

Article

A Wargame-Augmented Knowledge Elicitation Method for the Agile Development of Novel Systems

Stephen L. Dorton ^{1,*}, LeeAnn R. Maryeski ¹, Lauren Ogren ², Ian T. Dykens ¹ and Adam Main ¹

¹ Sonalysts, Inc.—Human-Autonomy Interaction Laboratory, Waterford, CT 06385, USA; lmaryeski@sonalysts.com (L.R.M.); idykens@sonalysts.com (I.T.D.); amain@sonalysts.com (A.M.)

² Naval Undersea Warfare Center Division Newport, Newport, RI 02841, USA; lauren.ogren@navy.mil

* Correspondence: sdorton@sonalysts.com

Received: 5 May 2020; Accepted: 10 August 2020; Published: 12 August 2020



Abstract: There are inherent difficulties in designing an effective Human–Machine Interface (HMI) for a first-of-its-kind system. Many leading cognitive research methods rely upon experts with prior experiences using the system and/or some type of existing mockups or working prototype of the HMI, and neither of these resources are available for such a new system. Further, these methods are time consuming and incompatible with more rapid and iterative systems development models (e.g., Agile/Scrum). To address these challenges, we developed a Wargame-Augmented Knowledge Elicitation (WAKE) method to identify information requirements and underlying assumptions in operator decision making concurrently with operational concepts. The developed WAKE method incorporates naturalistic observations of operator decision making in a wargaming scenario with freeze-probe queries and structured analytic techniques to identify and prioritize information requirements for a novel HMI. An overview of the method, required apparatus, and associated analytical techniques is provided. Outcomes, lessons learned, and topics for future research resulting from two different applications of the WAKE method are also discussed.

Keywords: knowledge elicitation; wargaming; requirements; design thinking

1. Introduction

Developing Human–Machine Interfaces (HMIs) for systems is a “bread and butter” task for human factors engineers, or Human Systems Integration (HSI) practitioners. There is no shortage of research methods to elicit information requirements (i.e., what information is needed by end users to make informed and timely decisions); however, they each have their own limitations. The increasingly prevalent use of agile lifecycles in system development drives the need for new human factors methods to enable HSI practitioners to have inputs into the system development process more rapidly. The HMI design process (including the elicitation of users’ information requirements) is even more difficult when developing a first-of-its kind system. We will explore the different challenges in this problem space and review possible approaches and research goals, before describing one such novel human factors method for capturing information requirements for HMI development.

1.1. Problem Space

1.1.1. Extreme Novelty

Designing and developing an effective HMI is an integral part of system development. HMI development is made challenging by the complexity and interconnectedness of modern military systems. Current trends in multi-domain, multi-mission operations drive the need for developing HMIs that are equally effective across multiple, disparate missions, which can be particularly difficult.

HMI development is an especially unique endeavor when the system is a first-of-its-kind prototype. Eliciting and understanding operator information requirements for a prototype is uniquely challenging due to the fact that nobody has used the system, and similarly, there are no Concepts of Operations (CONOPS) for how the system would be used in a relevant operational environment. Because of this, a task analysis alone would be of limited efficacy, because there is no understanding of the exact operational sequence of use. Thus, there is a need to design and develop the operational workflow and HMI in parallel.

For example, imagine that you are trying to design the HMI for the first-ever smartphone. The engineers have a range of features that can be included or excluded depending on size, weight, and power limitations, such as different batteries, a gyro, Global Positioning System (GPS), different screens, keys/controls, cameras, etc., but they want to know how users will employ the smart phone in different use cases before finalizing the system requirements. When you ask some potential end users about how they envision using the system, they tell you that they would first need to know what features will or will not be included. The extreme novelty of the system creates a relatively intractable problem, where the system (including its HMI) and the CONOPS need to be concurrently developed.

1.1.2. Constraints to Traditional Approaches

Developing HMIs by collaborating with end users is not a new concept, and as such, there is no shortage of participative methods that can be used to capture, represent, and exploit expert knowledge [1]. More specifically, Knowledge Elicitation (KE) methods have been used for decades as a means to capture tacit knowledge from domain experts to support the development of knowledge-based systems [2], and have been widely applied across other fields such as psychology, business, management, education, cognitive science, linguistics, philosophy, and anthropology, to name a few [3]. KE is a broad term that can be used to describe any number of structured or semi-structured methods to gather knowledge from experts [2], where each KE method has its own strengths, weaknesses, and appropriate applications [3]. Despite their relative strengths and weaknesses, they all have limitations in the context of this type of system development process (i.e., the rapid design of a prototype system for which there are no existing requirements, designs, or end users/experts).

Klein and Armstrong's Critical Decision Method (CDM) has been widely applied to identify information requirements in the form of a Critical Cue Inventory (CCI) [4–6]. The findings of the CDM can readily be applied to design interfaces that feature information used by domain experts to make critical decisions during the execution of their duties. A significant caveat to using the CDM for this effort is that it requires the expert to verbally recount a situation that has happened. In the case of a first-of-its-kind system, nobody has such an experience using the system. Additionally, techniques such as CDM can require a considerable amount of time to conduct and analyze, which may conflict with a rapid prototyping schedule.

Cognitive Walkthrough (CW) techniques such as the Wharton, et al. Cognitive Walkthrough (WCW) or mockup reviews can also be used to specify information requirements. CWs provide engineers and developers with the opportunity to observe how expert users perform sensemaking and decision making activities as they work to complete specified tasks using paper mockups or a working prototype of a system [6,7]. CWs are advantageous because researchers can explicitly capture information requirements as well as the perceived usability of the system through the use of querying probes that capture the expert's insights and perception of the mockup using standardized feedback forms. However, CWs have a relatively narrow focus on usability [8], therefore, they may be inefficient when there is a greater need for information requirements over usability feedback [9]. Additionally, CWs require existing mockups of the system under investigation, making them inappropriate for this scenario where no such mockups exist.

1.1.3. Agile/Scrum Development and the Fear of Missing out

As the level of complexity and interconnectivity of systems being developed increases, so too does the prevalence of agile processes such as “Scrum.” Scrum is the systems development process of iteratively developing and incrementally delivering software in small durations of time known as “sprints” (usually a month in duration, or less), while incorporating customer and end user feedback with each iteration [10]. The iterative process allows for constant feedback and reassessment, which makes it ideal for development of complex adaptive systems where there is a high degree of interdependency among system components. The merits of such approaches are obvious, and the Government Accountability Office (GAO) has recently issued guidance on adopting such systems development processes for future acquisitions [11].

This new emphasis on rapid development through processes such as Agile/Scrum has created a “Fear of Missing Out” among the HSI community of practice. The primary challenge is that some human factors methods have grown and flourished in an acquisition environment where the waterfall was the de facto systems development lifecycle model. The waterfall (or at least the “pure waterfall”) process involves the structured and ordered sequence of steps (concept development, requirements, design, implementation, test, rework, etc.), where formal reviews mark the end of each step and beginning of the next [12]. Many of the methods described in Section 1.1.2 take months to design, execute, and apply the results towards the development of requirements and design artifacts. While this schedule enabled more than enough time for HSI practitioners to contribute to requirements in a waterfall lifecycle, it results in HSI not having an impact on system requirements or designs until several iterations of the system have already been developed (Figure 1). Simply put, as the world continues to adopt more iterative development processes, the HSI community of practice will need human factors methods that can be more rapidly executed. Without these rapid methods, we will be sidelined until later iterations of system development. We are not advocating replacing these critical human factors methods, merely supplementing them with more rapid methods to enable HSI to have inputs during early iterations of system development.

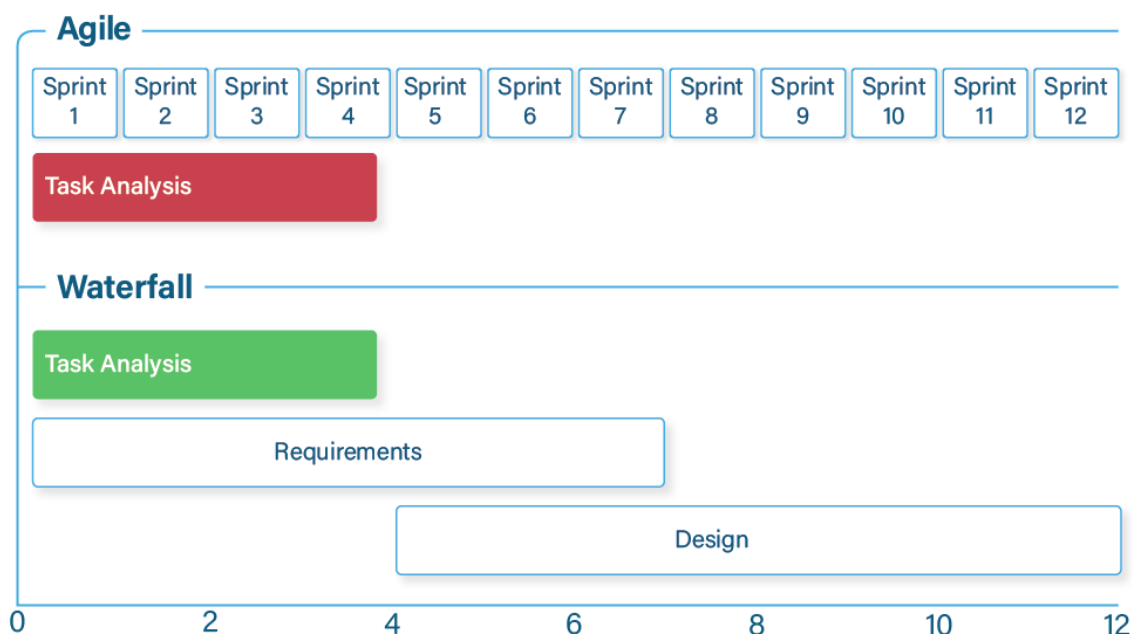


Figure 1. A comparison of how a hypothetical task analysis fits in an agile and a waterfall lifecycle model. While executing a task analysis in four months allows for Human System Integration (HSI) practitioners to be completely involved in the waterfall lifecycle, it would result in HSI having no inputs in the agile lifecycle until the fifth or sixth iteration of the system.

1.2. Possible Approaches

While there are many limitations of traditional user research methods for such a complex scenario (i.e., development of an HMI for an entirely novel system with no user base, and no existing CONOPS), there are a variety of possible constructs and methods that were considered when developing the Wargame-Augmented Knowledge Elicitation (WAKE) method.

1.2.1. Stretched Systems, Resilience, and Learning by Doing

The law of stretched systems is an especially relevant construct when developing a completely novel system with a high degree of interdependency and complexity. In essence, the law of stretched systems states that new capabilities (such as autonomy) do not simply reduce workload of operators, but instead, demand more complex forms of work such as coordinating activities, expanding perception over larger ranges, and projecting intent or goals [13]. Stretched systems is often associated with resilience engineering. In the context of resilience engineering, stretch may also refer to increased task load or mental workload [14]. Resilience engineering is increasingly important as interdependent networks of humans and systems can result in not only unanticipated side effects, but also massive failures with rapid onset and little warning [15].

While there are several means to measure the resilience of a system [16], we are interested in proactively identifying unintended consequences so that we can design a resilient system, rather than measuring the degree to which the system is resilient. The practice of “learning by doing” is one such approach that enables proactive identification of issues with complex systems [17]. Learning by doing has long been used in industrial engineering, where data from machines and interviews with end users of machines were analyzed to reduce the cost of maintenance over time [18]. The general premise is that by recognizing problems before a system is fielded, the likelihood of failure is greatly reduced, and/or the degree to which operators and systems will need to stretch to accommodate a failure. Another notable aspect of learning by doing is that it is exploratory in nature, using a grounded research approach to assess collected data, rather than by testing a priori hypotheses.

1.2.2. Design Thinking

Design thinking is another approach that continues to grow in prevalence along with rapid prototyping processes (e.g., Agile/Scrum). Design thinking can be applied to the development of processes or experiences as easily as it can be applied to designing software, hardware, or integrated systems. The design thinking approach is built upon core tenets, which include focusing on user needs, iterative development and testing, and continuous user engagement throughout the development lifecycle [19]. A general goal of design thinking is to develop systems that are desirable to users while also being technically feasible [20].

Furthermore, design thinking has become recognized for its ability to solve “wicked problems,” or problems of such high codependency that they are essentially intractable [21]. This is accomplished by rapidly generating concepts in divergent activities such as brainstorming, then immediately providing feedback and shaping concepts through convergent activities such as voting and structured feedback. Design thinking represents a conceptual bridge between the aforementioned constructs of iterative development, user-centered focus, and coping with complexity (i.e., resiliency). The tenet of focusing on user needs immediately resonates with HSI practitioners, while the ability to solve complex problems lends itself to the field of resilience engineering.

1.2.3. Wargaming

Wargaming has been defined as a type of simulation that requires human players [22]. Unlike a computer simulation, a wargame generates new knowledge and insights through the interactions of the human players. Further, Perla emphasizes the primacy of the interaction of human decision making and game events, even going so far as to declare that this interaction is the distinguishing feature that

differentiates wargames from other models or simulations [23]. As such, we believe wargaming is closely aligned with the concept of learning by doing, and has intrinsic value in developing novel KE methods. Although wargaming was originally developed as a means to assess military tactics, it has also been adapted for uses in education, economics, corporate decision making, and even for leisure by hobbyists [22].

Recently, methods have been developed to perform KE during low fidelity wargaming events focused on tactics and CONOPS development [24]. These methods are valuable in that they enable the elicitation of information requirements, the understanding of how that information is used in the context of a mission, and the identification of underlying assumptions (which affect the generalizability of findings). A shortcoming of these methods is that the primacy of research is on CONOPS and tactics development, thereby requiring researchers to collect information requirements opportunistically by making inference on the utterances of participants/experts, which is less valid than more directly conducting KE [25].

1.3. Goals

The overarching goal of this research effort was to develop a structured and repeatable KE method that would enable HSI practitioners to identify the decisions operators would need to make across multiple missions, and the pieces of information would be required to support effective operator decision making with a prototypical system. In doing so, such a method would enable the enumeration of requirements, design of system HMIs, and development of tactics and CONOPS for the system. To summarize the various challenges faced and elements of possible approaches to be leveraged, an ideal human factors method would:

- Provide insights without relying upon existing HMIs or experienced operators/users.
- Enable concurrent development and assessment of the HMI, operator workflow(s), and CONOPS for new use cases, all of which are highly interdependent.
- Use probes to directly capture the insights and perceptions of future system operators, and enable assessment of their decision making under various conditions.
- Assess how the human-system team would stretch under various conditions and assumptions, and proactively identify any unintended consequences of use.
- Be planned, executed, and analyzed in a matter of weeks, enabling timely HSI inputs into agile system development lifecycles.

2. Materials and Participants

2.1. Materials

The WAKE method was executed with few materials, all of which are readily available at common office supply stores. For both data collection events described herein, the data collection was facilitated with sticky notes of two different colors, permanent markers, and larger canvas-style (2 ft × 3 ft) sticky notes. The majority of the required materials were used to facilitate the wargaming scenario, including the game board, markers, and dice, which are readily available for a total less than USD 150.00 at an office supply store [24]. Digital game boards can also be used if the complexity of the scenario warrants it; however, this research was largely exploratory rather than confirmatory, making the analog paper tabletop game more appropriate [26].

WAKE can be conducted with almost any simulator or model that one may have, so long as it lends itself to having multiple participants observe the status or state-of-play, and provide their inputs to the system (thus facilitating the gameplay and KE process). It is generally encouraged to adopt the “targeted fidelity” construct when facilitating WAKE events, which simply means that simulations should only have as much fidelity or resolution as needed to enable primary research objectives [27]. Applied in the context of WAKE, one may only need fidelity in simulating the ranges or performance of sensors or weapons (which can be done with paper look up tables and dice rolls), and will likely not

need a high resolution geospatial picture. Furthermore, such high fidelity visualizations may not only distract participants and facilitators from the phenomenology of interest, but participants have also been known to erroneously place confidence in their work with flashier high fidelity systems, despite being unwarranted by their task performance [28].

Additionally, Audio/Video (AV) recording and still photography were used to document gameplay and participant commentary. These artifacts prove exceptionally useful when conducting analyses following the event, as the WAKE method typically generates a deluge of context-rich discussion from all parties, which can be difficult to accurately recall afterwards.

2.2. Participants

Participant group sizes ranged from six to eight participants across both events, which were larger than the intended user base (i.e., the system may be operated by three or four people). The larger samples were used to enable a diversity of thought on how to employ the system during operational use. Participants were recruited with varying levels of expertise, which was representative of the type of team that would use the prototype system (i.e., actual system employment would involve a mix of junior and senior personnel). Because WAKE is an applied method, emphasis was placed on participant representativeness as a means to increase the external validity of the analysis [29].

In use cases where there is not a need for more junior participants, we recommend recruiting participants with greater expertise in analogous missions, systems, or technical domains. Generally, experts (i.e., those with greater experience and tacit knowledge in a particular domain) have more developed mental models than novices, and can reliably identify critical cues or information in relatively novel situations that novices cannot [30]. Similarly, experts can notice the absence of cues, making them excellent at asking questions that identify assumptions being made. Expertise is valuable in identifying information requirements and assumptions in novel situations that WAKE is designed to address.

3. Methods

The following sections describe the resultant WAKE method that blends naturalistic observations, freeze-probe queries, and design thinking techniques. WAKE can be used to rapidly identify and prioritize information requirements for the development of prototypical HMIs. WAKE was deployed in two different events, where each event was dedicated to investigate how operators would use a first-of-its-kind system in a different mission. The WAKE process took between 4–6 h to execute depending on the complexity of the specific mission or scenario being examined, and was performed using a basic tabletop setup and process described in detail in Reference [24].

We present WAKE in the following sections first as a generic process in order to reinforce its wide applicability across military and non-military domains. However, in order to provide clarity to the reader and ground the WAKE method in practice, we also provide examples of how WAKE might work using the example mentioned in Section 1.1.1. For this use case, we assume that the system being designed is the first ever smart phone, and the mission being wargamed is to conduct your evening commute after working at an offsite meeting in a part of town you are unfamiliar with, while picking up pizzas from a local restaurant, and then throwing a party for 10+ people in your backyard.

3.1. WAKE Data Collection Procedure

The setup process prior to beginning data collection was quick and simple. Facilitators simply labeled and hung a large canvas sticky note for each major phase of the mission on available wall space where the event took place. For both events, the wargame was broken into three distinct mission phases that represented how the system would likely be employed. Although these scenarios used a three-phase design for the specific operational scenarios, any number of mission phases can be used to match the type of operations being performed, such as the Find, Fix, Track, Target, Engage, and Assess (F2T2EA) construct [31], or the Navy strike mission planning cycle [32]. A fourth canvas

labeled “Parking Lot” was hung on the wall to serve as a means to capture ideas that were outside the scope of the current wargame, but germane to the overall program. Whenever facilitators felt discussion had veered off topic, the idea was captured and placed in the “parking lot” to keep the sessions on track. Facilitators had two different colors of sticky notes (one color for information and another color for assumptions) and permanent markers to enable the rapid capture of data. For the provided example, the major phases might be navigating home from the offsite meeting, picking up the pizza, and throwing the party itself. There are multiple tasks to be performed in each of these major “mission” phases; however, they represent three large, temporally-distinct parts of the mission.

Participants were provided an orientation briefing prior to commencing gameplay. This briefing included content such as:

- An overview of the gameboard, the various geospatial boundaries and features, and their significance.
- A review of the assumed capabilities and limitations of the system (characteristics such as range, power, and reliability of different system components).
- Rules for gameplay (the length of each turn, the scale of the map, when and how information can be requested from the game master).

The vast majority of each WAKE event was dedicated to the documentation of assumptions and information requirements that were specified by participants playing the wargame. For each of the three mission phases described above there were two periods of data collection: a gameplay period and a freeze-probe KE period (shown in Figure 2). During gameplay, participants would think aloud and collaborate as a group to accomplish the mission. It was common, and often necessary, for participants to pose questions and request clarifying information from the game master (i.e., the person administering the game who had “ground truth” knowledge of the scenario) to inform their sensemaking and decision making processes. During gameplay, facilitators passively took notes without interrupting gameplay on what information was being requested, and how it was used by participants. These findings (information requirements) were captured on sticky notes that were then placed on the canvas that corresponded to the mission phase being played. In addition to capturing these information requirements, facilitators captured any assumptions that were made by the participants or the game master (e.g., environmental conditions, sensors and weapons available to players, enemy capabilities and intents). Assumptions were captured on sticky notes of an alternate color (to denote they are assumptions, and not information requirements), and placed onto the appropriate canvas based on the mission phase.

In the smartphone example, participants may say that they would need information such as maps (to navigate an unknown area), traffic status (to choose the best route), the phone number of the pizza parlor (to make the order), and weather forecasts (to determine if the party can be outside or inside). During the initial phase of gameplay, we might elicit some assumptions such as there being a highway from the meeting to the pizza parlor (not requiring much navigation), or that the phone’s battery only had a 50% charge when the mission began, or the general levels and pace of traffic.

Throughout the gameplay period we subjected the participants to a variety of stimuli based on the overarching goals of the research. Stimuli such as system failures or changes in the environment were injected during gameplay to identify what information participants need to effectively adapt to system anomalies or novel ecological conditions. For example, we might inject a flat tire into the gameplay to understand whether participants would use the smartphone to call for help, or if they would revert to asking a passerby for assistance.

When the participants reached the end of each mission phase, facilitators would pause the scenario and conduct a freeze-probe KE activity. First, the facilitators would perform a review of what information requirements and assumptions were captured during gameplay to ensure the accuracy and completeness of the collected data. Verifying the accuracy of captured information requirements assisted the researchers in multiple ways. First, it ensured that the researchers did not unintentionally mischaracterize the specific articles of information, while verifying that no domain specific acronyms

and slang were incorrectly defined. Furthermore, the verification process allowed researchers to understand the rationale behind why each piece of required information was needed. After reviewing the captured requirements, participants were asked to further think and discuss what additional information would generally aid in decision making that may not have been expressed due to an idiosyncrasy of the particular mission that was being gamed. For example, we may probe participants on why they chose to hail a passerby to help with their flat tire, and come to find out that they would have used the smartphone to call for help if the weather was inclement.

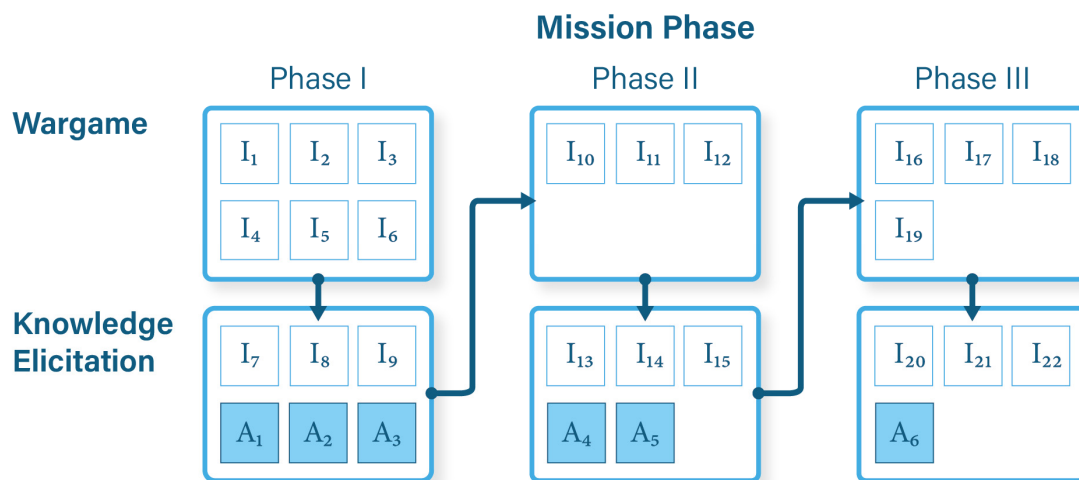


Figure 2. Overall flow of a Wargame-Augmented Knowledge Elicitation (WAKE) Event. For each mission phase there is a wargaming period where information requirements (I_x) and assumptions (A_x) are captured through naturalistic observation by the facilitators. After wargaming, a dedicated KE period is used to validate captured information, and further elicit other information and assumptions from participants. The arrows indicate the temporal flow of data collection, and are not meant to indicate that information and assumptions flow from one phase to another.

Researchers may choose to use a third color of sticky notes to differentiate between observed information requirements (from gameplay) and elicited information requirements (from the freeze-probe KE); however, this approach was not used for this particular application. This process of gameplay and KE was repeated for each mission phase until the wargame ended.

3.2. WAKE Analysis Procedure

Once the scenario and data collection was complete, we then guided participants through a series of participatory analytical activities (i.e., activities where the participants or end users conduct the analysis, not the research team). These activities involved having the participants physically move sticky notes (resembling pieces of information) and placing them on different templates to help characterize the information in some way. These methods resemble classic card sorting techniques, which have been successfully implemented for decades across commercial and academic applications [33]; however, card sorting is limited in that it simply uses semi-tacit knowledge to categorize or bin items on a single criterion (usually conceptual proximity) [34]. We overcame this limitation by employing a series of simple sorting activities that allowed us to characterize the cards (i.e., pieces of information) across numerous nominal and ordinal criteria.

Furthermore, participants were asked to think aloud (i.e., verbalize whatever crossed their mind) during each of these activities to enable the research team to gain insights towards their reasoning processes. Such think aloud protocols have been widely used in conducting KE for decades, due to their simplicity and effectiveness in accessing the thoughts of users during task performance [35,36]. This approach also promoted diversity of thought across the participants, as they were able to build upon (or refute) various casual thoughts verbalized by others, which might not have occurred if they were simply asked to perform the sorting task without the think aloud protocol. Because each activity concluded with a final placement of sticky notes and a verbal explanation, it implicitly forced the group to come to a consensus. Participants did not always wholly agree during these activities; however, the think aloud approach allowed the research team to capture these diverse, and sometimes dissenting, opinions.

For the first participatory analytical activity, participants were asked to sort information requirements based on what medium they would prefer to receive each piece of information (Figure 3). On each canvas (one for each phase), participants sorted information into three groups based on the following criteria:

- HMI: Information that participants would like to access natively through the HMI being developed.
- Verbal: Information that participants would like to access by verbal communications with the crew, rather than through an HMI.
- Other: Information that participants would like to access through another system or interface, rather than through the HMI or verbally from someone else.

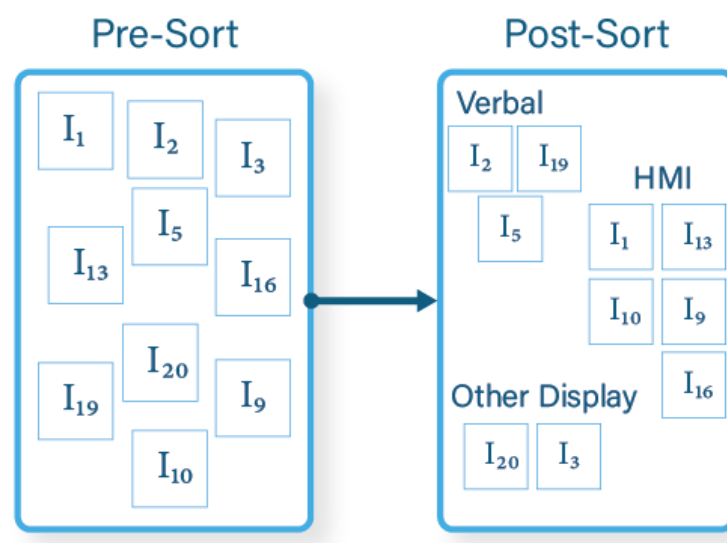


Figure 3. Information sorting process. Three categories are shown, but participants can allocate information to an arbitrary number of sources depending on the context of the system being investigated.

This simple analytical process enabled the research team to better understand what information requirements exist for the HMI. There was a general desire from participants to integrate disparate systems into the new HMI to enable rapid decision making. However, this process unveiled specific pieces of information that, for a variety of reasons (such as preserving screen space), they would prefer to keep on a separate console. Identifying early on that operators would prefer specific information to not be on their HMI enabled the development team to save considerable time and resources by not developing unwanted features. Still photography was used to document the results of this task before undertaking the next.

Facilitators then provided participants with another canvas, which was marked with two axes (Figure 4). Participants were asked to each select the two most important pieces of information from any of the canvases (i.e., across all mission phases), and to place them on the new canvas with the two axes. Participants were asked to conduct the “2 × 2 Matrix” method [37], where they would place the sticky notes representing information requirements on the canvas corresponding to their relative frequency of use (along the X-axis) and their relative criticality (along the Y-axis), while verbalizing their reasoning. This analytical method enabled the researchers to rapidly gain an understanding of the relative priorities of each piece of information required for sensemaking and decision making in a successful mission. The resultant placements of sticky notes were captured with still photography for later analysis and reporting. This method can also be used to characterize the assumptions identified during WAKE, where the axes are simply reworded such that the criticality (Y-axis) represents the degree to which the assumption would change the real-world outcome if the assumption did not hold, and the frequency (X-axis) is re-worded to “likelihood,” and represents the likelihood of each assumption holding true in a real-world scenario.

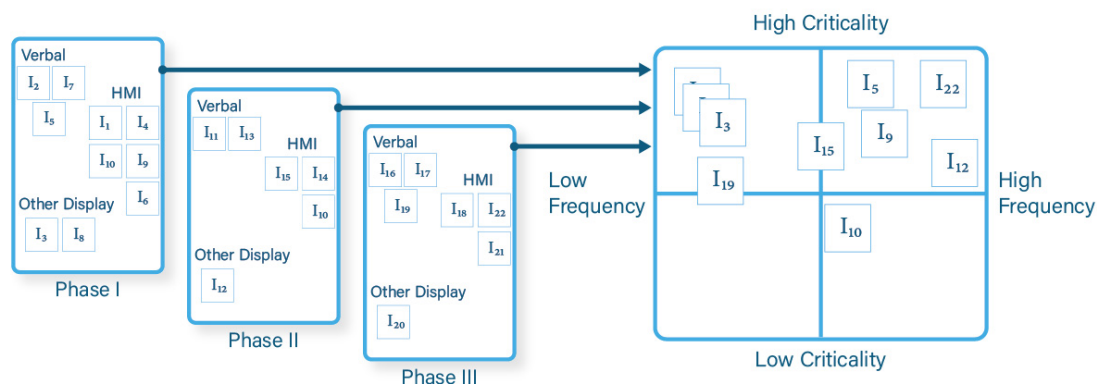


Figure 4. Information characterization process. This analytic technique allows for prioritization of each information requirement, which directly informs Human–Machine Interface (HMI) design decisions (i.e., more important and/or frequently used information is commonly made more prominent).

In the smartphone example, participants may indicate that information such as the weather forecast is of the highest criticality (it informs traffic patterns and whether or not the party can be hosted outdoors), but needed infrequently as it is unlikely to change in the duration of a drive. Conversely, traffic information is of moderate criticality, but might be needed very frequently as it is subject to regular change.

The final analytical method can be performed after the event itself, without the involvement of the participants. Facilitators simply create a large Venn diagram (on a table with tape, or on a large white board) with circles corresponding to each phase of the mission. Then, the sticky notes containing information requirements (and assumptions, if desired) are placed in corresponding circles to show what information is required across multiple mission phases (Figure 5).

The outcome of this analysis is a mapping of when each piece of information is critical for Operators, which provides valuable insight into the development of requirements and designs for the HMI. This process can also be performed where each circle in the Venn is a different mission/scenario, which aids in identifying which information is universally important across all missions, and which information is only critical in certain mission profiles. In the evening commute example, this method might highlight that weather forecasts are needed early on (Phase I), while traffic updates are needed during all phases.

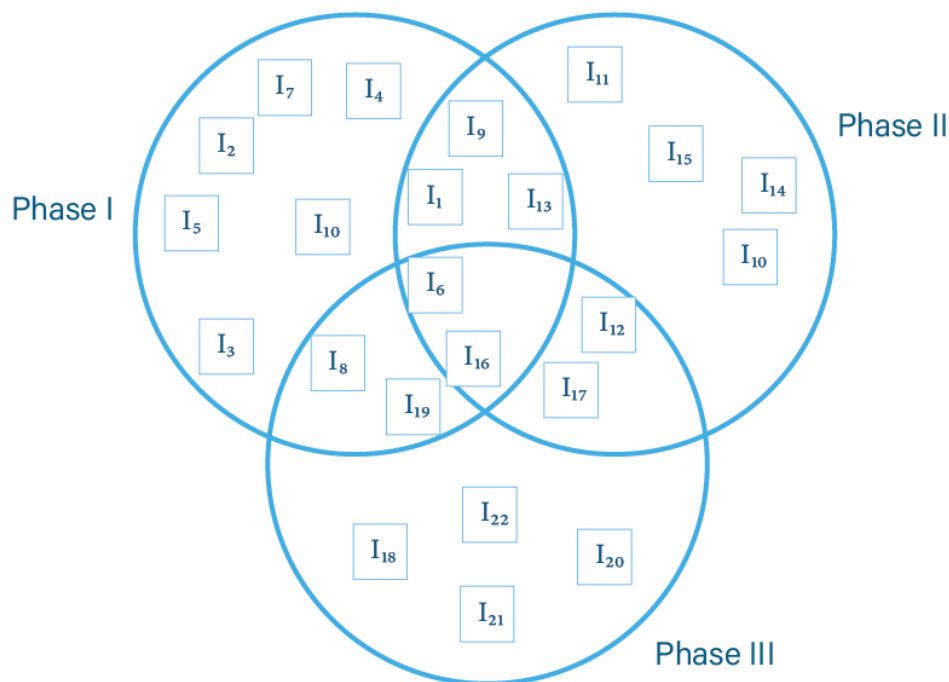


Figure 5. Temporal analysis using a Venn diagram. Three phases are shown in this exemplar; however, any number of mission phases can be used (although complexity increases exponentially with each additional phase).

4. Results and Conclusions

The two pilot events as described in Section 3 have enabled the development of a scalable and repeatable multidisciplinary research approach to understand human decision making, and apply that understanding to rapidly develop HMIs for novel technologies. Furthermore, the scenario-driven approach of WAKE aids the system engineering process above and beyond generating HMI requirements, as the results substantially aid the development of CONOPS and use cases. These points are discussed further in the following sections.

4.1. Outcomes

The WAKE method has provided a unique approach to obtain user inputs up-front in the design of a novel HMI. The transfer of knowledge throughout the process is captured via multiple methods, and is synthesized into documentation that is quantifiable and usable by system developers. Typical documentation generated from user interviews is often poorly understood by engineers and management, as it often consists of identified themes, frequencies of supporting or dissenting comments, and direct quotes to provide context. However, the wargaming approach provides a contextual overview in the more readily understandable format of a story. The documentation from WAKE events includes turn-by-turn imagery of the game board (i.e., the positions of all entities) and major decisions and outcomes made at each turn (e.g., maneuvering or employing various systems). Furthermore, we included other information captured via passive note taking and recordings, including other courses of actions considered, but not taken. These other candidate ideas and discussion greatly increased the generalizability of findings across other scenarios, and provided engineers with valuable context as to *why* users need specific pieces of information.

This process generates information in a format that is easily translated into user-centric requirements in the system's requirements documentation and technical data package. It was reported that documentation from the WAKE events became the preferred onboarding material to orient new engineers and developers on the project, engendering a Shared Mental Model (SMM) among

the development team, which aids in such complex tasks that require a SMM to be successful [38]. The development of a method that provides direct input from users to engineers has reduced the overall design time and streamlined the HSI input into the requirements process.

Additionally, WAKE has created an opportunity to inform other areas of the system design process. The scenario-based tabletop wargaming enables discovering edge cases that may not have been identified by users or engineers, either through the decisions made in the game itself, or other ideas captured in the parking lot. These edge cases were used to drive HMI mockups which were ultimately included in the HMI design and later tested with users. As a result, user testing was not hindered by unfortunate discovery of edge cases/conditions, since they were previously identified during WAKE events.

Finally, the overall outputs of the WAKE events provided direct input into how the HMI will be used in the field. The early generation of scenario-based usage documentation (i.e., early testing and analysis of CONOPS) before the HMI was developed was critical to a successful system development. Informing the CONOPS from an information and decision making-based approach improved the overall understanding of the HMI for users, engineers, and program managers, and provided alternative perspectives for future HMI operation.

4.2. Participant Engagement

One notable difference between WAKE events and more traditional user research methods (e.g., task analysis, CDM, or CW) is the degree to which participants are engaged and enjoy the process of participating. On multiple occasions participants voluntarily informed researchers of their schedules and asked when the next event would be. One participant confided that they were upset they were being assigned to a new unit, and would no longer be able to participate in future events for the project. This type of genuine, intrinsic user engagement is beneficial in a variety of ways. In the simplest sense, it combats participant (and data) attrition—once the program was established, more people wanted to participate than there was room to accommodate. At a larger level, it creates a true sense of ownership in the project, which results in developing a cadre of end user champions to advocate for the resultant system.

Surveys were conducted with participants ($N = 50$) to assess perceptions on a variety of topics, including interpersonal dynamics and benefits of the toolset and process that preceded the WAKE method (i.e., a similar research method/event that was focused on testing novel concepts/tactics prior to codifying the WAKE method). Although the primacy of the survey was focused on assessing the efficacy of different tools to facilitate the wargaming process [26], the results are indicative of perceptions toward the wargame-driven user research process, more generally. The surveys contained a series of positively-oriented statements where participants would mark their level of agreement on a Likert scale from 1 = “Strongly Disagree” to 5 = “Strongly Agree” (3 = “Indifferent To”). As shown in Table 1, participants in the WAKE precursor events consistently agreed or strongly agreed with each statement regarding personal or group dynamics, as well as the benefits of the tools and process itself.

Table 1. Survey Results Regarding the WAKE Precursor.

Positively-Oriented Survey Statement	<i>M</i>	<i>SD</i>	Agreement Level
... Promotes discussion and critical thinking	4.68	0.47	Strongly Agree
I feel involved in gameplay during ... sessions	4.50	0.58	Strongly Agree
The entire group gets involved in playing ...	4.00	0.93	Agree
I have fun and enjoy the ... sessions	4.52	0.65	Strongly Agree
... helps uncover and discuss assumptions	4.54	0.50	Strongly Agree
... helps identify spatiotemporal issues with concepts	4.38	0.64	Agree
... supports determining effectiveness of concepts	4.36	0.69	Agree
... is a useful tool for developing advanced concepts	4.54	0.65	Strongly Agree

Most notably, participants strongly agreed that they had fun and enjoyed participating in the sessions. Additionally, participants found that the tool (and method) was useful for developing advanced concepts. There was a significant positive correlation between agreement that the sessions were fun and agreement that the sessions were useful ($R_S = 0.66$, $p < 0.01$) and that sessions promoted discussion and critical thinking ($R_S = 0.57$, $p < 0.01$). In other words, those who found the WAKE precursor event to be fun likely did so because of the associated discussion and feeling of accomplishment.

4.3. Limitations and Future Work

Several lessons learned and areas of future research were identified after employing WAKE across the two different events, including:

- When assessing a new system or technology, participants may need prompting or “nudging” to arrive at the capabilities and limitations of the new system, and not existing analogical systems. Future research and employment of WAKE will explore priming participants to think about the new technological capabilities and limitations in an interview phase before the wargaming event.
- Intermittent questions and probes by the facilitators were helpful in keeping discussion on track during gameplay.
- The assumptions and Parking Lot canvases were useful not only in capturing information, but as a means to constrain superfluous conversation and maintain focus.
- Some social loafing was observed among participants, which may be expected as teams approach or exceed 10 members [39].
- The participative analytical methods are, by their team-oriented nature, susceptible to the adverse effects of group think and mutual influence among participants. This can be exacerbated with military participants where rank is a substantial factor in team dynamics (i.e., junior participants are less likely to contradict senior participants). We managed these risks through proactive facilitation (e.g., asking the most junior participant to start the analysis or present the findings to the group). Future work will investigate how to apply one or more existing methods or tools to gather individual votes or priorities [37,40].
- The 2×2 method had limited diagnosticity, since several required pieces of information were all-or-nothing criticality (i.e., the mission is guaranteed to fail without them), causing an n -way tie at the top of the board. Future work will look at alternative methods to prioritize information requirements to avoid such issues, such as implementing a forced-choice rule where no two pieces of information can be parallel to each other on one or more axis.
- Data analysis can be time intensive beyond reporting enumerated information requirements (i.e., reviewing video footage and transcribing data). New methods for collection, processing, and exploitation of data will be investigated for future applications of WAKE. Such technologies may include, but are not limited to, speech-to-text transcription and digital sticky notes through a touch screen display or interactive projector (obviating the need for transcription).

Author Contributions: Contributions including the following: Conceptualization, S.L.D., L.R.M., L.O.; Data curation, L.R.M., S.L.D., I.T.D., A.M.; Investigation, S.L.D., L.R.M., L.O., I.T.D., A.M.; Methodology, S.L.D., L.R.M., L.O.; Project administration, L.O., L.R.M., S.L.D.; Visualization, A.M., S.L.D.; Writing—original draft, S.L.D., I.T.D., L.O., L.R.M.; Writing—review and editing: S.L.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: The authors would like to thank our colleagues at the Naval Submarine Medical Research Lab (NSMRL) for their collaboration in conducting both WAKE events.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Sinclair, M.A. Participative assessment. In *Evaluation of Human Work*, 3rd ed.; Wilson, J.R., Corlett, N., Eds.; Taylor and Francis: Boca Raton, FL, USA, 2005; pp. 83–112.
2. Shadbolt, N.; Burton, A.M. The empirical study of knowledge elicitation techniques. *SIGART Bull.* **1989**, *108*, 15–18. [[CrossRef](#)]
3. Cooke, N.J. Varieties of knowledge elicitation techniques. *Int. J. Hum.-Comput. Stud.* **1994**, *41*, 801–849. [[CrossRef](#)]
4. Klein, G.; Armstrong, A.A. Critical decision method. In *Handbook of Human Factors Ergonomics Methods*; Stanton, N.A., Hedge, A., Brookhuis, K., Salas, E., Hendrick, H., Eds.; CRC Press: Boca Raton, FL, USA, 2005; pp. 58:1–58:6.
5. Shadbolt, N.R.; Smart, P.R. Knowledge elicitation: Methods, tools, and techniques. In *Evaluation of Human Work*; Wilson, J.R., Sharples, S., Eds.; CRC Press: Boca Raton, FL, USA, 2015; pp. 174–176.
6. Stanton, N.A.; Salmon, P.M.; Walker, G.H.; Baber, C.; Jenkins, D.P. Cognitive task analysis methods. In *Human Factors Methods: A Practical Guide for Engineering and Design*; Stanton, N., Salmon, P., Eds.; Ashgate Publishing: Surry, UK, 2005; pp. 75–108.
7. Polson, P.G.; Lewis, C.; Wharton, C. Cognitive walkthroughs: A method for theory-based evaluation of user interfaces. *Int. J. Man-Mach. Stud.* **1992**, *36*, 741–773. [[CrossRef](#)]
8. Wharton, C.; Rieman, J.; Lewis, C.; Polson, P. The cognitive walkthrough method: A practitioner's guide. In *Usability Inspection Methods*; Nielsen, J., Mack, R., Eds.; Wiley: Hoboken, NJ, USA, 1994. Available online: <http://www.colorado.edu/ics/technical-reports-1990-1999> (accessed on 11 August 2020).
9. Rogers, Y.; Sharp, H.; Preece, J. Evaluation: Inspections, analytics, and models. In *Interaction Design: Beyond Human-Computer Interaction*; Rogers, Y., Ed.; John Wiley and Sons: West Sussex, UK, 2013; pp. 514–517.
10. Doshi, H. *Scrum Insights for Practitioners: The Scrum Guide Companion*; Hiren Doshi: Lexington, KY, USA, 2016.
11. Ludwigson, J. *DOD Space Acquisitions: Including Users Early and Often in Software Development Could Benefit Programs*; GAO-19-136; Government Accountability Office (GAO): Washington, DC, USA, 2019; pp. 1–55.
12. McConnell, S. *Rapid Development: Taming Wild Software Schedules*; Microsoft Press: Redmond, WA, USA, 1996.
13. Woods, D.D. The law of stretched systems in action: Exploring robotics. In *Proceedings of the 2006 ACM Conference on Human-Robot Interaction, HRI 2006*, Salt Lake City, UT, USA, 2–3 March 2006. [[CrossRef](#)]
14. Sheridan, T.B. Risk, human error, and system resilience: Fundamental ideas. *Hum. Factors* **2008**, *50*, 418–426. [[CrossRef](#)] [[PubMed](#)]
15. Woods, D.D. Four concepts for resilience and the implications for the future of resilience engineering. *Reliab. Eng. Syst. Saf.* **2015**, *141*, 5–9. [[CrossRef](#)]
16. Hoffman, R.R.; Hancock, P.A. Measuring resilience. *Hum. Factors* **2017**, *59*, 564–581. [[CrossRef](#)] [[PubMed](#)]
17. Arrow, K. The economic implications of learning by doing. *Rev. Econ. Stud.* **1962**, *29*, 155–173. [[CrossRef](#)]
18. von Hippel, E.; Tyre, M. How “learning by doing” is done: Problem identification in novel process equipment. *Res. Policy* **1995**, *24*, 1–12. [[CrossRef](#)]
19. Brown, T. *Change by Design*; Harper Business: New York, NY, USA, 2009.
20. IDEO—Design Thinking. Available online: <http://www.ideo.com/pages/design-thinking> (accessed on 5 February 2020).
21. Liedtka, J. *Evaluating the Impact of Design Thinking in Action*; Darden Working Paper Series; Darden School of Business: Charlottesville, VA, USA, 2018; pp. 1–48.
22. Oriesek, D.F.; Schwarz, J.O. *Business Wargaming: Securing Corporate Value*; Gower Publishing: Aldershot, UK, 2008.
23. Perla, P.P. *The Art of Wargaming*; Naval Institute Press: Annapolis, MD, USA, 1990.
24. Erwin, M.; Dorton, S.; Tupper, S. Rolling the dice: Using low-cost tabletop simulation to develop and evaluate high-value CONOPS in uncertain tactical environments. In *Proceedings of the MODSIM World 2017*, Virginia Beach, VA, USA, 25–27 June 2017.
25. Johnson, R.B. Examining the validity structure of qualitative research. *Education* **1997**, *118*, 282–292.
26. Dorton, S.; Tupper, S.; Maryeski, L. Going digital: Consequences of increasing resolution of a wargaming tool for knowledge elicitation. *Proc. 2017 Int. Annu. Meet. HFES* **2017**, *61*, 2037–2041. [[CrossRef](#)]
27. Morabito, T. Targeted fidelity: Cutting costs by increasing focus. In *Proceedings of the MODSIM World 2016*, Virginia Beach, VA, USA, 26–28 April 2016.
28. Smallman, H.S.; Cook, M.B.; Manes, D.I.; Cowan, M.B. Naïve realism in terrain appreciation: Advanced display aids for routing. *Proc. 2008 Int. Annu. Meet. HFES* **2008**, *52*, 1160–1164. [[CrossRef](#)]

29. Bracht, G.H.; Glass, G.V. The external validity of experiments. *Am. Educ. Res. J.* **1968**, *5*, 437–474. [CrossRef]
30. Klein, G. *Sources of Power: How People Make Decisions*, 20th Anniversary ed.; MIT Press: Cambridge, MA, USA, 2017.
31. Tirpak, J.A. Find, Fix, Track, Target, Engage, Assess. *Air Force Mag.* **2000**, *83*, 24–29. Available online: <http://www.airforcemag.com/MagazineArchive/Pages/2000/July%202000/0700find.aspx> (accessed on 11 August 2020).
32. Menner, W.A. The Navy's tactical aircraft strike planning process. *Johns Hopkins APL Tech. Digest* **1997**, *18*, 90–104.
33. Burton, A.M.; Shadbolt, N.R.; Rugg, G.; Hedgecock, A.P. The efficacy of knowledge elicitation techniques: A comparison across domains and levels of expertise. *Knowl. Acquis.* **1990**, *2*, 167–178. [CrossRef]
34. Fincher, S.; Tenenber, J. Making sense of card sorting data. *Expert Syst.* **2005**, *22*, 89–93. [CrossRef]
35. Jaaskelainen, R. Think-Aloud Protocol. In *Handbook of Translation Studies*; Gambier, Y., van Doorslaer, L., Eds.; John Benjamins Publishing Company: Amsterdam, The Netherlands, 2010; pp. 371–373.
36. Cooke, L. Assessing concurrent think-aloud protocol as a usability test method: A technical communication approach. *IEEE Trans. Prof. Commun.* **2010**, *53*, 202–215. [CrossRef]
37. Dorton, S.L.; Smith, C.M.; Upham, J.B. Applying visualization and collective intelligence for rapid group decision making. *Proc. 2018 Int. Annu. Meet. HFES* **2018**, *62*, 167–171. [CrossRef]
38. Harper, S.; Dorton, S. A context-driven framework for selecting mental model elicitation methods. *Proc. 2019 Int. Annu. Meet. HFES* **2019**, *63*, 367–371. [CrossRef]
39. Hackman, J.R. *Collaborative Intelligence*; Berrett-Koehler Publishers: San Francisco, CA, USA, 2011.
40. Van de Ven, A.H.; Delbecq, A.L. The effectiveness of nominal, Delphi, and interacting group decision making processes. *Acad. Manag. J.* **1974**, *17*, 605–621. [CrossRef]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).