gianni, & Divitini, 2017) employs workshop techniques to foster creative thinking in novices for augmented objects in IoT. These participatory design techniques have been used successfully in the past to solve the aforementioned problems, but they do not provide any specific guidance for providing consumer services directly to end-users. For example, we imagine a system that can propose several use case scenarios to smart home users when they explore and select devices or services from the repository of available devices. Through this exploration phase, end-users will get several ideas to address their own contexts and needs. One can generate several predefined IoT scenarios through the aforementioned participatory design techniques by running several workshops in a physical setting with non-experts. But, the problem is that they are not scalable and the whole process is quite resource intensive (e.g., hiring of participants and moderators, materials and space) and time consuming. Furthermore, some of those techniques focus on tech professionals rather than non-experts (e.g., IoT Service Kit (Brito & Houghton, n.d.).

Recently, researchers have leveraged crowd creativity as a method to generate and evaluate creative content in different domains. Examples include but are not limited to: sketches of chairs (Yu, Kraut, & Kittur, 2016; Yu & Nickerson, 2011), feedback on posters (Luther et al., 2015), user interfaces of interactive systems (Lasecki et al., 2015) and ideating by collaborating with synchronous group of workers (Andolina et al., 2017); These crowd-based ideation techniques allow for both gathering ideas from crowd workers by showing them diverse examples (Siangliulue, Chan, Gajos, & Dow, 2015), and assessing the creativity of individual ideas (Yu & Nickerson, 2011). In this paper, we leveraged crowd creativity as an established methodology to generate IoT scenarios. An IoT-specific advantage for leveraging crowd creativity is hiring workers on-demand (24 x 7) with specific technical expertise (e.g., programming skills or prior IoT experience). Furthermore, using crowdsourcing as an ideation tool for IoT has several benefits that would otherwise difficult to attain; (1) *Scalability in Ideation & Evaluation*: one can engage hundreds of participants in an ideation task by organizing and presenting to them smart home devices. (2) *Variety of Ideas*: crowdsourcing helps to involve people from different backgrounds, cultures and education levels to generate a diversity and variety of IoT scenarios addressing the context and needs of smart home inhabitants in different parts of the world.

Our study builds upon the prior work done in personalization in the context of IoT (Clark et al., 2017; Ur, McManus, Pak Yong Ho, & Littman, 2014). For instance, Clark et al. (2017) studied the effect of personification–with agent-mediated devices and data and unmediated devices and data–on end-user's mental model. Amazon Mechanical Turk (MTurk) workers were shown a list of smart home sensors, actuators and online services and then they were asked to imagine what kind of application they would desire in their homes (Clark et al., 2017). Their work differs significantly from our work in that they studied those scenarios with the lens of mental models and showed that different abstractions (agent-mediated and unmediated devices and data) have different priming effect on user's mental models. In another study, Ur et al. (2014) asked MTurk workers to create five applications by picturing themselves in a smart home equipped with different IoT devices (Ur et al., 2014). Nevertheless, the main focus of their study was whether inexperienced users can express applications in trigger action programming style. Most of their participants (62.6%) described scenarios in a programming style and 77.9% of created IoT applications typically used one-to-one mapping (combination of one-trigger device with one-action device). This shows that creating a program that involves multiple triggers and actions is still challenging.

The focus on IoT necessarily differentiates our work with prior work in subtle yet salient ways. First, IoT scenarios must include input and output devices. In our study, we find that the number of devices and the number of combination of devices impacts how creative the ideas are perceived. Second, IoT implies some programming or end-user programming (e.g., of setting up a device or interoperating several devices). In our study, we find that workers with self-reported intermediate programming knowledge wrote more creative scenarios. Finally, the IoT industry is already a multi-billion industry and is expected to further grow. According to a recent survey, "The European smart home market is expected to grow from USD 22.8 billion in 2018 to USD 44.0 billion by 2024 at a Compound Annual Growth Rate (CAGR) of 11.58%" European2024 (European Smart Home Market Size, Growth, Trend and Forecast to 2024, (n.d.)). Therefore, we expect our study to be particularly interesting to IoT companies. Furthermore, we analyze in-depth text parameters of the crowd-created scenarios for informing the design of future systems. More specifically, based on our analysis, future systems can automatically evaluate the quality of IoT scenarios in real-time and can adapt the interface of crowd workers to better support them in the creative task, e.g., by nudging them to combine devices, or by proposing replacing common words with unique ones. Finally, our in-depth qualitative analysis of the scenarios revealed some novel features which can be used by IoT designers to design more novel and natural interfaces for end-users.

Practically, our research can fulfill smart home owners' diverse needs by presenting them with initial smart home scenarios fit for their specific context, who would otherwise limit applications to just a few devices and services (Dixon et al., 2010). In other words, this paper has shown the potential value by addressing the challenge described in Jakobi et al. (2017):

> … We have seen people standing in [a consumer market] in front of smart home products, and I literally saw the question marks in their eyes.

Furthermore, the method of generating IoT scenarios that we present in this paper could be used either by vendors or distributors to assist smart home end-users in addressing their needs and inspiring novel use cases.

In summary, in this paper we make the following four contributions:

(1) By systematically investigating the differences between creative and ordinary IoT scenarios in terms of devices/services used, their combination and text metrics (e.g., number of words), we present computational metrics that future systems could use to automatically compute the creativity of IoT scenarios.
(2) By investigating whether workers' knowledge (e.g., programming or smart home experience) would impact the creativity of their scenarios, we find that workers with intermediate programming knowledge wrote more creative ones when compared to expert or novice programmers.
(3) By exploring crowd created IoT scenarios in detail we discover their hidden features and recurrent patterns which could inform the future design of EUD systems (e.g., novel operators, types of rules, etc.). Finally, we make public our dataset with the actual scenarios along with their meta-data to promote further research. Our dataset is available in two formats: MS Excel[1] and SPSS.[2]
(4) We triangulate the crowd-based evaluations with actual smart home users demonstrating the validity of the crowd-based evaluation.

1.1. Scenario for situating the problem domain

Let's suppose a household composed of Maria, the mother; John, the father and Alice, the daughter. They plan to renovate their home, so they visit a consumer electronics store aiming to purchase electronics for their home (left part in Figure 1). While strolling through the store's aisles, slowly but surely, they start feeling overwhelmed by the choices regarding the number of IoT devices, their capabilities, and how they can be configured or interconnected. Confused and puzzled they have difficulty in making up scenarios that would fit their purposes. This keeps them indecisive about the set of devices that they need to purchase.

Now imagine a system, which is installed in the shop (right part in Figure 1), which can assist customers to play around with different IoT devices by dragging them to the screen. When they select devices and drag them to the screen, the system presents them with IoT scenarios highlighting various categories that cover several contexts within the domestic domain. The system can recommend scenarios either from an already available corpus, which was created through crowdsourcing, or it can forward an on-demand request to crowd workers when it does not find any relevant scenarios. Through this exploration process, John's family is both able to recognize various contexts and needs and also acquire knowledge about the set of devices and various possibilities to connect them in a unique and novel way.

In the current study, we focus on what is needed to design such a system through outsourcing IoT scenarios to crowd workers. For this paper the implementation of such a system is out of our scope.

2. Background and related work

In this section we structure the related work around three main topics:

- We outline some end-user challenges for IoT that hinders the pace of creating such novel and user-centered solutions.
- We present participatory design approaches for ideation and creativity in IoT and summarize the available methods for nurturing creative solutions.
- We present crowd-sourced ideation and crowd creativity as an established methodology for creating IoT scenarios. We also briefly outline some available methods for large scale ideation when leveraging crowdsourcing.

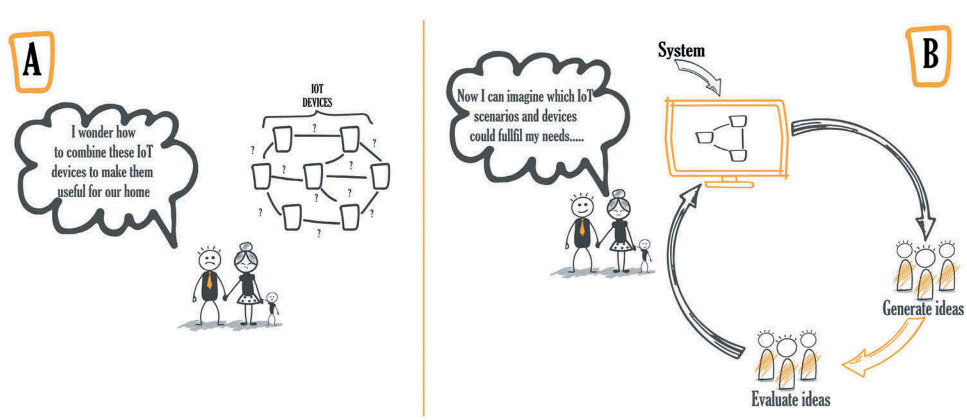

Figure 1. A: while visiting a consumer electronics shop this family has hard time understanding which IoT devices would be suitable to make their home smarter. B: The same family now uses a system in-store that can assist them to recognize various contexts and needs as well as give them ideas to connect different devices in a unique way.

2.1. End-user challenges for IoT

Considerable advances have been made in the field of End-User Programming technologies for IoT with the aim to lower the threshold for configuration of smart objects in IoT space. Most of these efforts focus on the esthetic aspects of EUD environments, i.e., how to make languages syntactically correct through natural or visual interfaces and semantically meaningful by making them transparent, intelligent and accountable. As a result, two levels of abstraction appear in the literature to empower end-users for personalizing their IoT devices: low level and high-level abstractions. First, we review some limitations of low-level abstractions including automation tools (e.g., IFTTT) and then we review some high-level abstractions. Finally, we review some challenges that have been overlooked in the past regarding IoT personalization and smart homes.

2.1.1. Lower-level abstractions

The most popular and widely used interaction paradigm for end-user configuration of smart objects is object composition. It can be defined as "synchronizing the behavior of multiple objects to create new, added-value services" (Desolda et al., 2017). The most popular IoT services like IFTTT, Zapier, Node-RED follow the object composition paradigm. There are two popular object composition models: wizard, that allows a step by step approach to define rules as offered by IFTTT; and wired which allows end-users to connect services or devices (also called nodes) through wires or lines. Desolda et al. (2017) further investigated the object composition paradigm by analyzing some task automation tools (IFTTT, Atooma, Tasker, etc.) to identify new visual notations and rules for non-programmers in IoT. Through an elicitation study, they identified some new notations by analyzing the event-condition-action rules based on the 5W model (Who, What, When, Where, Why) with end-users. They then compared the usability of three prototypes (E-wizard, E-wired and E-Free) that emerged from the elicitation study and then further validated the best prototype with smart home expert users. Their participants found E-free tool more expressive as it allowed them to define rules without following a strict order and gave them flexibility to modify or extend their rules.

Like the object composition paradigm, some process-oriented (Brich et al., 2017) approaches have emerged for programming smart objects. A process-oriented approach can be considered as multi-stage automation rule chains where multiple processes–involving different set of devices and conditions–can be joined to create more complex rules. To explore the potential of the process-oriented paradigm, Brich et al. (2017) conducted a contextual inquiry study with 18 participants in 12 households to explore the end-user needs for smart home automation for two-popular configurations: rule-based (if-this-then-that) and process-oriented. Their results indicate that the rules-based notation was ideal for simple scenarios but was not suitable for more complex configurations involving several devices due to the strictness of rules (joining triggers with operators) that were not easily comprehended by their participants. On the contrary, their participants enjoyed the process-oriented approach due to its expressiveness in terms of more options that they had, e.g., grouping of rules, line connections, temporal and hierarchical dependencies.

The jigsaw puzzle metaphor (Danado & Paterno, 2015) allows the end-user to configure compositions of services controlling smart objects, smartphones and web services. This framework is similar to the Scratch[3] programming environment that allows jigsaw to ease the development process.

Such programmatic control of IoT devices–either through object composition, jigsaw puzzle blocks or process-oriented chain rules – requires great deal of comprehending and formulating of complex logical expressions which is a challenging task for novices. Furthermore, end-users want clear explanations and accountability about context-sensitive behaviors of their IoT devices. More recently, researchers developed an ontology and corresponding mark-up language called context range semantics (CRS) (Metaxas & Markopoulos, 2017). The CRS does not only allow end-users to reason about the context but it can assist them (with the help of context range editor (CoRE)) to automatically construct correct logical expressions from the context. These pre-fabricated logical expressions can later be edited to create new logical expression with point and click interaction upon operands and operators. This way, it supports intelligibility and accountability by providing clear explanation about system behavior, supporting 'why' and 'why not' questions which can assist end-users to modify the system behaviors if required.

2.1.2. Problems with automation tools

In this section, we highlight some limitations of current automation tools. More specifically, we focus on the popular IFTTT service.

2.1.2.1. One-to-one-mapping. The first problem with popular automation tools like IFTTT is that they only support one-to-one-mapping between a single trigger device and a single action device. In other words, IFTTT currently does not support conjunction of multiple triggers and actions. For example, if a user wants to activate the Philips Hue bulb only when he arrives home after 18:00 (two triggers: location is home and time is after 18:00). In IFTTT, it is difficult to setup the recipe to capture the above scenario (Huang & Cakmak, 2015; Rahmati, Fernandes, Jung, & Prakash, 2017; Ur et al., 2014).

2.1.2.2. Duplicate recipes. Furthermore, the work of Ur et al. (2016) argued that end-users created "duplicate" recipes and not the original or new ones on the IFTTT platform because of the interface which forced them to do so (Ur et al., 2016).

2.1.2.3. IFTTT and mental model inaccuracies. A prior study (Huang & Cakmak, 2015), revealed that users were not able to distinguish correctly between instantaneous (e.g., sending an e-mail), extended (e.g., brewing coffee) and sustained actions (e.g., turning the lights on) with regard to IFTTT. Furthermore, they were also in disagreement regarding the difference between states (e.g., it is raining) and events (at 3:00 pm). These findings suggest that IFTTT does not provide any clues to their users for construction of correct recipes and does not capture the mental models of their users.

2.1.2.4. Other related tools. In a prior work (Dahl & Svendsen, 2011), authors tested the usability of three end-user composition tools in a usability lab including filter list, wiring diagram and jigsaw puzzle. The test participants were given some predefined composition tasks that they solved using three prototypes. They identified several problems with these tools. In the wiring diagram prototype, participants reported that the interface become cluttered when they add more connecting arrows between components. Besides, it was difficult to visually "trace" the connecting arrow to see which components it linked. Additionally, it was difficult to plan new compositions due to lack of overview of compositions or saved compositions (which can give clue to user). In jigsaw puzzle, it was difficult to extract information as the number of compositions increase. Some participants found it challenging to differ between trigger and response device when using the jigsaw puzzle prototype. Test participants also gave indications that it was challenging to quickly understand and extract meaning out of compositions represented as jigsaws.

In summary, these findings show that automation tools (in all composition interfaces) have usability issues and do not capture the exact mental model of end-users.

2.1.3. Higher-level abstractions

In a recent work (Corno et al., 2019), a novel way of personalizing IoT devices and services was explored. Corno et al. (2019) argued that existing task automation tools such as IFTTT do not support two well-known problems in the IoT domain: a) *abstraction*: which means devices and services which have similar functionalities but belong to different brands or vendors are treated as different or separate objects and thus could not be programmed through same set of generic rules, b) *Adaptation*: due to rapidly growing ecosystem of IoT, adding a new device or service would further increase the complexity of lower-level abstractions because one has to define new rules to personalize each new device or service. To overcome the aforementioned problems, they developed an ontological based high-level representation for EUD in the IoT called EUPont. Their higher-level semantic model enabled features to practically execute the IoT applications. The evaluation study with 30 participants demonstrated that end-users were able to create rules in higher-level abstraction correctly in less amount of time as compared to those who created rules in lower-level abstractions.

In another study (Clark et al., 2017), the authors studied the effect of priming end-users with different kinds of smart home abstractions: unmediated devices/data and agent-based devices/data. Their results indicate that different smart home abstractions effect strongly on the mental models of users which resulted in different kinds of interactions and smart home applications for IoT. They also found that priming with different abstractions affects more strongly older users and those who do not have computer science background. They also found that the Unmediated Device (described by manual one-to-one mapping between single input and output devices) which is one of the most common abstraction adopted by contemporary IoT services (e.g., IFTTT) produced the most limited kinds of interactions.

2.1.4. Challenges

Mennicken et al. (2014) discuss several challenges but also new promising directions for smart homes specifically related to user experience. With regards to complex domestic spaces, they argued that apart from solving interoperability issues between devices and services, it is important to consider how end-users can identify and configure meaningful connections of devices and services. Concerning the infinite possibilities that exist between devices, they further argued that:

The number of possible combinations of and interactions between devices are huge, and so adding a new sensor, device or robot to the home could have unforeseeable results.

Another qualitative study conducted with smart home owners and professionals, Mennicken and Huang (2012) included the notion of involving end-users in all phases of setting up smart home system – from conception to daily use. They pointed out several challenges of planning for "unfamiliar" smart home technology, but the most relevant ones to our study are:

(1) Information provided on vendor's website or in the manuals offer technical details and is not usable for less tech savvy persons. For example:

> Information about home automation technologies, such as that found on websites, brochures, or manuals, often offers technical details but is less informative about its potential effects on every-day life. At the start of the planning phase, participants reported not understanding potential benefits of technologies and therefore had difficulties prioritizing those technologies against other needs in the home

(2) People could not anticipate all future use case scenarios covering several aspects of their lives and they want flexible solutions that allow them to tailor the functionality.
(3) Less tech-savvy persons need to rely upon other experienced persons and their expertise to find useful inter-connection between devices.

An 18 month long living-lab study with 14 households equipped with smart home technology (Jakobi et al., 2017) investigated the different stages that smart home users go through. Those included system set-up (planning), installation and configuration, domestication, daily use, reconfiguration and extension. Part of their study focused on the system setup and the planning stage that was neglected in the past. One of the findings they reported regarding system setup:

When households were picking components for setting up their future smart home, it became obvious that many had very little knowledge of the various features of sensors and actors, as well as of their potential for combination …

This shows that many smart home owners are unable to think beyond simple configurations (usually involving one trigger and one action device) that are provided by vendors.

In relation to creating end-user scenarios for smart-home technology, Ogonowski, Hennes, and Seiffert (2016) developed an online system called Shop&Play.[4] Using that system, end-users can select specific requirements related to security, comfort and energy saving. Then based on those requirements, they can select some pre-defined scenarios for automation of smart home systems and then order a complete pre-configured package from the vendors. Although Shop&Play is a step in the right direction, there are still several questions that persist: Where should these pre-defined scenarios come from? How can vendors anticipate and gather useful scenarios and provide them on their websites? How can inexperienced inhabitants, who are already living in smart homes and want to enhance their homes with new devices, get useful ideas?

Based on these earlier investigations into the end-user configuration for IoT devices and services in a domestic context we wish to explore the combination of two elements:

(1) Lowering the threshold for non-tech savvy end-users to formulate configurations of IoT devices and services by using non-technical representations, such as in a scenario-based description.
(2) Expanding the range of relevant combinations of IoT devices and services that end-users can conceive themselves by crowdsourcing the creation of these scenarios.

2.2. Ideation and creativity in relation to IoT

During the past few years, some efforts have been made to put ideation and creativity at the core of design process for making user-centered and novel IoT applications.

Know Cards AspialaLearn.Cards (Aspiala, n.d.) allow people having limited background in IoT and electronics to combine several components together to create novel solutions. The main objective of Know Cards is to come up with unusual ideas by experimenting with random combination of input and output devices. In this methodology, the first step for participants is to shuffle the input (trigger devices) and output cards (action devices). Then they need to pick at least one to two input and one to two output devices and place them on the table. Next, they are required to write down the most novel and unusual scenarios by combination of those input and output cards. In the follow up group activity, they vote on the most interesting and unique ideas and assign points according to their novelty.

Our proposed approach in this paper is inspired by the Know Cards; nevertheless, we did not use physical cards to engage our participants (crowd workers), rather we used imagery of actual input and output devices and we also showed them detail descriptions of their capabilities, sample scenarios and a video to boost their inspiration before engaging them in the actual task.

Other similar techniques include the IoT Service Kit BritoIoTKit which enables domain experts to explore and physically demonstrate several IoT scenarios with cards, maps and 3D printable pieces. Similarly, the Tiles Ideation Toolkit (Mora et al., 2017) follows a process-oriented approach to foster creativity and ideation for augmented objects using design cards and a workshop technique.

Negri, Trousse, and Senach (2012) used a mixed method approach combining both a participatory design approach and a diary-based technique to foster creativity for IoT. They asked participants to generate ideas by placing probes – fake sensors and actuators–at their actual environments and then report and share their ideas with other groups going through same process.

Apart from card-based ideation toolkits, there are various other design thinking approaches for creating user-centered (Wilson, 2005) and novel solutions in IoT. For example, Fauquex, Goyal, Evequoz, and Bocchi (2015) presented a methodology where user needs and expectations drive the core of the design process. Their methodology has seven stages: discovery, capturing, research, design, prototype, evaluate and refine. They combined several conventional tools such as surveys, brainstorming, interviews, etc., and invented several other IoT-specific design considerations to build IoT solutions. The outcome of their proposed approach included well-described scenarios covering user needs and expectations in various contexts and a complete toolbox to implement and test those IoT scenarios.

There is no doubt that authors of card based ideation tools have identified the importance of the problem in IoT, its dimensions and proposed a practical way to generate creative solutions. Nevertheless, these card-based ideation tools – adapted from current participatory design techniques – do not provide any specific guidance for the challenge identified in the introduction of the paper, namely: how to provide consumer services directly to end-users. For this reason, we choose to work with crowdsourcing with the view that a crowd can be recruited through readily available platforms to provide direct services for consumers of IoT products, e.g., in the vendor's shop or over the Internet to a large-consumer audience.

2.3. Crowd creativity

Crowdsourcing is an emergent approach to collaborative work that provides easy access to the skills and talent of a large number of individuals spread geographically, to solve tasks that would otherwise be difficult or costly to solve using more traditional approaches (Mao, Capra, Harman, & Jia, 2017). The typical approach to crowdsourcing is "microtask" crowdsourcing which allows people to globally distribute "micro" tasks to various online workers (also known as crowd-workers) and let them to solve on the order of few minutes or even seconds. Microtask crowdsourcing has been used in various tasks that require human intelligence, for example, identification of objects in images, natural language processing, finding relevant information or generating creative contents. Various platforms have been developed to support microtask crowdsourcing including but not limited to: Amazon's Mechanical Turk, [5] Figure-Eight[6] (old CrowdFlower), MicroWorkers[7] etc. In this paper, we have used Amazon's Mechanical Turk as a microtask crowdsourcing platform. More recently, a notion of MacroTask crowdsourcing (Schmitz & Lykourentzou, 2018) has emerged where a task itself could not be decomposed into subtasks due to its complexity in nature. Furthermore, unlike Boolean crowdsourcing where there is only one definite answer or certain range of answers, open-ended tasks mostly don't have 'right' or 'wrong' answers akin to the scenario creation task. We believe that writing a creative scenario where each comprising of average 60 words and then judging their creativity is not a "microtask" in nature. Hence, we position this creative task under macrotask crowdsourcing.

Over the last few years, some literature has investigated the link between crowdsourcing and creativity. For example, In CrowdBoard (Andolina et al., 2017), collocated ideators develop ideas on a digital whiteboard with the help of synchronous crowd workers who also shares the whiteboard with ideators. Additionally, the crowd can hear and see the live broadcast of ideators working on a brainstorming session. Ideators can freely draw and erase on the board using IR pens and can create movable sticky notes while having a chat function enabled. Crowd workers can also create notes or comments directly on the whiteboard and can initiate a live chat with the ideators.

In a distributed analogical idea generation technique (Yu et al., 2016), authors draw inspirations from abstract examples (analogy) to generate designs for kindergarten chairs through human computation. Results indicate that crowd workers generate more creative solutions when they are shown domain independent information and concrete constraints with visual examples as a source of inspiration. In another similar work (Yu, Kittur, & Kraut, 2016), authors claim that people show more creativity when they use relevant examples from outside of the problem's domain (abstract/schematic representation). They randomly displayed to crowd workers either an abstract representation of a problem (cup or power strip problem) or an original problem (from within a domain) to generate domains of expertise. Results indicated that crowd generated more diverse domain of expertise when shown abstract representation of problems than an original problem. In a subsequent study, they tested whether the found domains of expertise can produce relevant and useful inspirations. Crowd workers were randomly assigned to four conditions to propose solutions pertaining to either an original problem, schematic or abstract problem, self-selected (either cup or power strip) or irrelevant schema-driven problem (e.g., propose solutions for cup while drawing inspirations from power strip and vice-versa). Results indicated that the schematic or abstract driven condition increased the possibility of creating more creative solutions.

In the field of creative sketching, Yu and Nickerson (2011) introduced an approach based on a human-based genetic algorithm through crowdsourcing. In their approach, a first generation of sketches are evaluated based on originality and practicality using 7-point Likert scales (fitness ranking). Then few designs are selected to become parents of the next generation which produce new off-springs (designs) through combination. These off-springs then serve as a new generation and the whole process repeats. This work has been inspiration for ours as we also evaluated originality and practicality of IoT scenarios using the same 7-point Likert scale.

Hive (Thusoo et al., 2014) implements a novel approach by intermixing people instead of ideas to produce high-quality proposals. They proposed network rotation – an optimization algorithm which helps to improve the efficiency of a team by rotating team members regularly between small teams. The main theme of the research is that rotating the team members helps to bring novel perspectives to team which can stimulate creativity.

Luther et al. (2015) sourced design critique from the crowd to help designers improve their graphic designs. In their method, first, designers submit their initial designs and crowd workers critique them against known design principles. Their system aggregates critiques and allows designers to review the crowd-generated feedback efficiently.

Similarly, Apparition is an advanced sketching system (Lasecki et al., 2015) that can employ crowd workers in near real-time to transform low fidelity prototypes described in sketches and simple natural language to high fidelity prototypes.

Instead of recruiting unskilled and anonymous crowd workers, the Wish system (Kulkarni, Narula, Rolnitzky, & Kontny, 2014) leverages expert crowd workers to further enhance the creativity of three complex tasks – designing, writing and coding. The Wish system uses the "CrowdSearch" algorithm to find expert crowd workers even if there is no expert available by continuously sending requests to non-experts and requesting them to find experts.

Inspired by this prior work, in our research, we adopt a two-stage evaluation process of IoT scenarios involving first a larger and less qualified crowd before recruiting expert crowd workers. For this, we used a recruitment tool for hiring expert crowd workers as reported in (Smedema, 2017). We have briefly discussed these stages (Figure 2) in Section 3. Further details can be found in the procedure Section 3.4.

Previous studies evaluated design ideas using originality and practicality measures, but they do not provide much insights into what can contribute to an idea being characterized as creative. Only one study (Ahmed & Fuge, 2017) has examined several factors that contributed to classifying ideas into winning and non-winning ideas, although that work focuses on design communities in OpenIDEO, which is a crowdsourcing platform aiming to solve complex societal problems, whereas in our work we focus on non-expert end-users in an IoT context.

To address this gap our study examines several scenario features, such as: text metrics (e.g., word count), complexity of scenarios, prior knowledge of crowd-workers, number of trigger and action devices, and how these relate to creativity. The contribution of such an analysis is that it could potentially lead to authoring guidelines or an automated assessment of the scenarios' creativity.

3. Study 1: Crowdsourcing IoT scenarios

3.1. Aim of study 1

With Study 1, we aim to demonstrate the feasibility of engaging the crowd in this creative task and evaluate the quality of the output. Further, we analyze the factors which can contribute to the overall creativity of ideas generated through crowdsourcing. Based on our results, we discuss the feasibility of using crowdsourcing as a service to support ideation and creativity in IoT.

3.2. Structure of the study

We have briefly describe the stages of Study 1 as shown in Figure 2;

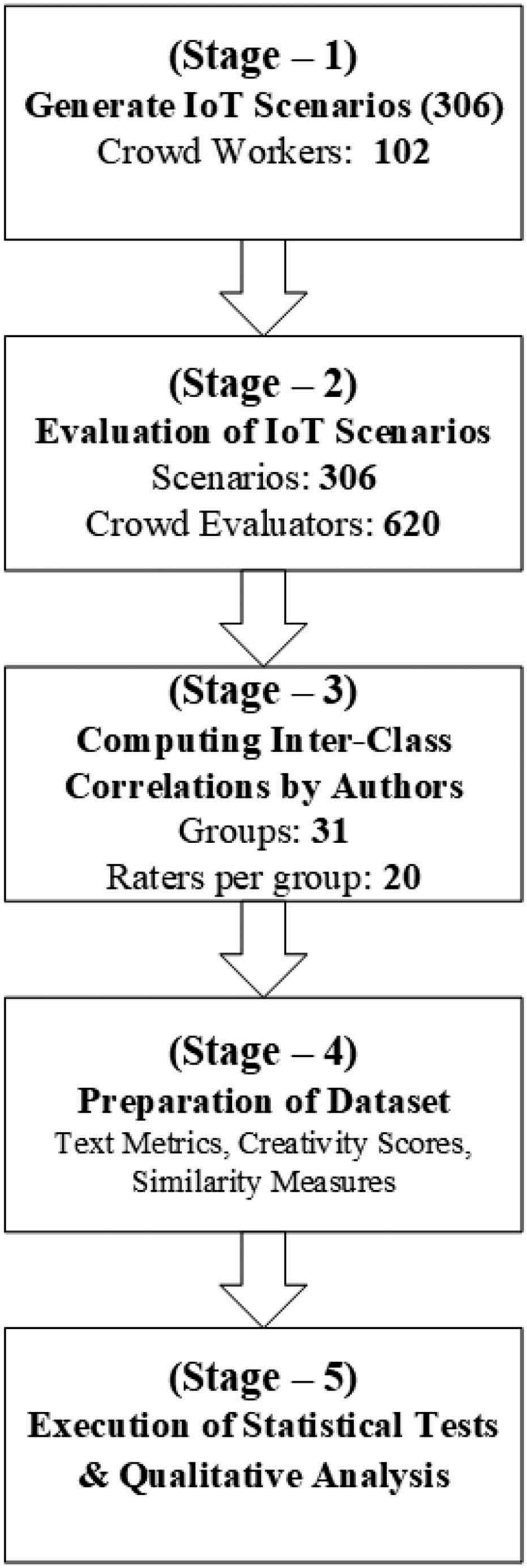

Figure 2. Stages involved in the first study.

Stage–1: In this stage, we invited 102 workers from MTurk to create the IoT scenarios. The input to this stage was mere instructions of the task (both video and text) to crowd workers. The outcome of this stage was 306 scenarios in more than 7 categories (avg. 60 words per scenario).

Stage–2: In this stage, we invited 620 workers from MTurk to evaluate the IoT scenarios, from the previous stage, on their creativity (originality and practicality). Each worker evaluated 10 scenarios and could not repeat the task.

Stage–3: Before executing statistical tests, we checked the reliability of ratings for both originality and practicality that we got from crowd workers (through intraclass correlation coefficient (ICC) in SPSS).

Stage–4: Two researchers then prepared the data set. This included counting the number of trigger and action devices as well as their triggers and actions. Additionally, text metrics (e.g., word count, long words, etc.) were also calculated through the Linguistic Inquiry and Word Count tool.

Stage–5: In the last stage, we executed all statistical tests and reported our findings. Furthermore, we conducted qualitative analysis of IoT scenarios to find features, which can be used to create novel IoT personalization techniques.

3.3. Participants

We conducted the study on Amazon Mechanical Turk (hereafter: MTurk) and restricted the survey to US crowd workers with over 98% approval rating and at least 5000 Human Intelligent Tasks (HITs)[8] approved. Based on our prior experience with MTurk, we decided to use approval ratings greater than 98% to ensure higher quality results. As an extra precaution, we set the HITs approved over 5000 to further increase quality results. We compensated each participant with $1.8 (matching the US minimum wage, 7.25$/hour); our survey average response time was 15 min and the total cost of this study was $226.80.

In total 102 MTurk crowd workers wrote 306 smart-home usage scenarios; we asked each one to write three. We removed from further analysis one worker whose three scenarios were all spam. The remaining sample was 48.5% male ($N = 49$) and 51.5% female ($N = 52$) with ages ranging from 23 to 68 ($M = 37.28$, $SD = 9.78$). Among the 101 participants, 48 (47.5%) had no programming experience, 32 (31.7%) had beginner-level programming experience, 15 (14.9%) had intermediate-level programming experience, and 6 (5.9%) were advanced. Furthermore, 48.5% (49) workers had smart home experience ranging from few months to a maximum of 120 months ($M = 12.83$, $SD = 22.7$).

To assess whether the crowd can provide novel and practical scenarios, we hired a separate group of 620 crowd workers from MTurk to evaluate the creativity of scenarios. We did not collect the demographics for this group of crowd workers, nevertheless, we used the same qualification criteria (over 98% approval ratings and over 5000 HITs approved).

3.4. Procedure

As we described in the study stages, the procedure was divided into five steps: 1) scenario generation through crowd, 2) evaluating the originality and practicality of scenarios, 3) checking the reliability of crowd ratings for both originality and practicality, 4) preparation of data set, 5) and finally, execution of statistical tests and qualitative analysis.

3.4.1. Stage 1: Smart home scenario generation

Through MTurk we investigated whether an anonymous crowd could assist smart home inhabitants in the creation of meaningful scenarios by combining the functionalities of given devices. In Figure 3, we show visuals of five selected input and output devices. We list all input and output devices in Table 1. As aforementioned, we were inspired by the Know Cards approach, but instead of using input and output cards for trigger and action devices, we used a picture of the devices presenting along their features. The idea behind using pictures as an "input" was also inspired by GenIOT project (Negri et al., 2012) which uses fake sensors and actuators as a visual stimulus to foster creative thinking.

On the first page of the survey, crowd-workers were asked to watch a short video (1 min and 29 s) that described the purpose of the research in detail. In the next page, we asked crowd workers to imagine that:

You have a home with devices that are connected to the Internet. You are required to pick any combination of the devices and then write three meaningful scenarios related to the selected devices.

We highlight that we purposely did not describe any high-level needs and context of the intended end-users to the crowd workers for the scenario creation process; Our intention was to capture the high-level needs of end-users from crowd workers who are essentially potential end-users of domestic IoT technologies themselves. Thus, the process of "needs discovery" and "scenario creation" was combined in one step. We assume that people will create scenarios to fulfill their needs in real settings. For example, a scenario from Home Security category describes user needs as:

> The basement in my home is prone to leaking during extreme periods of rainfall or after the melt from a lot of snowfall. If I could detect the leak as soon as it begins, it can help me prevent a significant amount of damage. As soon as the water sensor detects a leak, I would want the water sensor to send me an alert on my smartphone. If possible, the sensor could send estimates of the amount of water being leaked. This would help me stop the problem before it becomes a larger issue.

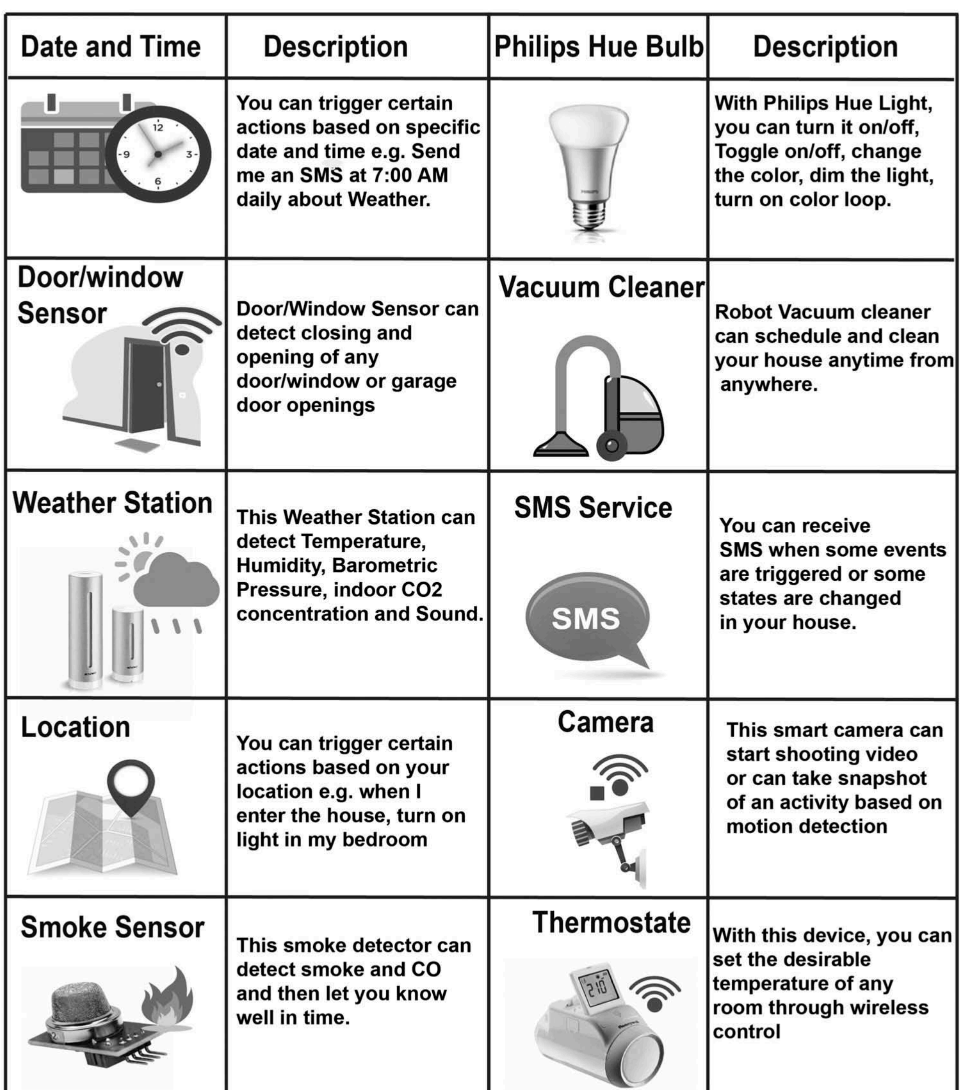

Figure 3. Five selected input and output devices among the ones that crowd workers used to create smart home scenarios.

Table 1. List of trigger and action devices used in the study

Trigger devices	Action devices
Date and Time	Philips Hue Light
Location	Thermostat
Door/Window Sensor	Smart Lock
Smoke sensor	Smart Plug
Weather Station	Vacuum cleaner
Motion Sensor	Smart Coffee Maker
Light/Brightness Sensor	Smart TV
Temperature Sensor	Google Chromecast Audio
Water Sensor	Air Conditioner
Doorbell Sensor	Smart Kettle
	Smart oven
	Smart Washer
	Siren
	IP Camera
	SMS
	E-Mail

To avoid confusion, we used the term "input devices" for trigger devices/channels and "output devices" for action devices/channels. We also provided a field for entering new devices, which were not already on our list. We then showed them an example, sample scenario:

> When I leave for work, I would like my front and back doors to lock automatically after I am at certain distance away. When I get back home from work, I want them to unlock as I approach and then that signal would be sent to the smart thermostat to turn the heating on as necessary depending on the outside temperature.

We also cautioned them to avoid writing generic, meaningless scenarios. To reduce the cognitive load, we showed trigger devices (10) and action devices (16) in two separate pages and showed their selected input from the previous page to next page and on the actual scenario creation page – to avoid them having to recall the devices they had selected in previous pages (See Figure 4). We then asked them to categorize each scenario by selecting some predefined categories (*Comfort, Energy Saving, Home Security, Elderly Care, Child Care, Entertainment, Health, Cooking*) or by adding their own category. At the end of the survey, we asked them demographic-related questions such as: smart home experience, age, gender, family size, educational level and programming experience.

3.4.2. Stage 2: Smart home scenario evaluation by anonymous crowd

To evaluate the collected scenarios, we used the measure of creativity as previously used in a similar study (Yu & Nickerson, 2011). According to this two-item measure of creativity, a creative idea is rated on two seven-point scales for: originality (novelty) and practicality (usefulness).

First, we divided the scenarios in 31 groups where each group contained 10 scenarios; except the last group, which contained six scenarios. We broke down this task to increase feasibility by making it less tedious and speed-up its completion. For each group, we hired 20 crowd workers to evaluate the creativity of

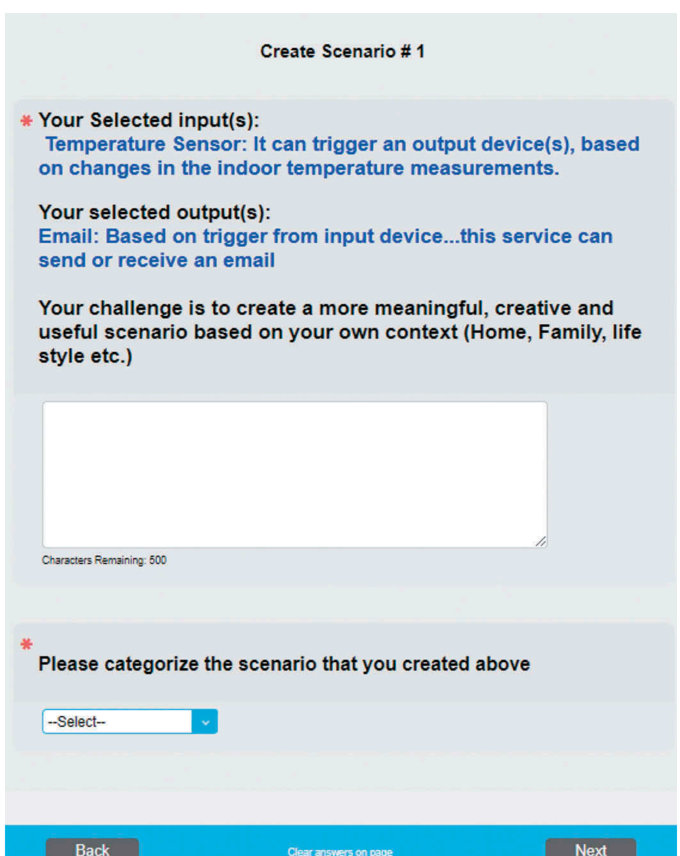

Figure 4. Scenario creation page. It shows all selected input and output devices (from previous pages) on the scenario creation page to avoid recall.

scenarios on a seven-point Likert scale. According to a previous study (Riedl, Blohm, Leimeister, & Krcmar, 2013), a minimum of 20 crowd workers per object of evaluation are required to achieve reliable ratings from crowdsourcing platforms. According to our calculations, the average completion time of the evaluation was roughly 3 min. Therefore, to satisfy the average US minimum hourly wage, we compensated each worker with 0.40$. The total cost of the evaluation was $347.20 (roughly 11.20$ per group). We received 12,240 ratings from 620 crowd workers for the 306 smart home scenarios, which were generated in the previous phase (Each scenario has two dimensions: originality and practicality–two-item measure of creativity leading to two separate 7-point Likert scales per scenario. Hence total ratings are calculated as: 306 scenarios x 20 worker ratings x 2 items = 12,240.

3.4.3. Stage 3: Computing inter-class correlation

Following a previous study (Wang, Wang, & Tao, 2017), we computed the inter-rater reliability of evaluations, using inter-class correlation coefficient (ICC) for originality and practicality for each group separately.

3.4.4. Stage 4: Preparation of data set

Next, we computed the mean scores for both originality and practicality from 12,240 ratings. Following the approach in (Yu & Nickerson, 2011), we also selected the mean score of 4.0 as the threshold value for both originality and practicality. In other words, we considered a scenario as creative if the scores for both originality and practicality were greater than 4.0. We then computed a new variable 'creativity' by calculating the mean of originality and practicality to get the overall creativity scores Wang et al. (2017). Further, we created a dichotomous variable 'creative', which we encoded as 1 if the scores for both the originality and practicality were greater than 4.0 (on a seven-point scale) and 0 otherwise. We used SPSS for data analysis.

To get an overview for the scenario's complexity, two researchers (coauthors of this paper) counted the number of trigger and action devices as well as the total number of triggers and actions used for each scenario. We also computed text metrics like word count per scenario, number of unique words per scenario, words per sentences, long words per scenario (more than six letters) and difficult words per scenario using Linguistic Inquiry and Word Count (LIWC) tool (Tausczik & Pennebaker, 2010) and through an online web tool.[9] A word is considered difficult if it does not belong to 3000 familiar simple words.[10] We also computed the structural complexity of scenarios by counting the number of trigger devices that were combined with output devices.

We were also interested to know whether similarity measures can be applied to evaluate crowd-generated scenarios for their originality. This was assessed through a cosine-based similarity matrix and TF-IDF ("term frequency" and the "inverse document frequency") (Manning, Ragahvan, & Schutze, 2009). Cosine similarity is a vector-based measure of the similarity of two strings. In this method, we transform each string into a high-dimensional vector space in which strings that are closer to each other are considered more similar and vice versa. Moreover, we checked whether the crowd workers' subjective evaluation of originality could be gauged by such a similarity metric, i.e., its inverse was expected to relate to originality. We used the scikit-learn Python library for implementing TF-IDF and cosine similarity. The result of this analysis was a score from 0 (not similar) to 1 (exactly similar) for each scenario. We calculated the mean of pairwise comparisons generated from the Cosine similarity matrix.

3.4.5. Stage 5: Execution of statistical tests and qualitative analysis

Since we asked each crowd worker to write three scenarios, we were expecting some variance in the creativity scores and other text metrics (e.g., word count, unique words, etc.) between the first, second and third scenario. Since we showed their created scenarios in the next page after each scenario creation and encouraged them to use a different set of input/output devices and think of different scenarios, we were expecting that the next scenario would be more creative than the previous one. A repeated measures one-way ANOVA was carried out to test within subject differences in the mean scores for creativity.

We also ran a multi-factorial 3×4 mixed ANOVA to examine the between-subject (Gender, Programming Experience, Smart Home Experience, etc.) and within-subject differences for the mean scores for creativity and other text metrics.

Finally, one researcher executed all the tests and reported the findings. Furthermore, an inductive thematic analysis (Braun & Clarke, 2012) of the collected scenarios was carried out using the Atlas.ti[11] software. The purpose of this analysis was to find the recurrent themes and patterns which can be used to create the more natural and expressive End-User Development environments for personalization of IoT devices and services.

4. Results from study 1

4.1. Introduction

In total, we collected 306 scenarios (three from each MTurk worker). Out of those 306, we removed 23 scenarios: three scenarios were meaningless/spam and 20 scenarios had low-reliability scores among raters (we expand on this point in Section 4.3). Hence our final sample contained 283 smart home scenarios. Out of 283 scenarios, 153 scenarios were creative while 130 were rated as ordinary by crowd workers. Among the 283 scenarios, 164 were rated as highly novel and original while 264 were rated as highly practical and useful.

Figure 5 shows the distribution of scores for both originality and practicality; the normal distribution curves indicated that our data is approximately normally distributed in both cases. We also calculated the z-score tests of skew and kurtosis to confirm normality; since the size of our data was more than 100, we decided $\pm$ 3.29 as a cutoff value. The z-score for skewness and kurtosis was -1.67 and -1.71, respectively (within ± 3.29 range). By looking at the graph, about half of the scores (164 out of 283) for originality were below threshold (4.0), while in case of practicality (264 out of 283), scores were tilted toward higher values. This finding makes sense: if a scenario is not original this does not imply that it is impractical. Furthermore, this finding implies that it is easier to come up with practical than original scenarios.

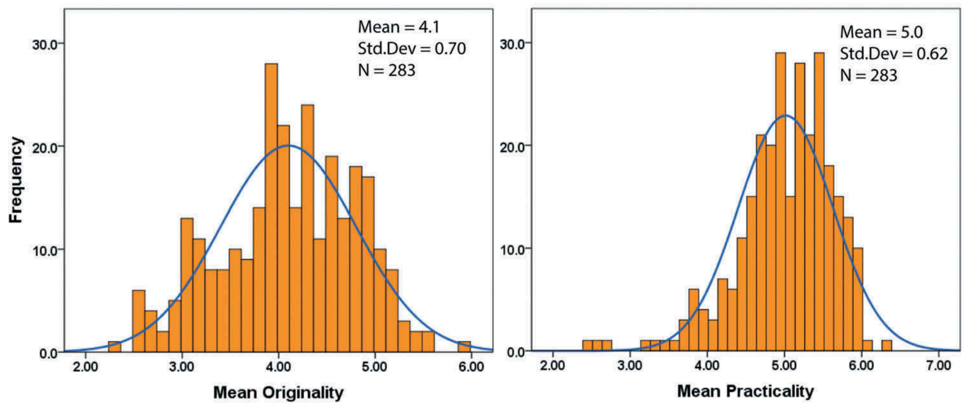

Figure 5. Normal distribution curves showing the spread of creativity scores across both dimensions: Originality and practicality.

Table 2. Workers created several different categories from the predefined categories we presented to them

Worker-defined categories	
Aquarium Monitoring	Pet Care
Maintenance	Home Arrival
Multiple categories (Scene)	Cabine Damage
Weather Conditions	Convenience
Save Time	Protection from Water Damage
Automobile	Cleaning
Housework	Home Utilities
Keep Plants Alive	Lost property

4.2. Categories

As aforementioned, crowd workers could classify their scenarios by choosing from a list of predefined categories or create their own new category. Examples of worker-defined categories are: *Pet Care, Water and Energy Savings, Aquarium Monitoring, Home Utilities and Maintenance* to name a few (see Table 2). For our analysis, since most of the worker-defined categories were redundant or sub-categories of predefined ones, one researcher carefully examined them and collapsed them into predefined categories. Due to this, we left only one worker-defined category: *Pet Care*. Among others, the most prominent chosen categories were: *Home Security, Comfort, Cooking and Energy Saving* (See Figure 6). The Other category includes one worker-defined category *Pet Care* (1.06%) and two predefined categories: *Health* (1.41%) and *Elderly Care* (0.71%).

These results show the diversity and classes of IoT scenarios that we got from crowd workers. Crowd workers used our predefined categories and also created new categories, which indicates their motivation and interest toward the task.

4.3. Reliability of crowd ratings for originality and practicality

To find the important factors or correlates of creativity, it was important to first check the reliability of ratings for both originality and practicality that we got from crowd workers. This was checked through intraclass correlation coefficient (ICC) in SPSS. In Table 3, we have reported detailed analysis showing the average values for Cronbach's alpha, Avg. ICC scores and P values of the F test.

As we mentioned in the procedure section, we made 31 groups and each group contained 10 scenarios for crowd workers to evaluate. We invited 20 crowd workers for each group to ensure high reliability in the ratings Riedl et al. (2013). Another reason to use large number of groups and crowd evaluators for creativity evaluation was to generalize our results.

Out of 31 groups, only two groups failed to achieve high reliability in ratings (gray-colored rows in Table 3), so we excluded all those scenarios that were in those groups for further analysis. The ICC scores for those groups were less than the minimum 0.4 threshold Cicchetti (1994). In total, we had to remove 20 scenarios from the analysis in addition to three scenarios that were spam, ending up with a total of 283 smart home scenarios.

4.4. Finding 1: Computational metrics for crowd-created IoT scenarios

In this section, we report a series of salient metrics to differentiate creative IoT scenarios from the ordinary scenarios. First, we explored what features make IoT scenarios more original and practical. Our analysis includes the following computational metrics: (1) *Text Metrics* (e.g., number of words, unique words, etc.); (2) *Number of Trigger Devices* (e.g., door sensor, smoke sensor, etc.); (3) *Number of Triggers* (e.g., when temperature is greater than 37 degrees); (4) *Number of Action Devices* (e.g., thermostat, smart lock, etc.); (5) *Number of Actions* (e.g., turn on the light); (6) *Services* (e-mail, SMS, location, etc.); (7) *Complexity*–combination of input and output devices. For instance, a scenario can contain one-trigger device and one-action device (1:1) or it can combine many trigger devices to many action devices (X:X). In total, we analyzed four combinations or structures (1:1), (1:X), (X:1), (X:X); (8) *Diversity*–How often a particular trigger device was combined with other action devices and vice versa). In other words, we explored the design space of all possible combinations of input and output devices and services through a heatmap.

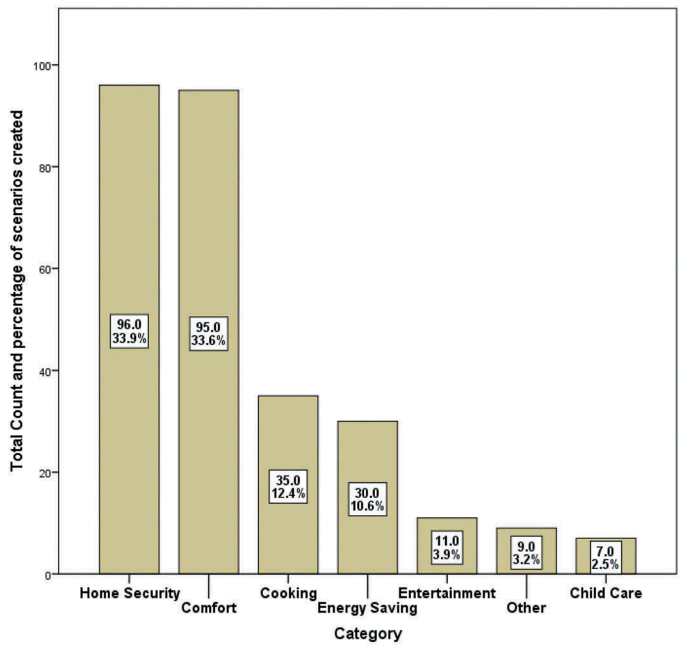

Figure 6. Distribution of categories in which crowd workers classified their scenarios. It closely resembles a bimodal distribution with home security and comfort topping the list.

Table 3. Reliability of crowd's ratings for both originality and practicality. It presents values for Cronbach's alpha, ICC average scores and F-Tests for originality and practicality. The average values for Avg. ICC and *P* values of the F test indicates that there was high degree of reliability in the ratings provided by the crowd for both originality and practicality. N = 20 means that there were 20 crowd evaluators for each group. Each group contains 10 scenarios which were evaluated by crowd. We have only removed two groups (gray-colored rows) from further analysis due to low-reliability ratings.

	Practicality (N = 20)		F-Test (DF1 = 9, DF2 = 171)		Originality(N = 20)		F- Test (DF1 = 9, DF2 = 171)	
#	Cronbach's Alpha	Avg. ICC	value	P	Cronbach's Alpha	Avg. ICC	Value	P
1	0.872	0.834	7.797	$p < .001$	0.835	0.748	6.079	$p < .001$
2	0.875	0.861	8.010	$p < .001$	0.906	0.864	10.658	$p < .001$
3	0.791	0.711	4.782	$p < .001$	0.715	0.588	3.504	$p < .001$
4	0.877	0.856	8.098	$p < .001$	0.882	0.860	8.508	$p < .001$
5	0.794	0.695	4.846	$p < .001$	0.727	0.601	3.664	$p < .001$
6	0.223	0.123	1.288	$p = .247$	0.262	0.134	1.355	$p = .212$
7	0.571	0.532	3.329	$p = .017$	0.569	0.457	2.318	$p = .017$
8	0.543	0.460	2.189	$p = .025$	0.819	0.730	5.300	$p < .001$
9	0.623	0.549	2.655	$p = .007$	0.816	0.712	5.421	$p < .001$
10	0.856	0.827	6.942	$p < .001$	0.816	0.775	5.435	$p < .001$
11	0.913	0.902	11.46	$p < .001$	0.877	0.867	8.133	$p < .001$
12	0.808	0.772	5.204	$p < .001$	0.618	0.575	2.618	$p = .007$
13	0.481	0.389	1.926	$p = .051$	0.784	0.733	4.638	$p < .001$
14	0.964	0.951	27.87	$p < .001$	0.939	0.915	18.245	$p < .001$
15	0.616	0.424	2.607	$p = .008$	0.895	0.836	3.429	$p < .001$
16	0.615	0.498	2.595	$p = .008$	0.726	0.643	3.648	$p < .001$
17	0.409	0.362	1.692	$p = .094$	0.919	0.883	12.411	$p < .001$
18	0.603	0.495	2.521	$p = .010$	0.775	0.674	4.437	$p < .001$
19	0.585	0.469	2.408	$p = .013$	0.850	0.771	6.648	$p < .001$
20	0.725	0.695	3.640	$p < .001$	0.893	0.862	9.307	$p < .001$
21	0.726	0.649	3.647	$p < .001$	0.887	0.843	8.839	$p < .001$
22	0.720	0.614	3.570	$p < .001$	0.602	0.529	2.511	$p = .010$
23	0.535	0.401	2.152	$p = .028$	0.872	0.855	7.791	$p < .001$
24	0.627	0.535	2.680	$p = .006$	0.273	0.199	1.376	$p = .202$
25	0.907	0.876	10.77	$p < .001$	0.911	0.895	11.178	$p < .001$
26	0.742	0.603	3.877	$p < .001$	0.825	0.777	5.723	$p < .001$
27	0.774	0.702	4.433	$p < .001$	0.881	0.851	8.379	$p < .001$
28	0.537	0.359	2.158	$p = .027$	0.849	0.733	6.608	$p < .001$
29	0.890	0.863	9.107	$p < .001$	0.902	0.862	10.220	$p < .001$
30	0.849	0.832	6.639	$p < .001$	0.861	0.819	7.194	$p < .001$
31	0.820	0.668	5.551	$p < .001$	0.857	0.774	3.829	$p = .003$

Table 4. Correlates of originality and overall creativity

Variable 1	Variable 2	R	P
Creativity	Word Count**	r = .40	$p < .001$
	Unique Words**	r = .40	$p < .001$
	Difficult Words**	r = .32	$p < .001$
Originality	Word Count**	r = .43	$p < .001$
	Long Words**	r = .30	$p < .001$
	Unique Words**	r = .43	$p < .001$
	Difficult Words**	r = .39	$p < .001$
	Action devices*	r = .28	$p < .001$

**moderate correlation.
*weak correlation.

4.4.1. Correlation between creativity, text metrics, devices & similarity measures

The purpose of this test was to investigate whether computational metrics (e.g., word count, number of devices, etc.) could be used to automatically assess the creativity of IoT scenarios. Finding significant correlations means guiding the design of more sophisticated crowd interfaces: one could summon crowd workers to give input that comprises of more features (e.g., word count) and then one could respectively remunerate the crowd.

We examined IoT scenarios based on the number of devices mentioned in them, the number of triggers, number of actions and other text metrics. As shown in Table 4, we found a positive correlation between the creativity (mean of originality and practicality) and word count, unique words and difficult words.

We also found a positive correlation between originality and: word count, long words, unique words and difficult words. Surprisingly, practicality was not strongly associated with either devices or text metrics. We only found a weak positive correlation between practicality and: word count (r = 0.12, p = .04) and unique words (r = .14, p = .018).

These findings suggest that text metrics like word count, unique words and difficult words have a strong influence on creativity (mean score of originality and practicality) as well as the originality of scenarios. Additionally, a weak correlation between practicality and other metrics suggests that practicality is something which could not be judged computationally. A future study, could further explore the underlying factors that can influence the practicality of IoT scenarios.

We also found that there was practically no correlation between the crowd's subjective evaluation of originality and the computational evaluation of originality measured through cosine-based similarity matrix and TF-IDF (r = –0.12, p = .843). This indicates that assessing originality through similarity metrics is not yet straightforward and more research is needed to find suitable AI techniques to automate the ranking of scenarios by creativity – something which is needed to automate this process without the crowd's intervention.

4.4.2. Difference between creative and ordinary scenarios in terms of text metrics, devices & their triggers/actions

We were interested in examining the differences between creative and ordinary scenarios for text metrics, number of devices and their triggers and actions. If we find a significant difference between two groups, then this would assist to automatically classify scenarios into creative and ordinary ones.

We consider a scenario as creative when originality and practicality scores are greater than 4.0. Since each worker wrote three scenarios, we divided scenarios into three groups to meet T-test's assumption of the independence of observations: first group contains all scenarios which were created by independent crowd

Table 5. T-Test results for creative and ordinary scenarios. Several variables are significantly different between the creative and ordinary scenarios

T- Test results for First group of Scenarios.				
Variables	Creative (N = 45)	Ordinary (N = 46)	T-Test	P
Word Count***	60.76 ± (23.6)	43.20 ± (22.5)	t(89) = 3.638	p ± .001
Unique Words***	41.27 ± (13.7)	30.59 ± (12.3)	t(89) = 3.901	p ± .001
Difficult Words**	11.71 ± (6.5)	8.50 ± (5.8)	t(89) = 2.505	p = .014
Action Devices**	1.73 ± (0.75)	1.35 ± (0.70)	t(89) = 2.524	p = .013
Actions**	2.07 ± (1.18)	1.52 ± (0.91)	t(89) = 2.473	p = .015
T- Test results for Second group of Scenarios				
Variables	Creative (N = 56)	Ordinary (N = 45)	T-Test	P
Trigger Devices**	1.70 ± (0.73)	1.38 ± (0.61)	t(99) = 2.325	p = .022
Action Devices**	1.64 ± (0.77)	1.29 ± (0.46)	t(91.7) = 2.858	p = .005
Actions*	1.84 ± (0.83)	1.51 ± (0.73)	t(99) = 2.092	p = .039
Word Count**	58.07 ± (23.8)	43.27 ± (22.8)	t(99) = 3.168	p = .002
Word per Sentences**	25.82 ± (11.5)	20.5 ± (6.20)	t(87.7) = 2.944	p = .004
Long Words**	8.00 ± (5.53)	5.62 ± (3.72)	t(87.7) = 2.469	p = .015
Unique Words**	39.46 ± (12.8)	31.60 ± (13.0)	t(99) = 3.046	p = .003
Difficult Words**	11.11 ± (6.15)	7.56 ± (5.18)	t(99) = 3.092	p = .003
T- Test results for Third group of Scenarios				
Variables	Creative (N = 52)	Ordinary (N = 39)	T-Test	P
Trigger Devices** [1]	1.71 ± (0.72)	1.31 ± (0.47)	t(87.3) = 3.227	p = .002
Triggers**	2.06 ± (0.94)	1.49 ± (0.68)	t(89) = 3.212	p = .002
Action Devices**	1.58 ± (0.77)	1.23 ± (0.43)	t(82.5) = 2.716	p = .008
Word Count**	55.63 ± (13.2)	41.90 ± (25.7)	t(89) = 2.693	p = .008
Unique Words*	38.27 ± (12.8)	31.74 ± (15.5)	t(89) = 2.165	p = .033

*** $p < .001$, ** $p < .01$, * $p < .05$.

workers and so on. The dependent variable was the dichotomous creative variable which contains two groups: creative group, which includes all those scenarios whose scores for originality and practicality were greater than or equal to 4.0, and similarly all those scenarios that did not fulfill this criterion were in the ordinary group. Then we performed the T-test for each group separately. We confirmed that there is indeed a significant difference in the mean scores for different variables used in the study for all three groups (see Table 5). We can see that some variables appeared repeatedly in all groups, e.g., Word count, unique words, action devices, etc. These findings suggest that the previous metrics can potentially be used to provide automated estimates/feedback regarding the creativity of scenarios they contribute (we expand on this point in the discussion).

4.4.3. Complexity of generated scenarios

As mentioned in the method section, we encouraged crowd workers to use more than a single input and output device to generate creative IoT scenarios. We related the complexity of scenarios with their structure (i.e., the combination of input and output devices in a scenario). We hypothesized that more creative scenarios would contain more combinations of input and output devices. If we find positive results, then we can encourage the crowd to combine more input and output devices for authoring scenarios. Furthermore, this would also help in automatizing the classification of scenarios into creative and ordinary.

We analyzed the following combinations per scenario (we name this "structure"):

(1) One input device connected to one output device (1:1) – i.e., one scenario had one and only one input and output device
(2) One input device connected to many output devices (1:X) – i.e., one scenario had one and only one input device but more than one output devices
(3) Many input devices connected to a single-output device (X:1) – i.e., one scenario had more than one input device but only one output device
(4) Many input devices connected to many output devices (X:X) – i.e., one scenario had more than one input device and more than one output device

For the "structure" of scenarios, we created a nominal variable in SPSS with four potential values, which denoted the structure of scenarios. Next, we related these structures to the overall creativity of scenarios (a dichotomous variable: 1 creative, 0 ordinary). Table 6 shows the frequency of scenarios, which we classified into the structures.

Table 6. Structures/combinations of created scenarios

Structures	Ordinary	Creative	Total
1:1	71	37	108
1:X	16	33	49
X:1	26	42	68
X:X	17	41	58
Total	130	153	283

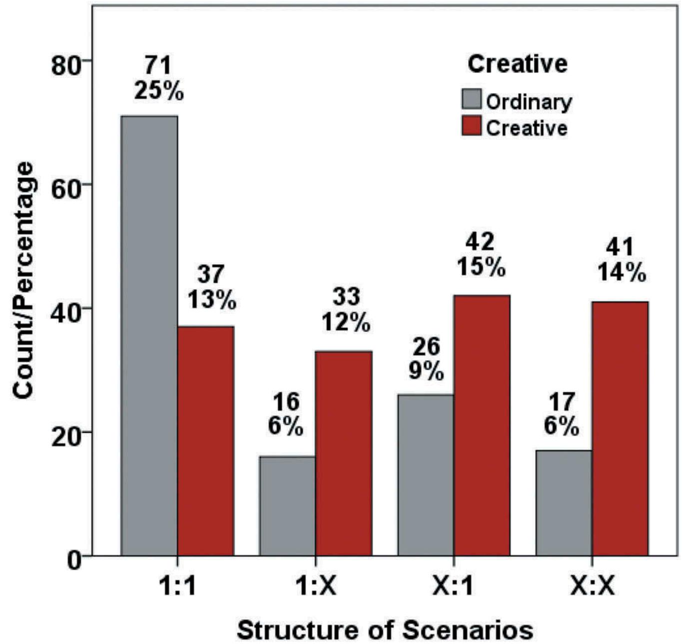

Figure 7. This figure shows different structures of scenarios emerged from study. Through Chi-square test, we find that more creative scenarios contained more combinations of trigger and action devices. Two such combinations were X:1 and X:X.

We found that more creative scenarios were those that fall into structures involving more than a single sensor and actor. We calculated a Chi-square test of independence comparing the frequency of creative and ordinary scenarios in all four structures. We found a significant difference $\chi^2(1, N = 283) = 28.6$, $p < 0.001$. As evident from Figure 7, most contributed scenarios involve more than a single trigger and action devices. The most scenarios (108) fall into a one-to-one structure and a large number (71) of those scenarios were ordinary. This also confirms our finding (previous section) where we showed through a T-test that a greater number of trigger and action devices was an important element of creative IoT scenarios. Surprisingly, among all other groups, X:1 and X:X groups contained more creative scenarios which shows that people like to create two kinds of scenarios: 1) triggering a single action device based on triggers from multiple devices; 2) triggering many action devices based on many trigger devices.

We further want to analyze the difference in the creativity scores between all the pairs of structures described in the previous section. Since each worker wrote three scenarios, we had repeated measures in the sample. Therefore, we divided our scenarios into three groups to make them single measures and to avoid the repetition of subjects. We intended to use the one-way ANOVA analysis, but our dependent variable (creativity) was not normally distributed in a group according to Shapiro-Wilk normality of test. For that reason, we used the Kruskal-Wallis test which is a non-parametric equivalent. The Kruskal-Wallis test indicated that there was a significant difference in creativity scores between all structures of scenarios in the first group ($N = 91$): $H(3) = 8.270$, $p = .041$. The median scores can be seen from Table 7. We subsequently ran the Mann-Whitney U tests to see the exact difference between each pairs of structures. These tests revealed that the difference was only significant between (1:1) and (X:X) structures ($U = 173.0$; $N1 = 37$;

Table 7. Creativity scores based on the structure of scenarios. Med: denotes the median of creativity. CI: confidence interval

	First Group				Second Group				Third Group			
Structure	N	Med	95% CI		N	Med	95% CI		N	Med	95% CI	
1:1	33	4.5	4.3	4.6	38	4.4	4.2	4.6	37	4.4	4.2	4.6
1:X	17	4.7	4.5	4.9	19	4.6	4.2	4.9	13	4.7	4.3	4.9
X:1	22	4.5	4.1	4.8	23	4.7	4.5	5.0	23	4.7	4.5	5.0
X:X	19	4.7	4.5	5.0	21	4.7	4.6	4.9	18	4.8	4.5	4.8

N2 = 18; p = .004; effect size(r) = 0.37). Likewise, we also found a significant difference in the creativity scores for the second group (N = 101): H(3) = 9.450, p = .024. The Mann-Whitney U tests indicated that the difference was significant between (1:1) and (X:1) structures (U = 283.0; N1 = 38; N2 = 23; p = .022; effect size(r) = 0.29) as well as between (1:1) and (X:X) structures (U = 228.5; N1 = 38; N2 = 21; p = .007; effect size(r) = 0.35). Lastly, we did find a difference in the creativity scores for third group (N = 91) as well: H(3) = 9.553, p = .023; nevertheless, this difference was only prominent between the (1:1) and (X:X) groups (U = 173.0; N1 = 37; N2 = 18; p = .004; effect size(r) = 0.38).

Again, these tests revealed that more creative scenarios had more combinations of trigger and action devices for the first, second and third ordered scenarios. This difference was much prominent between the (1:1) and (X:X) structures in all groups. These findings show that describing a scenario with a greater number of trigger and action devices would make it more creative no matter whether it is created first, second or third by crowd workers.

4.4.4. Diversity of generated IoT scenarios

In this section, we analyzed how different devices and services were combined together for creative and ordinary scenarios. Furthermore, we also wanted to know whether creative scenarios included more devices or services.

If we closely look at the heatmap for creative scenarios (Figure 8), we can see that the physical devices that were frequently combined were: motion sensor and IP camera (N = 15); temperature sensor and thermostat (N = 13); brightness sensor and hue light (N = 13). Regarding most combined services, location was combined with smart lock (N = 17); SMS with motion sensor (N = 13) and doorbell sensor (N = 13); and E-Mail with water sensor (N = 7). Additionally, the date/time service was combined with Hue light (N = 9) and thermostat (N = 10). Moreover, we observed that Location, which is an input service, was mostly combined with two output services: SMS (N = 10) and E-Mail (N = 6).

Regarding the ordinary scenarios (Figure 8), the temperature sensor was combined frequently with thermostat (N = 20) and

Output Devices / Input Devices	Hue Light	Thermostat	Smart Lock	Smart Plug	Vacuum cleaner	Coffee Maker	Smart TV	Google Chromecast Audio	Air Conditioner	Smart Kettle	Smart Oven	Smart Washer	Siren	IP Camera	SMS	Email
Date and Time	9	10	2	2	7	9	4	3	4	3	4	0	0	0	5	1
Location	12	10	17	2	8	3	7	2	4	0	2	1	2	6	10	6
Door/Window Sensor	1	4	8	0	4	1	1	1	2	1	2	0	1	4	6	3
Smoke sensor	3	0	4	0	1	1	1	0	1	0	2	0	5	3	8	3
Weather Station	2	9	2	2	1	1	1	1	4	1	1	1	0	1	3	1
Motion Sensor	10	5	5	1	3	8	3	1	3	3	1	1	4	15	13	3
Brightness Sensor	13	4	3	0	2	4	2	2	2	2	1	2	2	6	3	4
Temperature Sensor	3	13	1	4	2	1	2	0	8	0	3	1	2	1	8	1
Water Sensor	1	0	2	1	2	0	1	0	0	0	2	3	2	3	5	7
Doorbell Sensor	9	1	7	0	1	2	3	0	1	0	1	1	2	11	13	4

Output Devices / Input Devices	Hue Light	Thermostat	Smart Lock	Smart Plug	Vacuum cleaner	Coffee Maker	Smart TV	Google Chromecast Audio	Air Conditioner	Smart Kettle	Smart Oven	Smart Washer	Siren	IP Camera	SMS	Email
Date and Time	1	7	3	2	3	16	3	3	4	7	8	2	2	8	2	2
Location	2	6	3	1	3	4	4	0	5	1	3	0	1	4	5	1
Door/Window Sensor	0	3	2	0	1	0	2	0	0	0	2	2	2	3	2	3
Smoke sensor	1	2	1	0	1	1	2	0	1	0	1	0	2	2	3	0
Weather Station	1	7	0	1	1	2	1	4	3	1	1	0	0	4	4	4
Motion Sensor	3	1	0	1	1	6	1	2	2	2	3	0	3	9	3	1
Brightness Sensor	3	2	0	0	4	4	1	0	0	1	1	0	0	1	3	1
Temperature Sensor	1	20	1	1	2	3	1	2	10	0	3	0	1	5	5	3
Water Sensor	0	0	0	1	1	1	0	0	0	1	0	1	0	0	1	2
Doorbell Sensor	1	2	2	0	1	1	3	0	1	0	2	1	1	11	2	2

Figure 8. Heatmaps for creative and ordinary scenarios: it shows the combination of input devices and services with the output devices and services. On top, the heatmap for creative scenarios is shown while at the bottom, the heatmap for ordinary scenarios is shown. The rows represent the input devices and services while columns represent the output devices and services (input and output services are shown in red color). The numbers within cells represent the frequency of combination between input and output devices and services. The darker the color is, the more frequently two devices/services at the intersection were combined together while the lighter the color is the less two devices were combined.

Table 8. Repeated measures one-way ANOVA for order (First, second and third scenario). There are no significant difference found, which means that the order does not impact creativity – i.e., a third scenario is not more or less creative than the first one.

Variable	Test results
Creativity	$F(2, 160) = 0.340$, $p = .713$
Originality	$F(2, 160) = 0.152$, $p = .860$
Practicality	$F(2, 160) = 0.045$, $p = .875$
Word Count	$F(2, 160) = 1.234$, $p = .294$
Words Per Sentences	$F(2, 150) = 0.529$, $p = .590$
Unique Words	$F(1.8, 146) = .767$, $p = .455$
Difficult Words	$F(2, 160) = 2.287$, $p = .105$
Long Words	$F(2, 158) = 1.674$, $p = .191$
Trigger Devices	$F(2, 160) = 0.048$, $p = .954$
Triggers	$F(2, 160) = 0.019$, $p = .981$
Action Devices	$F(2, 160) = 1.084$, $p = .341$
Actions	$F(2, 160) = 0.357$, $p = .700$

air-conditioner (N = 10) while the doorbell sensor was combined with the IP camera (N = 11). Regarding services, the most frequent combination was between date/time (input service) with the smart coffee maker (N = 16).

These findings show that creative scenarios covered a good range of both input and output devices as well as services. Furthermore, creative scenarios also combined services more often than the ordinary scenarios.

4.4.5. Influence of order on the creativity of IoT scenarios

Since we asked three scenarios from each crowd worker, we were expecting some variance in the creativity scores and other text metrics (e.g., word count) between the first, second and third scenario. An order effect for the creativity and text metrics could be important for deciding how many scenarios to request from each coworker. A repeated measures one-way ANOVA did not reveal any significant differences in the mean scores for all metrics for the first, second and third scenario (see Table 8).

In prior work (Abbas, Khan, Tetteroo, & Markopoulos, 2018), the overall creativity increased with each subsequent scenario χ^2 (1, N = 40) = 5.06, $p < 0.05$. Therefore, we were curious to know whether this finding can be generalized on a larger dataset (N = 283). To test our hypothesis that creativity increases with each subsequent scenario (practice); we related each scenario with an order variable (categorical with three levels) and tested order effects on the dichotomous variable "creative" . Consistently with the within subject's analysis above, we found that the mean score for both originality and practicality did not increase with the order χ^2 (1, N = 283) = 1.080, $p = .299$. It could be that because we hired crowd workers with over 98% approval ratings on MTurk, this helped ensure a consistent quality for all three scenarios.

4.5. Finding 2: Impact of individual differences on the creativity of IoT scenarios

We also ran a multi-factorial mixed ANOVA to find the between-subject (Gender, Programming Experience, Smart Home Experience, etc.) and within-subject differences in the mean scores for creativity and other text metrics. The purpose of this analysis was to determine whether demographics and knowledge of the crowd could also affect the creativity of scenarios. We assumed that smart home experience and programming experience will significantly contribute to creativity of scenarios while gender will not. If we find positive results, then this could lead to establishing qualification requirements to hire potential crowd workers.

We found no significant differences in the creativity scores between crowd workers of different gender $F(1, 77) = .247$, $p = .621$; effect size (d) = .05; we also found no difference between creativity scores and different levels of smart home experience $F(1, 77) = .491$, $p = .486$; effect size (d) = .07. But we did find a difference in relation to programming experience. We observed a significant interaction effect between programming experience and mean creativity, $F(6, 154) = 2.279$, $p = .039$; effect size (d) = 0.3. Mauchly's Test of Sphericity indicated that the assumption of sphericity had not been violated, $\chi(2) = 0.637$, $p = .727$.

We also observed an overall between-group difference in creativity for programming experience, $F(3, 77) = 2.816$, $p = .045$; effect size (d) = 0.33. The Bonferroni Post Hoc test indicated that there was a difference in the mean creativity scores between 'intermediate' and 'expert' programming group $p = .04$ (see Figure 9). We observed no difference between all other groups (See Tables 9 & 10 for more details).

From Figure 9 (Profile-plot), we found that scenarios of expert crowd workers were slightly ordinary, and this difference was only prevalent between intermediate and expert crowd workers. We were expecting such a difference between crowd workers having no programming experience and expert programmers: Because people who have interest in programming were expected to create more creative scenarios. We will expand on this difference between the 'intermediate' and 'expert' crowd workers in the discussion section.

4.6. Finding 3: Features of IoT scenarios for the design of future end-user development systems

An inductive thematic analysis (Braun & Clarke, 2012) of the collected scenarios was carried out using the Atlas.ti software. We found many interesting themes for end-user development: types of rules, automation styles, novel operators, value-based rules, self-disclosure, narrative styles and novel security solutions.

4.6.1. Type of rules

4.6.1.1. Personal rules. We categorized all those scenarios as personal rules that authors created for themselves alone and were expressed in first person singular (i.e., I) or personal pronoun (i.e., me). For example:

> When I get up in the morning and it is past a certain time like 5AM or so, turn on the electric kettle that I will use to make coffee. (56 year-old male with one year of smart home experience; worker-id: 92)

4.6.1.2. General rules. We categorized all scenarios as general that were created for everyone in the family. These rules either imply "nothing" or "you" as a subject pronoun. For example, there was no indication of subject pronoun in the rule created below:

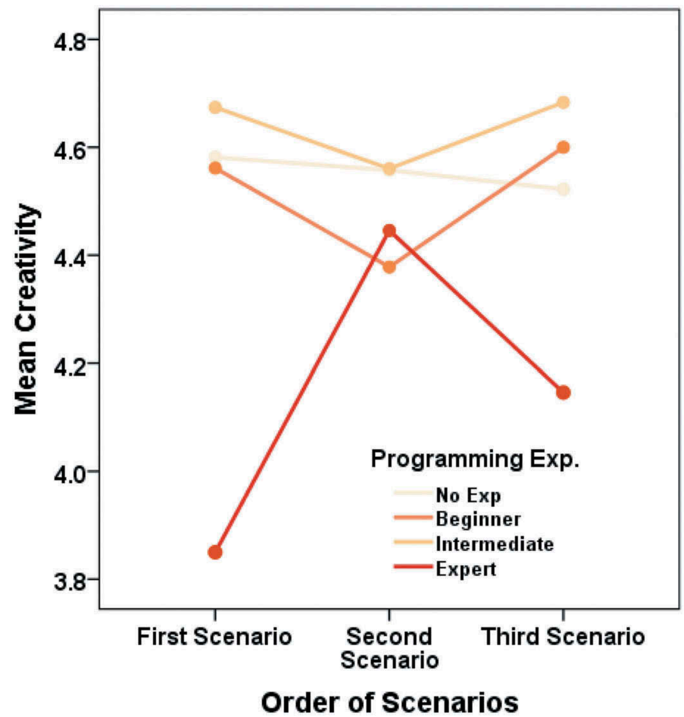

Figure 9. The goal of this figure is to examine whether programming experience and order of writing impacted creativity. Since each worker created three scenarios, we had repeated measures for creativity scores in our sample. The X-axis shows the within subject factor with three levels (first scenario created by a worker, second scenario created by the same worker and so on). The Y-axis represents the creativity scores. We ran a mixed 3 × 4 multi-factorial ANOVA to find between-subject differences (programming group) and within-subject differences in the creativity scores. Our result indicates that there was a difference in the mean creativity scores between the 'intermediate' and 'expert' programming groups. Overall, workers with intermediate programming knowledge were more creative.

Table 9. Programming exp. Vs. creativity: creativity scores for different programming groups (Mixed 3 × 4 ANOVA): three represents creativity scores for three scenarios while four represents programming groups

No. Exp.			Beginner			Intermediate			Expert		
Mean	SE	N	Mean	SE	N	Mean	SE	N	Mean	SE	N
4.58	.075	39	4.56	.095	24	4.67	.135	12	3.85	.190	6
4.55	.077	39	4.37	.099	24	4.56	.139	12	4.44	.197	6
4.52	.070	39	4.60	.089	24	4.68	.125	12	4.14	.177	6

Table 10. Within group main effects for programming groups and creativity (Creativity (1) = mean creativity score for scenarios which were created first and so on)

Creativity	Mean	SE	N	Programming Exp.	Mean	SE	N
Creativity(1)	4.417	.066	81	No Exp.	4.554	.056	39
Creativity(2)	4.486	.068	81	Beginner	4.513	.072	24
Creativity(3)	4.488	.061	81	Intermediate	4.639	.102	12
				Expert	4.147	.144	6

If the motion sensor detects motion from 9AM to 5PM from Monday to Friday, experience; worker-id: 16)

4.6.1.3. Social rules.

These rules involve other family members in addition to the author of the rule. For example, rules that involve wife, kids, mother, or even a workman or a courier (i.e., package delivery person) were classified as 'social' rules. For example:

> When I am not home, I would like the system to notify me if any doors open, if anyone rings a doorbell, and if there is any water leakage in the house. This will keep me informed of what may be happening in the house as it pertains to children or others that may want to intrude on my domain. I will know when the kids leave and get home from school and will know when the wife goes to and from work. I would like the alarm to be able to set it off if someone enters the house unwanted to deter them.(39 year-old male with one year of smart home experience; worker-id: 135)

Another scenario created for a workman:

> While away from your home program sensors to alert you of any changes like a water leak. This would allow you to get repairs done immediately by being able to unlock your door and let workman in. When they leave you can re-lock the doors. (65 year-old female with three years of smart home experience; worker-id: 259)

4.6.1.4. Cascading rules.

These rules get activated automatically after recognizing some activity. For example, "leaving a home" triggers a set of devices which in turn triggers another set of devices automatically. For example:

> I would like a technology that allows me to better control the temperature of my house while also controlling whether the lights are on and off. So let's say I'm leaving the house, and the garage door closes – and after a few minutes the garage light goes off, which would trigger the light brightness sensor (set on reverse to trigger when there is no light rather than more light), which then in turn turns the AC off while I am gone, saving energy. (42 year-old male with 10 years of smart home experience; worker-id: 68)

4.6.2. Automation styles

We have also observed several automation styles in the scenarios including agent-based rules similar to Alexa or Google Home, programming or IFTTT style, depicting the desire for manually handling devices and remotely controlling devices while away from home.

4.6.2.1. Agent-based. We classified all those scenarios as agent-based rules where a house automatically executes some action(s) based on the preferences set by the author or described "system" as a subject (e.g., like Alexa or Google home). The majority of scenarios fell under this category. For example:

> I would love a system that automated the coffee making process at working, saving me time and effort. So basically, when I come in the front door at work, a motion sensor sends a signal to the smart coffee maker, which then begins the brewing process so that I can go about my morning routine while the coffee is being made. (42 year-old male with 10 years of smart home experience; worker-id: 170)

4.6.2.2. Programming. Programming scenarios were those which were brief and resembled the traditional trigger-action programming (TAP) style or were inspired by IFTTT recipes. For example:

> If the motion sensor detects motion from 9:00am to 5:00 pm from Monday to Friday, activate the Siren and play the sound. (28 year-old male with one year of smart home experience; worker-id: 16)

4.6.2.3. Manual control. In this type, End-user wants to control devices manually, and not based on any pre-defined schedule or rules. In addition to that, user wants to execute some actions automatically while initiating manual control. For example:

> My smart tv disconnects from the Internet every time I turn it on (32 year-old male with one year of smart home experience; worker-id: 1)

4.6.2.4. Remote control. We also observed that authors desired for remotely controlling some devices when they were away from home. For example:

> I would like the Philips Hue Light to come on when I access it from my phone. I am away from home quite a bit at night and would like to turn it on and off at different times so that it seems that I am at home. (68 year-old female with one year of smart home experience; worker-id: 4)

4.6.3. Novel operators

From an EUD language perspective, crowd workers seemed to require some operators to express logic conditions, which alludes to the relevance of the natural language programming approach as discussed by Myers (Myers, Pane, & Ko, 2004). They mentioned novel conditional operators such as (*until, before/after, as soon as, once*), innovative programming operators (e.g., *ALL, ANY*) and unique iterations (*again, every, whenever*).

4.6.4. Value-based

Similar to a previous study (Clark et al., 2017), we also observed that crowd workers described a value or an ultimate goal in a scenario. For example:

> When I get out of bed in the morning, I would like the motion sensor to trigger the thermostat to turn on the heater if the weather is cold or turn up the thermostat if it is warmer outside as I tend to keep it colder when I am sleeping. When I reach my staircase, I would like the motion sensor to turn on my coffee pot and begin brewing a pot of coffee. I would also like it to start preheating the oven. This will save my family time in the morning and enable us to multitask for the upcoming day. (36 year-old female; no smart home experience; worker-id: 52)

4.6.5. Self-disclosure

We also noted that authors revealed information about their past experience and desired to tackle the situation with IoT devices. For example:

> My dog stays outside during the day when I am at work, but I would like him to have control to enter my home if he triggers the sensor on his doggy door. I want him to have the ability to come in and out of the door as he pleases. Only his sensor will work for the doggy door. (60 year-old female; no smart home experience; worker-id: 145)

Another example of self-disclosure is:

> We have a cabin in the mountain that we close down for winter but want to make short term trips there for a weekend. We want to keep the base temperature at a certain level so the water pipes will not freeze. We do not want to leave the furnace on all the time. We want it to be turned on when the temperature in the cabin gets below 40 degrees. When the temperature falls it will turn on the furnace so the water lines will not freeze. when we get to the cabin, we can then turn up the heat. (55 year-old female; no smart home experience; worker-id: 135)

4.6.6. Narrative styles

We also observed that authors created interesting narratives, i.e., story-like narratives while describing scenarios. For example:

> It is the winter, and I have just left for work in the morning. I get halfway to work and realize I never turned down the thermostat. I do not want the heat running all day because that would be a waste of money. I use my SMS system and send a message to turn the heat down to 63 until I get out of work, and then bring the temp back up to 68 when I am ten minutes away from home. (35 year-old male with one year of smart home experience; worker-id: 54)

4.6.7. Creative security solutions

Authors also created rules to signify that "they are at home" for deceiving potential intruders. For example:

> When the doorbell sensor is triggered, it activates Hue Lights in the home sequentially, making it appear that someone in the home is activating the lights in order to make the home appear to be occupied and a less ready target for crime. The doorbell sensor also activates the IP camera which takes video of the individual activating the doorbell, and a snapshot of the person is attached to an SMS sent to the phone, to allow the owner to monitor who is at the door. (34 year-old male with 6 months of smart home experience; worker-id: 125)

We will further expand on these results in the discussion section within the realm of EUD and what we have learned from these IoT scenarios.

5. Study 2: Validating the creativity of crowdsourced IoT scenarios

The purpose of Study 2 was to explore whether scenarios rated as 'creative' by the previous crowd workers would indeed be useful and practical for actual smart home users. We operationally defined "actual smart home users" as those who have been using IoT technology for at least 1 year or more. Furthermore, less-than-amateur's scenarios were already filtered out in the scenario evaluation stage (evaluated by 620 workers).

5.1. Participants

In total, 27 crowd workers participated in our study but only 20 were forwarded to the actual survey due to our inclusion criteria (we recruited 20 crowd workers for the same reason as we mentioned in Section 4.3). Furthermore, crowd workers in our study had varied smart home experience ranging from 1 year to maximum of 3 years.[12]

5.2. Procedure

We randomly picked two scenarios (one creative and one ordinary) from each of the four most frequent categories (*Comfort, Home Security, Cooking and Energy Saving*) from our first study. The reason to choose scenarios from these categories was that we had a larger number of scenarios to ensure that we could generalize our results. We used the same binary measure of creativity (Originality and Practicality) as we used in the scenario evaluation stage.

For the survey, we used the SoGoSurvey professional online tool SoGoSurvey (n.d.). Since we wanted to target specific users who have smart home experience, the SoGo survey application was useful in setting up a "skip logic" to pre-screen potential crowd workers. We also randomly shuffled the order of creative and ordinary scenarios for each worker.

We carefully set-up screener questions to recruit potential crowd workers from MTurk. In our first screening question we asked crowd workers to indicate whether they have smart home experience and anyone who responded with "no" was re-directed to a page explaining that they were not qualified for this survey. We made sure to also clearly mention this requirement on the MTurk task description page. The next screening question was: "How long you have been using smart home technology?" All respondents who indicated "less than a year" also did not qualify. We only selected crowd workers who had at least 1 year of smart home experience or more. We finally added a question which served as a final-quality check. We asked crowd workers to: "Write down the names of any two devices that you are currently using. If possible, also write the name of their vendor". This was not a screening question to disqualify crowd workers but helped us to double-check and have crowd workers with actual smart home experience.

We also carefully assessed the duration of the survey and set the cost equal to \$0.5 per HIT for total of 3-min survey – which amounts more than the USA minimum wage: \$7.25/hr. The total cost of this study was \$15. We set the same qualification as we set in the scenario creation and evaluation stage – approval rate more than 98% and number of HITs approved greater than 5000. We also calculated the Inter-Class Correlation coefficient for both originality and practicality measurements among the 20 crowd workers.

5.3. Results

Figure 10 shows some commonly used devices mentioned by crowd workers in our study; this shows that they were indeed actual smart home users and were familiar with the IoT technology.

We found a high degree of reliability between originality and practicality measurements for the eight smart home scenarios among our 20 smart home users. The average measure Inter-Class Correlation (ICC) for originality was .827 with a 95% confidence interval from .593 to .959 ($F(7,133) = 5.855$, $p < .001$). For practicality, the average measure ICC was .838 with a 95% confidence interval from .614 to .961 ($F(7,133) = 6.171$, $p < .001$).

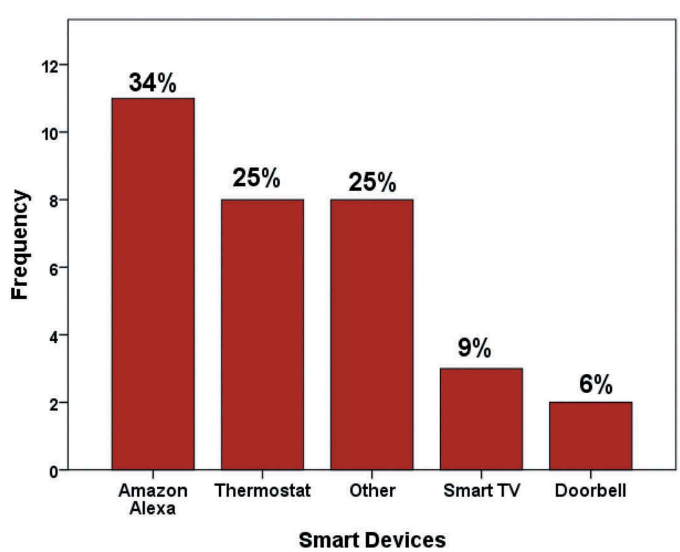

Figure 10. Smart home devices reported by crowd workers: Some devices were merged, e.g., **Amazon Alexa** = Echo + Echo dot, **Thermostat** = Nest + Honeywell + Vivint, **Smart TV** = Apple + Samsung + Vizio, Doorbell = Vivint + Ring. **Other** includes: Nest camera, Apple Homepod, Nest Hello, Alexa remote, Serena shades, ZIP DEVICE, Philips smart LED Bulb and simplisafe security.

Table 11. The comparison of scores between originality and practicality created by the initial crowd and then the 20 smart home users (evenly numbered scenarios represent the ordinary ones). To interpret this table correctly, consider both values separately (originality-before (OB) with practicality-before (PB) provided by ordinary crowd and originality-after (OA) with practicality-after (PA) provided by expert crowd. If both values are equal or greater than 4.0, then scenario is creative (You should not consider mean of Originality and Practicality).

#	Originality-before	Originality-After	Practicality-Before	Practicality After	Category	Creativity
1	5.15	5.25	4.00	4.15	Comfort	Creative
2	4.30	4.45	3.95	3.50	Comfort	Ordinary
3	4.60	4.55	4.55	5.20	Home Security	Creative
4	3.95	4.10	4.15	3.95	Home Security	Ordinary
5	4.50	4.70	5.15	4.60	Cooking	Creative
6	3.15	3.20	4.70	5.60	Cooking	Ordinary
7	4.45	4.10	4.90	5.65	Energy Saving	Creative
8	3.00	3.05	4.40	4.45	Energy Saving	Ordinary

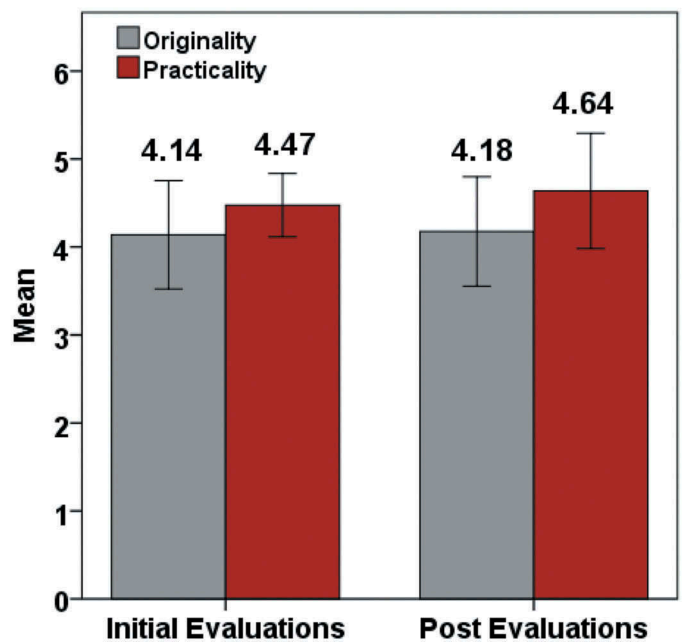

Figure 11. The Comparison of originality and practicality scores between the initial crowd and smart home users.

5.4. Finding 4: Strong agreement between the crowd and smart home users' evaluation concerning the creativity of IoT scenarios

As hypothesized, all four creative scenarios were rated as creative and vice versa – all four ordinary scenarios were rated as non-creative. From Table 11 and Figure 11, it is evident that originality and practicality scores of scenarios (evenly numbered scenarios represent the ordinary ones) highly correlated with the initial evaluation and the post-evaluation. We found a strong-positive correlation between the two: the evaluation created by the initial crowd and that by the 20 actual users of smart home technology (post group, recruited through screener survey) – for both originality ($r = .972$, $p < 0.001$) and practicality ($r = .730$, $p < 0.05$).

We also examined these eight scenarios based on the number of devices mentioned in them, the number of triggers, the number of actions and other text metrics as we did before with previous evaluations. We found a strong-positive correlation between the practicality and number of action devices ($r = 0.852$, $p < .001$). We also found that neither originality nor practicality scores were correlated with the text metrics (e.g., word count). Like previous evaluations, we found no correlation between originality and similarity measures assessed through cosine similarity – hence that strengthens our previous argument. We also found a positive correlation between creativity (mean of originality and practicality) and the number of action devices ($r = 0.716$, $p = .046$) as well as between creativity and difficult words ($r = 0.745$, $p = .034$).

6. Discussion

In this section, we compare our findings with prior work in the fields of crowd creativity and IoT to highlight the novelty of our work. First, we expand why computational metrics are important for IoT scenarios and how these metrics can assist in classifying scenarios into creative and ordinary groups. Next, we explain the relationship between complexity and creativity of IoT scenarios and how this complexity metric can help in designing intelligent interfaces for nudging crowd workers to select more IoT devices while creating IoT scenarios or IoT applications. After that, we describe why workers having intermediate programming skills performed better than expert crowd workers and why this skill matters for IoT scenarios. Then, we explain the relationship between scenarios and IoT personalization techniques and how our findings could inform the design of more natural personalization or EUD techniques for end-users. Then we explain how our work differs from prior in crowdsourced ideation. Finally, we close this discussion by highlighting the importance of the visual presentation of IoT devices and examples and their impact on findings.

6.1. Computational metrics for assessing crowd created IoT scenarios

In this section, we discuss three important metrics (text metrics, devices and complexity) that emerged from our studies, which can assist the automated assessment and classification of IoT scenarios.

6.1.1. Text metrics and smart home devices

Our results suggest the possibility to automate the assessment and classification of IoT scenarios based on computational metrics. We found the most common factors to differentiate between creative and ordinary scenarios were: number of words, unique Words, difficult words and action devices. We propose these metrics to predict the creativity scores of scenarios to further classify the scenarios into creative and ordinary ones. Our findings are in line with a previous study on OpenIDEO, an online creative community, in which the number of words, complex words, similarity measures, long words and sentences played a central role in classifying the winning and creative ideas (Ahmed & Fuge, 2017). Another related example from prior work is the Automated Essay evaluation system where the number of words and essay scores (showing quality of the essay) showed a positive correlation (Burstein, Chodorow, & Leacock, 2004). However, in our study we find a salient difference for the specific context of IoT scenario generation; namely the number of devices is also a relevant metric in addition to the text metrics reported in previous studies.

Furthermore, current crowdsourcing quality mechanisms usually improve the quality of ideas by first generating a large quantity of ideas and then employing a second unit of crowd-workers to evaluate their quality on different parameters Yu and Nickerson (2011). Another way to improve the quality is to combine the ideas generated from other crowd-workers to create more novel ideas Wang et al. (2017). Although these mechanisms help to improve the quality, they require extra time and costs. Furthermore, it is not feasible to use these mechanisms directly on the user interface to provide services to end-users in near real-time. For example, lets re-consider the context presented in Section 1.1, in which end-users request IoT scenarios from crowd workers on-demand when any relevant scenarios are unavailable. The task of assessing and classifying the scenarios can be automated based on the parameters presented in our study without involving extra crowd workers. In this way, we can leverage the power of both crowds and machines to improve the efficiency in terms of cost, time and quality.

6.1.2. Complexity of scenarios

We found that complexity of scenarios which we define as the 'number of trigger/input and action/output devices used per scenario' was also a significant factor in making those scenarios more creative and practical. Furthermore, we found that more creative scenarios were those that fall into structures involving more than a single trigger and action devices. Such were mostly found in two kinds of structures or combinations: A "X:1" structure in which a scenario had more than one input device but only one output device, and a "X:X" structure in which a scenario had more than one input device and more than one output device. Also, by comparing different scenario structures in terms of their creativity scores, we found a significant difference in the mean creativity between (1:1) and (X:X) structures in all three groups of scenarios. This finding alludes to guiding and encouraging the crowd to use a greater number of devices or services when creating novel IoT solutions for consumers. Since, it is not computationally hard to capture the input or output devices from the text using natural language processing techniques, we argue that this could be an important factor for automated assessment of creativity of IoT scenarios.

Furthermore, crowd-workers did not only create scenarios in different structures, but they also used both predefined and their own categories, which is an indication that they thoroughly explored the design space. This indicator addresses the variety in the scenarios, which is an important element of creativity (Nelson, Wilson, Rosen, & Yen, 2009).

In relation to the number of devices and the issue of variety, Ur et al. (2016) analyzed 200,000 recipes of IFTTT and showed that Philips Hue is the only IoT device that has been used in most of the recipes. This finding shows that IFTTT has not reached its maturity for the broader IoT domain because the possible space for device combinations is still sparse. Hence, we can fill this gap in functionality and possible combinations through crowdsourcing.

6.2. Crowd workers' intermediate knowledge in programming helps in creative input

We observed an overall between-group difference in creativity for programming experience (four levels) using analysis of variance. This difference in the mean creativity scores was prevalent between the 'intermediate' and 'expert' programming group. One possible reason for this difference was that expert crowd workers created scenarios that were more inclined toward formal representations or pseudocode styles. An example scenario of an expert in our sample:

> When I arrive home and my location is detected, I want the Philips lights to brighten up and I want my smart tv to tune to a certain channel and play music.

On the contrary, while people who had moderate or less programming experience also mentioned an actual situation or circumstances where such scenarios could help and as a result, created more one-off solutions. For instance, in an example below, the worker first described the actual situation and then mentioned each step that he/she would wish to follow in a less formal or structured way:

> One of my pet peeves in the morning is waiting for my car to defrost on cold mornings before work. Ideally, once I respond to my morning alarm on my smart phone, I would want the Weather Stations to check the weather conditions in my local area. If conditions for frost are detected, I want the Weather Station to send a request to my phone asking if I want to start my car now, so it can defrost. If I select yes, the car will run until there is no frost detected and then shut off.

One possible reason that impeded the creativity of expert crowd workers could be design or functional fixation (Crilly, 2015). Fixation refers to situations where novelty is blocked due to high level of knowledge or expertise (Bilalić, McLeod, & Gobet, 2008; Linsey et al., 2010); as a result, this can hamper the ability to think beyond familiar methods or knowledge which can serve as a basis for stereotypical thinking. For example, Gero (2015) conducted an experiment for the design of a novel device for supporting the elderly in a bath for the domestic context. For that purpose, they hired both advanced mechanical engineering and industrial design students (some students from each discipline were also shown some familiar examples). The results indicated that industrial design students produced more solutions and many of them were novel regardless of whether they had seen familiar examples or not as compared to mechanical engineering students.

These findings show that crowd-workers with intermediate programming experience can generate more creative scenarios than other crowd-workers. We highlight this as ambivalent finding, since we further need to investigate about the possible factors which helped crowd-workers having intermediate knowledge in programming to generate more creative input.

6.3. Relationship between scenarios and IoT personalization

In Section 4.6, we described some recurrent characteristics of IoT scenarios that can guide IoT designers to invent more natural, personalized and expressive interfaces. Now we further expand on the findings of Section 4.6.

6.3.1. Types of rules

We found that crowd workers created different types of rules including personal, general, social and cascading. In traditional TAP-like programming interfaces, usually IoT devices are controlled by only one user at a time and typically only one device is connected to another device to create a program or recipe. However, we have seen a variety of configuration styles in IoT scenarios; Personal rules are similar to trigger-action programming styles where only a single user creates a rule to personalize a set of devices for himself/herself. Social rules present more novel representations which involve other family members in the rule. We are only aware of the project We@Home (Caivano et al., 2017), which promotes collaboration among the family members of the smart home by associating rules to them. For example, one family member requests for an event-condition-action style rule and then other family members respond to those requests by creating rules for him/her. Furthermore, based on our finding, we argue that putting family members names directly in rules would promote transparency among family members. For example, if John is at home and Maria is out, then set the temperature according to John's preferences. If both John and Maria are at home, then activate the general rule which is applicable to both. Furthermore, cascading rules are somehow related with scenes. "Scenes are actually sequencing of actions that the user may manually activate" (Fogli, Peroni, & Stefini, 2017). We have seen various examples from our corpus where users desire to trigger several devices based on the initiation of a single activity. For example, when a user is leaving for work, he/she desired to activate many actions such as triggering a smart lock to close all doors and windows, triggering a washer and vacuum cleaner.

6.3.2. Automation styles

We observed different automation styles in our IoT scenarios. These automation styles somewhat relate to prior work in trigger action programming (Ur et al., 2014) in which smart home applications were coded to the following categories: programming, self-regulation, remote control and specialized functionality. We have seen from our corpus that many IoT scenarios resemble the self-regulation style or agent-based (in our case). In an agent-based style, a house automatically executes some action(s) based on the preferences set by the author of the rule. However, some users preferred a manual control (e.g., pressing a button) to activate some devices. Furthermore, the remote control style was described when a user was away from a home (e.g., lock the door through mobile device). In a programming style, users created very brief scenarios using an if-this-then-that style. It would be very useful to further analyze scenarios which fall into this category for the design of natural languages for end-users, which is beyond the scope of this paper.

We would argue that IoT personalization techniques that support different automation styles will give more freedom and flexibility to the owner of the smart home to control their devices according to the situation.

6.3.3. Novel notations

We identified several novel notations in our findings which further confirm the vision of Myers (Myers et al., 2004) about adding the element of naturalness in the end-user programming. In a prior study (Coutaz & Crowley, 2016), it was reported that the expression of compound conditions is difficult for users without programming experience. Here we claim that programming languages that use notations close to mental model of end-users would be more acceptable to end-users. A prior study (Huang & Cakmak, 2015) also proposed novel operators closest to our notations including *while, as long as, otherwise* etc. We claim that these operators in addition to our proposed novel operators can provide more expressiveness to automation rules which could then enable less tech-savvy users to construct complex logical expression in natural language. Furthermore, these notations could further be investigated in the light of prior work in building syntactically correct and semantically meaningful logical expressions (Metaxas & Markopoulos, 2017).

6.3.4. Associating values or goals with rules

We also noticed that crowd workers created rules with values or goals associated with them. This resembles the HOME (Benzi, Fogli, & Guida, 2017) – an intelligent agent based on a Belief-Desire-Intention (BDI) model with value as an additional component. According to this paper, "values represent abstract and permanent needs of the user that the system should satisfy, like, for example, security, comfort, savings, or health." For instance, each context (e.g., at home, away from home) must be associated with the values (e.g., energy saving, security) and when that specific context happens, a system should be able to recognize values set by users and then perform the most suitable action(s) to satisfy those values. Another study (Clark et al., 2017) also reported that users created smart home applications with goal or reasoning associated with them. They argue that one purpose of associating goal(s) with the IoT applications was to inform the home (considering it as an intelligent agent) about the potential goals of the users that it needs to fulfill. For example, if the rule is related to control the temperature along with energy saving goals, then the home should try to optimize the temperature in a such a way that it minimizes the monthly bill of energy consumption.

We claim that smart home agents that also understand users' goals and values could be a potential area of future research in designing new configurations for smart home users. More specifically in a context of social rules (described above) where the home (agent) must be able to resolve the conflicting goals of the family members.

6.4. Differences with other crowdsourced ideation techniques

Existing work in the realm of crowd creativity has mainly focused on creative writing (Siangliulue, Arnold, Gajos, & Dow, 2015), product design (Jang, 2014; Yu, Kittur, & Kraut, 2014; Yu & Nickerson, 2011), planning (Wang et al., 2017), problem solving (Yu et al., 2016) and mind mapping techniques (Andolina et al., 2017) (Table 12). In the IoT domain, researchers have developed participatory design techniques and card-based tools to foster creativity. However, crowd creativity, despite its potential, has not been explored to generate IoT scenarios. Our study is the first which leverages crowd creativity to generate IoT scenarios for the home context. The visual stimulation technique in creativity (Michalko, 2006), in which participants are presented with objects or pictures before generating as many ideas as they can, serves as inspiration for our work.

However, our work is different from prior work in many aspects; 1) Planning: mapping of input and output devices as well as their triggers and actions is not trivial and its quite different than visual stimulation technique (Michalko, 2006) where ideas are generated by mere looking at graphics. For example, this mapping process requires significant cognitive effort because workers need to think carefully which triggers from input devices should activate actions from output devices and in which context (e.g., leaving for work, going for sleep, etc.), before writing a scenario. Furthermore, workers can add new devices that they are accustomed to use in their own context; 2) Creative Writing: After that, workers need to write a new scenario. Unlike previous work of idea generation where short text snippets (Siangliulue et al., 2015) were produced, in this paper, each worker created three scenarios comprising of on average 60 words per scenario; 3) Categorization Task: In addition to writing, we asked as part of their main task to categorize their scenarios into either predefined categories or they can generate new categories if required. This is important for crowd ideation because the worker's task description can be adapted to nudge workers to write scenarios in categories that have fewer scenarios; For example, CROWDMUSE (Girotto, Walker, & Burleson, 2019) uses matrix form of visualization to show unexplored categories to crowd workers so that they can contribute thoroughly to the unexplored space. This really helps to reduce the redundant or ordinary ideas. 4) Programming Skills: IoT implies some programming knowledge to generate more creative scenarios. Indeed, we found that workers who had intermediate programming skills created more feasible and innovative scenarios than other workers. Our finding is somewhat similar to prior work where high exposure to CS domain helped to recognize the ability to perform automation (Clark et al., 2017).

With regard to re-purposing, there are many findings from our work which can be generalized to other domains; for example, we identify text metrics (namely number of words, unique words, etc.) for assessing crowd created IoT scenarios. These

Table 12. Differences with other crowdsourced ideation techniques

Ref.	Input or stimulus	Output	Domain
Siangliulue et al. (2015)	Example ideas described in text	Birthday messages	Creative writing
Yu and Nickerson (2011)	Sketches of chairs	Sketches of chairs	Product design
Yu et al. (2016)	Description of domain knowledge and constraints, Visual examples of chairs	Kindergarten chairs	Product design
Wang et al. (2017)	Playing a game	Game rules and reward strategies	Planning
Andolina et al. (2017)	Plain description of task as a brainstorming prompt	Concept map, recalling a name	Mind mapping
Siangliulue et al. (2015)	On-demand examples of ideas in text	Ideas for imaginary technology	Product design
Yu et al. (2016)	Abstract and concrete description in text of a problem	New product ideas for Power strip and cup problem	Problem solving
Jang (2014)	Image and text stimulus	Idea sketches	Product design
Siangliulue et al. (2016)	Idea map, seeding ideas/examples	Features for novel Mturk microtask	Planning
Our Work	Pictorial representation of input and output devices, good and bad IoT scenarios, smart home video, probe question	Smart home scenarios	Planning, creative writing and categorization

measures can be used to differentiate between creative and ordinary scenarios in other domains. On the other hand, findings that exclusively apply to IoT includes the combination of input and output devices namely complexity and diversity of IoT scenarios and other features related to IoT personalization.

6.5. Effect of visualizing IoT devices and examples on IoT scenarios

We observed that tasking crowd workers with real input and output IoT devices and services resulted in realistic scenarios. We did not find any scenarios from our corpus which could not be implemented with the current IoT devices and services. Tasking workers with open-ended probes without real devices could result in lots of futuristic solution which are not supported by current IoT devices. For example, in a related study (Ur et al., 2014), where workers were tasked to create IoT applications without real IoT devices, resulted in unrealistic scenarios (10.8%) requiring specialized functionalities or devices. Furthermore, based on the pilot studies that we did with crowd workers, we gathered some good and bad examples and then present it to crowd workers in the actual study as examples. We were inspired from prior work in crowd creativity which claims that presenting people with creative ideas may stimulate creativity (Siangliulue et al., 2015). This helped us to circumvent receiving less than amateur scenarios from crowd workers. Hence, we argue that providing real input and output devices along with good and bad examples is really important for eliciting realistic IoT scenarios.

Furthermore, providing input and output devices for the idea generation task helped us to cover a broad range of combination of devices and services. We noticed (Figure 8) that workers did not only create scenarios combining physical devices but they also combined physical devices with services. Though, there was difference in choice of devices and services between creative and ordinary scenarios; workers in the creative group preferred to combine more services (with other services or physical devices) than in the ordinary group. Furthermore, the creative group had more coverage of devices than the ordinary group: there were total of 23-unexplored combinations in the creative group while the ordinary group contained 38-unexplored combinations.

6.6. Future directions

In this section, we explore some future directions based on the current study.

6.6.1. Artificial intelligence guiding crowd workers for creative input

We plan to work on intelligent crowd interfaces to further enhance creative input which goes beyond traditional UIs on crowdsourcing platforms. For example, we can highlight in the UI untapped IoT categories and devices in the form of visuals similar to a bibliometrics diagram or an Idea Map (Siangliulue et al., 2015) and encourage crowd-workers to contribute in those categories and devices. This can help to avoid duplication from crowd workers and could lead to a system where a secondary task of feedback (determining the similarity of ideas) can be integrated into the primary task of the scenario generation task (Siangliulue, Chan, Dow, & Gajos, 2016). Nonetheless, this previous study has shown the efficacy of this approach for generating and judging the similarity of short text snippets (e.g., birthday messages). However, it is yet unclear how can we judge the practicality of ideas, apart from similarity. We believe that creating a smart home scenario has more factors (e.g., complexity, programming knowledge, and other computational metrics) which makes it an interesting venue to extend existing studies.

Furthermore, we can use machine learning (ML) algorithms to classify the scenarios as being creative or not. With an increase in training data, the accuracy of the classifiers improves which then allows to hand-off classification tasks to the machine. It is also possible to automatically hand-off classification tasks back to crowd workers when ML performance is below a certain threshold. For example, when new devices or categories are introduced over time, it is possible that ML might not perform beyond certain threshold (e.g. (Laput et al., 2015)).

6.6.2. Exploring the effect of order on the creative input

The order in which crowd workers wrote scenarios did not affect the creativity of scenarios. However, the overall creativity score increased from the first scenario (16%) to the second scenario (19.8%) before it slightly decreased from the second to third (18.3%). While these differences are not significant, an explanation for order not leading to better results could relate to fatigue and fixation (Crilly, 2015). For practical purposes and in the absence of more conclusive evidence, it would appear parsimonious to require crowd workers to generate two such scenarios. It would be interesting for future research to check the effect of sequence of crowd input on the overall creativity scores.

6.6.3. Exploring new categories and building ontologies in the IoT space

As we mentioned previously, crowd workers classified their scenarios in predefined categories as well as their own categories. The latter could help IoT vendors to get ideas from untapped categories, e.g., Pet Care and design IoT devices accordingly. Moreover, scenarios or use cases represented for a specific category on popular vendor's website are very limited in numbers and often described very briefly with no information of devices that can satisfy those use cases. Additionally, since crowd workers defined scenarios in several predefined and self-created categories, this would also help vendors to build a structured representation of IoT scenarios by building and evaluating ontologies with the help of the crowd (Getman & Karasiuk, 2014). This would also help end-users to search different variants (structures) of IoT scenarios when many of them are similar in nature and presented in the same category (e.g., Health). Furthermore, this would also help end-users to re-use the IoT scenarios by searching through a website or other systems using Information Foraging Theory (Fleming et al., 2013) or other techniques. Otherwise, one would create many duplicate IoT scenarios without using them in the actual context as was the case with the IFTTT (Ur et al., 2016).

7. Conclusion

In this paper, we reported the feasibility of collecting and evaluating IoT application scenarios collected using the crowdsourcing platform MTurk. IoT scenarios were analyzed along various dimensions which revealed correlations between creativity and various other computational metrics. We further analyzed the creative IoT scenarios and identified several insights that make these IoT scenarios more practical with the help of statistical analysis. In the second study, we triangulated the crowd-based evaluations with actual smart home users demonstrating the validity of the crowd-based evaluation. More specifically, we found that creativity of IoT scenarios corresponds to: a) Number of IoT devices and the number of combination of devices; b) Intermediate programming knowledge; c) computational metrics namely text metrics, number of devices and complexity. Furthermore, an inductive thematic analysis of the scenarios revealed many interesting themes which can guide practitioners of End-User Development systems to design more natural and expressive IoT personalization techniques for end-users. Additionally, our findings show potential to automate the classification and assessment of IoT scenarios based on several computational metrics. To promote further research on crowd-created IoT scenarios, we are also releasing our dataset to other researchers to contribute in this rapidly growing area. Our proposed approach could provide valuable and timely information to: smart home inhabitants who are in the planning phase or have already installed devices and want to get some starter ideas about possible use case scenarios combining different devices; IoT vendors who want to provide use case scenarios to their website related to IoT devices and can address other salient end-user aspects such as safety or security; designers of end-user development systems that are prominent in the IoT industry.

Notes

1. https://doi.org/10.6084/m9.figshare.8327096.
2. https://doi.org/10.6084/m9.figshare.7140431.v1.
3. https://scratch.mit.edu/.
4. http://smart-live.info/showroom/?lang=en.
5. https://www.mturk.com/.
6. https://www.figure-eight.com/.
7. https://microworkers.com/.
8. *"A Human Intelligence Task (HIT) is a term introduced on Amazon Mechanical Turk (MTurk) for completing a request of some digital work (e.g., writing a scenario) that can only be performed by people working on MTurk. A HIT represents a single, self-contained task that a Worker can work on, submit an answer, and collect a reward for completing"* **source**: https://www.mturk.com/worker/help.
9. https://wordcounttools.com.
10. https://wordcounttools.com/list-of-3000-familiar-words.html.
11. https://atlasti.com/.
12. Fortunately, in our first study, we found that 48.5% of workers had smart home experience though we did not require this from workers for creation task. But in our second study, having a smart home experience was mandatory to evaluate the IoT scenarios; Therefore, we put screening survey in place to hire only experienced workers.

References

Abbas, T., Khan, V.-J., Tetteroo, D., & Markopoulos, P. (2018). How creative is the crowd in describing smart home scenarios? supported cooperative work. *Extended Abstracts of the 2018 CHI conference on human factors in computing systems* (pp. 1–6), Montréal, Canada.

Ahmed, F., & Fuge, M. (2017). Capturing winning ideas in online design com- munities. *Proceedings of the 2017 ACM conference on computer supported cooperative work and social computing - CSCW '17* (pp. 1675–1687). Retrieved from http://dl.acm.org/citation.cfm?doid=2998181.2998249

Andolina, S., Schneider, H., Chan, J., Klouche, K., Jacucci, G., & Dow, S. (2017). Crowdboard: Augmenting in-person idea generation with real-time crowds. *Pro- ceedings of the 2017 ACM SIGCHI conference on creativity and cognition* (pp. 106–118). doi:10.1016/j.ptsp.2017.01.002

Aspiala, T. (n.d.). *Learn. Play. Collect. – Know cards*. Retrieved from https://know-cards.myshopify.com/

Benzi, F., Fogli, D., & Guida, G. (2017). Towards natural interaction with smart homes. In *Proceedings of the 12th biannual conference on Italian SIGCHI chapter* (pp. 1–6), Cagliari, Italy.

Bilalić, M., McLeod, P., & Gobet, F. (2008). Inflexibility of experts-reality or myth? Quantifying the Einstellung effect in chess masters. *Cognitive Psychology*, *56*(2), 73–102. doi:10.1016/j.cogpsych.2007.02.001

Braun, V., & Clarke, V. (2012). Thematic analysis. *Handbook of Research Methods in Psychology*, *2*. Retrieved from http://arxiv.org/abs/1806.08064

Brich, J., Walch, M., Rietzler, M., Weber, M., & Schaub, F. (2017). Exploring end user programming needs in home automation. *ACM Transactions on Computer-Human Interaction (TOCHI)*, *24*(2), 1–35. Retrieved from http://dl.acm.org/citation.cfm?doid=3077620.3057858

Brito, R., & Houghton, P. (n.d.). *IoT service kit*. Retrieved from http://iotservicekit.com/

Burstein, J., Chodorow, M., & Leacock, C. (2004). Automated essay evaluation: The criterion online writing service. *AI Magazine*, *25*(3), 27. Retrieved from http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Automated+essay+evaluation:+The+Criterion+online+writing+service#0

Caivano, D., Cassano, F., Fogli, D., Lanzilotti, R., Piccinno, A., Informatica, D., & Moro, A. (2017). We@home: A gamified application for collaboratively managing a smart home. In *International symposium on ambient intelligence, 615 (ISAmI)* (pp. 79–86). Retrieved from http://link.springer.com/10.1007/978-3-319-61118-1

Chen, X., He, Z., Anderson, D., Keller, J., & Skubic, M. (2006, 10). Adaptive silou- ette extraction and human tracking in complex and dynamic environments. In *2006 international conference on image processing* (pp. 561–564). IEEE. Retrieved from http://ieeexplore.ieee.org/document/4106591/

Cicchetti, D. V. (1994). Guidlines, criteria, and rules of thumb for evalauting normed and standardized assessment instruments in psychology. *Psychological Assessment*, *6*(4), 284–290. doi:10.1037/1040-3590.6.4.284

Clark, M., Newman, M. W., & Dutta, P. (2017). Devices and data and agents, oh my: How smart home abstractions prime end-user mental models. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, *1*(3), 44:1–44:26. doi:10.1145/3132031

Corno, F., De Russis, L., & Monge Roffarello, A. (2019). A high-level semantic approach to end-user development in the Internet of Things. *International Journal of Human Computer Studies*, *125* (November 2016), 41–54. doi:10.1016/j.ijhcs.2018.12.008

Coutaz, J., & Crowley, J. (2016). A first-person experience with end-user development for smart homes. *IEEE Pervasive Computing*, *15*, 26–39. Retrieved from http://ieeexplore.ieee.org/abstract/document/7445783/

Crilly, N. (2015). Fixation and creativity in concept development: The attitudes and practices of expert designers. *Design Studies*, *38*, 54–91. doi:10.1016/j.destud.2015.01.002

Dahl, Y., & Svendsen, R. M. (2011). End-user composition interfaces for smart environments: A preliminary study of usability factors. In A. Marcus (Ed.), *Design, user experience, and usability. Theory, methods, tools and practice. DUXU 2011. Lecture notes in computer science* (Vol. 6770, pp. 118–127). Berlin, Heidelberg: Springer.

Danado, J., & Paterno, F. (2015). A mobile end-user development environment for IoT applications exploiting the puzzle metaphor. *ERCIM News*, *101*(April), 26–27.

Desolda, G., Ardito, C., & Matera, M. (2017). Empowering end users to customize their smart environments. *ACM Transactions on Computer-Human Interaction*, *24*(2), 1–52. Retrieved from http://dl.acm.org/citation.cfm?doid=3077620.3057859

Dixon, C., Mahajan, R., Agarwal, S., Brush, A. J., Lee, B., Saroiu, S., & Bahl, V. (2010). The home needs an operating system (and an app store). In *Proceedings of the ninth acm sigcomm workshop on hot topics in networks - hotnets '10* (pp. 1–6), New York. New York, NY: ACM Press. Retrieved from http://portal.acm.org/citation.cfm?doid=1868447.1868465

European smart home market size, growth, trend and forecast to 2024. (n.d.). Retrieved from https://www.marketsandmarkets.com/Market-Reports/european-smart-homes-market-1290.html

Fauquex, M., Goyal, S., Evequoz, F., & Bocchi, Y. (2015). Creating people-aware IoT applications by combining design thinking and user-centered design methods. *IEEE world forum on Internet of Things, WF-IoT 2015 - Proceedings* (pp. 57–62), Milan, Italy.

Fleming, S. D., Scaffidi, C., Piorkowski, D., Burnett, M., Bellamy, R., Lawrance, J., & Kwan, I. (2013 3). An information foraging theory perspective on tools for debugging, refactoring, and reuse tasks. *ACM Transactions on Software Engineering and Methodology*, *22*(2), 1–41. Retrieved from. http://dl.acm.org/citation.cfm?doid=2430545.2430551

Fogli, D., Peroni, M., & Stefini, C. (2017). Smart home control through unwitting trigger-action programming. *Proceedings of the 22nd International Conference on Distributed Multimedia Systems, 2016*, 194–201.

Getman, A. P., & Karasiuk, V. V. (2014). A crowdsourcing approach to building a legal ontology from text. *Artificial Intelligence and Law*, *22* (3), 313–335. doi:10.1007/s10506-014-9159-1

Girotto, V., Walker, E., & Burleson, W. (2019). CrowdMuse: Supporting crowd idea generation through user modeling and adaptation. *Proceedings of the 2019 on creativity and cognition,* San Diego, CA.

Huang, J., & Cakmak, M. (2015). Supporting mental model accuracy in trigger-action programming. *Proceedings of the 2015 ACM international joint conference on pervasive and ubiquitous computing - UbiComp '15* (pp. 215–225). Retrieved from http://dl.acm.org/citation.cfm?id=2750858.2805830

Jakobi, T., Ogonowski, C., Castelli, N., Stevens, G., & Wulf, V. (2017). The catch(es) with smart home – Experiences of a living lab field study. *Proceedings CHI*, 1620–1633, Denver, CO.

Jang, S. (2014). the effect of image stimulus on conceptual combination in the design idea generation process. *Archives of Design Research*, *112* (4), 59. doi:10.15187/adr.2014.11.112.4.59

Kulkarni, A., Narula, P., Rolnitzky, D., & Kontny, N. (2014). Wish: Amplifying creative ability with expert crowds. *AAAI conference on human computation and crowdsourcing* (Hcomp) (pp. 112–120). Retrieved from http://www.aaai.org/ocs/index.php/HCOMP/HCOMP14/paper/view/8977

Laput, G., Lasecki, W. S., Wiese, J., Xiao, R., Bigham, J. P., Harrison, C., & Boule- Vard, J. C. W. (2015). Zensors: Adaptive, rapidly deployable, human - Intelligent sens or feeds. *Proceedings of the 33rd Annual ACM conference on human factors in computing* systems. *ACM, 2015* (pp. 1935–1944), Seoul, Republic of Korea.

Lasecki, W. S., Kim, J., Rafter, N., Sen, O., Bigham, J. P., & Bernstein, M. S. (2015). Apparition: Crowdsourced user interfaces that come to life as you sketch them. *Proceedings of the ACM CHI'15 Conference on Human Factors in Computing Systems*, *1*, 1925–1934. doi:10.1145/2702123.2702565

Linsey, J. S., Tseng, I., Fu, K., Cagan, J., Wood, K. L., & Schunn, C. (2010). A study of design fixation, its mitigation and perception in engineering design faculty. *Journal of Mechanical Design*, *132*(4), 041003. Retrieved from http://mechanicaldesign.asmedigitalcollection.asme.org/article.aspx?articleid=1450009

Luther, K., Tolentino, J.-L., Wu, W., Pavel, A., Bailey, B. P., Agrawala, M., … Dow, S. P. (2015). Structuring, aggregating, and evaluating crowdsourced design critique. *Proceedings of the 18th ACM conference on computer supported cooperative work & social computing*. Retrieved from http://doi.acm.org/10.1145/2675133.2675283%0Ahttp://dl.acm.org/ftgatewaycfm?id=2675283&type=pdf%0Ahttp://dl.acm.org/citation.cfm?id=2675283

Manning, C. D., Ragahvan, P., & Schutze, H. (2009). An introduction to information retrieval. *Information Retrieval*, (c), 1–18, Cambridge, UK: Cambridge University Press.

Mao, K., Capra, L., Harman, M., & Jia, Y. (2017). A survey of the use of crowd- Sourcing in software engineering. *Journal of Systems and Software*, *126*, 57–84. doi:10.1016/j.jss.2016.09.015

Mennicken, S., & Huang, E. M. (2012, June). Hacking the natural habitat: An in-the-wild study of smart homes, their development, and the people who live in them. In J. Kay, P. Lukowicz, H. Tokuda, P. Olivier, & A. Krüger (Eds.), *Pervasive computing. Pervasive 2012. Lecture notes in computer science, 7319 LNCS* (pp. 143–160). Berlin, Heidelberg: Springer.

Mennicken, S., Vermeulen, J., & Huang, E. M. (2014). From today ' s augmented houses to tomorrow ' s smart homes: New directions for home automation research. *UbiComp '14 proceedings of the 2014 ACM international joint conference on pervasive and ubiquitous computing* (pp. 105–115), Seattle, Washington.

Metaxas, G., & Markopoulos, P. (2017). Natural contextual reasoning for end users. *ACM Transactions on Computer-Human Interaction*, *24*(2), 1–36. Retrieved from http://dl.acm.org/citation.cfm?doid=3077620.3057860

Michalko, M. (2006). *Thinkertoys: A handbook of creative-thinking techniques*. USA: Ten Speed Press.

Mora, S., Gianni, F., & Divitini, M. (2017). Tiles: A card-based ideation toolkit for the Internet of Things. *Proceedings of the 2017 conference on designing interactive systems - DIS '17* (pp. 587–598). Retrieved from http://dl.acm.org/citation.cfm?doid=3064663.3064699

Myers, B. A., Pane, J. F., & Ko, A. (2004). Natural programming languages and environments. *Communications of the ACM*, *47*(9), 47. Retrieved from http://portal.acm.org/citation.cfm?doid=1015864.1015888

Negri, A.-L., Trousse, B., & Senach, B. (2012). Ideation of IoT services with citizen: Coupling GenIoT and AloHa ! methods. *ServDes 2012-service design and innovation conference* (pp. 157–170), Helsinki, Finland.

Nelson, B. A., Wilson, J. O., Rosen, D., & Yen, J. (2009). Refined metrics for measuring ideation effectiveness. *Design Studies*, *30*(6), 737–743. doi:10.1016/j.destud.2009.07.002

Ogonowski, O., Hennes, P., & Seiffert, M. (2016). Shop&play erlebnis im smart home: Nutzung statt installationschaos. In *Workshop usability für die betriebliche praxis,* Bad Honnef, Germany.

Purcell, A. T., & Gero, J. S. (2006). *Design and Other Types of Fixation or Is Fixation Always Incompatible with Innovation?*.

Rahmati, A., Fernandes, E., Jung, J., & Prakash, A. (2017, 9). IFTTT vs. Zapier: A comparative study of trigger-action programming frameworks. *arXiv preprint arXiv:1709.02788*. Retrieved from http://arxiv.org/abs/1709.02788

Riedl, C., Blohm, I., Leimeister, J. M., & Krcmar, H. (2013). The effect of rating scales on decision quality and user attitudes in online innovation communities. *International Journal of Electronic Commerce*, *17* (3), 7–36. Retrieved from http://www.tandfonline.com/doi/full/10.2753/JEC1086-4415170301

Schmitz, H., & Lykourentzou, I. (2018). Online sequencing of non-decomposable macrotasks in expert crowdsourcing. *ACM Transactions on Social Computing*, *1*(1), 1–33. Retrieved from http://dl.acm.org/citation.cfm?doid=3178568.3140459

Siangliulue, P., Arnold, K. C., Gajos, K. Z., & Dow, S. P. (2015). Toward collaborative ideation at scale - Leveraging ideas from others to generate more creative and diverse ideas pao. *CSCW '15* (pp. 937–945). Retrieved from http://dl.acm.org/citation.cfm?id=2675133.2675239

Siangliulue, P., Chan, J., Dow, S. P., & Gajos, K. Z. (2016). IdeaHound: Improving large-scale collaborative ideation with crowd-powered real-time semantic modeling. *UIST* (pp. 609–624), Tokyo, Japan. doi 10.1145/2984511.2984578

Siangliulue, P., Chan, J., Gajos, K. Z., & Dow, S. P. (2015). Providing timely examples improves the quantity and quality of generated ideas. *In Proceedings of the 2015 ACM SIGCHI conference on creativity and cognition* (pp. 83–92), Glasgow, United Kingdom.

Smedema, S. M. (2017). Amazon mechanical turk as a rehabilitation research participant recruitment tool. *Journal of Rehabilitation*, *83*(4), 3–12.

SoGoSurvey. (n.d.). Retrieved from https://www.sogosurvey.com/

Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology*, *29*(1), 24–54. doi:10.1177/0261927X09351676

Thusoo, A., Sarma, J. S., Jain, N., Shao, Z., Chakka, P., Anthony, S., … Murthy, R. (2014). Hive: Collective design through network rotation. *Proceedings of the VLDB Endowment*, *2*(2), 1626–1629. Retrieved from http://dl.acm.org/citation.cfm?doid=3290265.3274420

Ur, B., McManus, E., Pak Yong Ho, M., & Littman, M. L. (2014). Practical trigger- action programming in the smart home. *Proceedings of the 32nd annual ACM conference on human factors in computing systems - CHI '14* (pp. 803–812). Retrieved from http://dl.acm.org/citation.cfm?doid=2556288.2557420

Ur, B., Pak, M., Ho, Y., Brawner, S., Lee, J., Mennicken, S., … Littman, M. L. (2016). Trigger-action programming in the wild: An analysis of 200, 000 IFTTT recipes. *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 3227–3231), San Jose, CA, USA.

Wang, K., Wang, H., & Tao, Y. (2017). Combining ideas in crowd-sourced idea generation. *Foundations of Management*, *9*(1), 203–212. doi:10.1515/fman-2017-0016

Wilson, B. (2005). User-centered design. *D-Lib Magazine*, *11*(1), 1–14.

Yu, L., Kittur, A., & Kraut, R. E. (2014). Distributed analogical idea generation. *Proceedings of the 32nd annual ACM conference on Human factors in computing systems - CHI '14* (pp. 1245–1254). Retrieved from http://dl.acm.org/citation.cfm?id=2611105.2557371

Yu, L., Kittur, A., & Kraut, R. E. (2016). Encouraging "outside-the-box" thinking in crowd innovation through identifying domains of expertise. *Proceedings of CSCW'.16* (pp. 1212–1220), San Francisco, CA, USA.

Yu, L., Kraut, R. E., & Kittur, A. (2016). Distributed analogical idea generation with multiple constraints. *Proceedings of the 19th ACM conference on computer- supported cooperative work & social computing - CSCW '16* (pp. 1234–1243). Retrieved from http://dl.acm.org/citation.cfm?doid=2818048.2835201

Yu, L., & Nickerson, J. V. (2011). Cooks or cobblers? In *Proceedings of the 2011 annual conference on human factors in computing systems - chi '11* (p. 1393). New York. New York, NY: ACM Press. Retrieved from http://dl.acm.org/citation.cfm?doid=1978942.1979147

About the Authors

Tahir Abbas is a PhD candidate at the Industrial Design department of Eindhoven University of Technology, the Netherlands. His research interests include real-time human computation, crowd computing and social robotics. He has a degree in software engineering and is a lecturer, currently on study leave, at Mirpur University of Science & Technology, Pakistan.

Vassilis-Javed Khan is an assistant professor at the Industrial Design Department of Eindhoven University of Technology in the Netherlands. His current research activities focus on the overlap between crowd computing and design. He has held in the past research positions at Philips Research and Vodafone R&D.

Panos Markopoulos is a computer scientist specializing in the field of Human Computer Interaction. He is a professor in Design for Behavior Change at the department of Industrial Design in the Eindhoven University of Technology. His current research concerns designing interactive technologies for rehabilitation and for playful learning.