2025 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)

MAGIC: A Method for Analyzing the Grammar of Incomplete Cues

Xinyu Hu*
University of Central Florida

Joseph J. LaViola†
University of Central Florida

Ryan P. McMahan‡
Virginia Tech

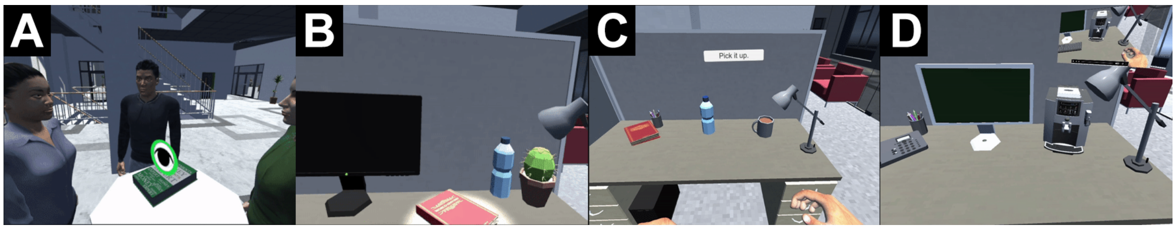

Figure 1: We present a novel method that is capable of analyzing and identifying interaction cues that lack critical information, including: A) who should complete the action (subject-incomplete), B) what action should be taken (predicator-incomplete), C) which object should be acted upon (object-incomplete), and D) how an action should be completed (modifier-incomplete) cues.

Abstract

Augmented reality (AR) and virtual reality (VR) applications commonly employ interaction cues that denote to the user what interaction to take. In this paper, we present a Method for Analyzing the Grammar of Incomplete Cues (MAGIC), which provides an approach for evaluating the design of interaction cues based on the completeness or incompleteness of the functional grammar that they convey through perceptual stimuli. To demonstrate the importance of complete cues, we also present a user study investigating the effects of complete and incomplete cues on which interactions participants choose. The results indicate that incomplete cues do not afford sufficient information, so users make assumptions about the intended interactions. Furthermore, the results indicate that users are more likely to choose intended interactions when the cues are complete. Hence, we present MAGIC as a potentially useful tool for helping interaction designers avoid usability issues with incomplete interaction cues.

Index Terms: Interaction cues, analytical evaluation, AR, VR.

1 Introduction

Interaction cues are crucial for providing guidance to users on what action to take next, particularly for virtual reality (VR) [20] and augmented reality (AR) [12] applications. They are often used in VR tutorials to explain how to interact with the system [22] or in VR training systems to guide trainees through the steps of a procedure [32] or how to practice psychomotor skills [6]. Similarly, interaction cues are frequently used in AR applications to show users how to physically execute real-world tasks [44] or how to navigate real-world environments [3].

Numerous empirical studies have been conducted to investigate and compare various interaction cues. Several studies have investigated interaction cues that guide the user's view to an object outside of their field of view (FOV) [7, 16, 41]. Multiple studies have also investigated guiding the user's attention to an object within their FOV [27, 45]. A few studies have compared interaction cues for tasks requiring the user to manipulate objects [24, 33]. Some studies have even investigated the presentation of multiple cues for multiple interactions in the near future, which is known as "precueing" [28, 29, 47]. Generally, empirical studies of interaction cues take extensive time and effort to design and conduct.

Outside of conducting empirical studies, researchers currently lack a method for assessing the pros and cons of different interaction cue designs. A number of taxonomies or frameworks have been presented for classifying interaction cues based on their visual qualities [12, 14] or other attributes [42, 48, 34]. In addition to classifying interaction cues, these frameworks can be used as design spaces for creating new interaction cues. However, they are not generally capable of analyzing the potential effectiveness of a cue design based on the information it conveys (or fails to convey).

In this paper, we present the *Method for Analyzing the Grammar of Incomplete Cues (MAGIC)*. Unlike the prior frameworks, MAGIC is specifically intended to be used as an analytical evaluation method for understanding interaction cue designs and generating hypotheses for empirical testing, in addition to serving as a classification and design tool. MAGIC can help researchers and developers analyze important functional differences between interaction cue designs to avoid lacking important information for completing their intended interactions. Furthermore, MAGIC can help form hypotheses for empirical evaluations of cue designs in advance.

MAGIC relies on the concept of functional grammar roles [46, 17]. It involves defining a text that represents the intended user interaction, identifying the functional grammar roles of that text, mapping any relevant sensory stimuli provided by the interaction cue to those roles, and then assessing which roles are not sufficiently conveyed (i.e., incomplete). MAGIC is capable of categorizing four types of incomplete cues based on four functional grammar roles: A) subject-incomplete, B) predicator-incomplete, C) object-incomplete, and D) modifier-incomplete cues.

After presenting the details of the MAGIC method, we then present an empirical evaluation of incomplete and complete cues to validate the theoretical implications identified through MAGIC. We conducted 40 participants (20 females, 20 males) through an evaluation of the different types of incomplete and complete interaction cues. The results of this study clearly indicate that incomplete interaction cues can have unintended consequences, as users attempt to determine what action to take, given incomplete information. Furthermore, the results indicate that users choose the intended interaction significantly more often when the presented cue is grammatically complete. Altogether, these results provide evidence that MAGIC can be a useful tool for interaction designers by highlighting potential usability issues with incomplete cues.

*e-mail: Xinyu.Hu@ucf.edu
†e-mail: jjl@cs.ucf.edu
‡e-mail: rpm@vt.edu

2473-0726/25/$31.00 ©2025 IEEE
DOI 10.1109/ISMAR67309.2025.00082

2 Related Work

There is substantial prior work that is pertinent to the concept of interaction cues, including perceived affordances [36, 23], feedforward [5, 34], and multimodal communication [4, 15]. As mentioned above, numerous empirical evaluations have been performed that compare a multitude of different interaction cues for a wide range of different tasks and contexts (e.g., [7, 16, 41, 27, 45, 24, 33, 28, 29, 47]). However, the magnitude of prior studies is beyond the scope of this paper, to which we recommend reading a recent systematic review of some interaction cues [40]. Instead, we focus on the prior taxonomies and frameworks that have been developed to classify interaction cues, which are most pertinent to MAGIC.

Dillman et al. [12] presented one of the first frameworks for classifying interaction cues. They developed their framework through the review and analysis of numerous contemporary video games. Their framework consists of three dimensions: purpose, markedness, and trigger. They define the *purpose* of a cue as the interaction being conveyed by the cue. They defined the *markedness* of a cue as how much it stands out in the scene. Finally, they defined the *trigger* of a cue as what causes the cue to appear.

Gattullo et al. [14] presented a taxonomy for classifying visual assets commonly used in industrial AR applications. Their taxonomy is based on the results of a systematic review of the literature. It includes three dimensions or questions: what, how, and why. For their *what* dimension, Gattullo et al. [14] identified eight types of visual assets, such as text, signs, and product models. For their *how* dimension, they described the frame of reference, color coding, and animation qualities of the visual assets. Finally, they classified their *why* dimension into four categories of uses: locating, operating, checking, or warning.

Rothe et al. [42] proposed a taxonomy for attention guidance in cinematic VR experiences, in which the user views 360° content and videos. They defined seven dimensions for attention guiding techniques based on their review of the literature: diegesis, senses, target, reference, directness, awareness, and freedom. While their taxonomy is mainly focused on simply viewing 360° content, some of their classification dimensions are very applicable to other types of VR or AR applications.

Wuertz et al. [48] presented a framework for designing awareness cues for distributed multiplayer games. They identified four types of information provided by awareness cues, including *who*, *what*, *where*, and *how*. In addition to identifying the types of information provided, they also discussed how that information is encoded into the cue, including its *modalities* (e.g., visual or auditory), *consistency* (i.e., whether it changes), and *representation* (i.e., information visualization). They also describe the composition of cues, including their *accessor* (i.e., how the player accesses the cue), *diegesis* (i.e., how the cue relates to the world), *interdependence* (i.e., the relationships between cues), and *attachment* (i.e., placement).

Most recently, Muresan et al. [34] have presented a feedforward framework for VR, which essentially encompasses interaction cues. Their framework is focused on three stages of presenting each cue, including triggering, previewing, and exiting. For *triggering*, their framework addresses the location of trigger relative to the action and outcome, the conditions and mechanisms that trigger previews, and optional signifiers. For *previewing*, their framework addresses playback, level of detail, perspective, rendering, targets, duplication, and representation. Finally, for *exiting*, their framework focuses on how the cue is untriggered and its exit transition.

While the above frameworks are useful for differentiating, classifying, and designing interaction cues, they do not afford the ability to analytically evaluate interaction cues in order to identify potential usability issues. On the other hand, our new method—MAGIC—can assist interaction designers by highlighting potential usability issues with incomplete interaction cues.

3 The MAGIC Method

As mentioned above, our method for analyzing incomplete cues was inspired by the concept of functional grammar [46]. More specifically, it involves identifying the functional grammar roles of an intended user interaction and mapping any relevant perceptual stimuli to those roles. First, we define a text that describes the intended user interaction. Next, we identify the functional grammar roles of that text. We then consider the perceptual stimuli afforded by an interaction cue and identify how the stimuli map to the functional grammar roles using the shared interaction state. Finally, we assess which functional grammar roles of the intended interaction are not addressed by the interaction cue's perceptual stimuli to determine whether, and if applicable, how the cue is incomplete.

Figure 2: After picking up the white bowl, a semitransparent green "ghost" bowl appears to an AR user to indicate where to place the physical bowl on the table.

3.1 Define the Intended Interaction

To demonstrate our method, consider the AR-based interaction cue depicted in Figure 2. First, we must define a text that describes the intended user interaction of the cue. Because interaction cues are essentially commands from the interactive system, we use imperative sentences to define our texts. For this example, we use the following imperative sentence to describe the intended interaction:

$$(You)\ Place\ the\ bowl\ on\ the\ table. \quad (1)$$

It is important to note that MAGIC is robust enough to handle nearly any variation of this text, as long as the variation is functionally equivalent in terms of what the user should do to complete the intended interaction. In fact, the text does not even need to be an imperative sentence. To demonstrate this robustness, we will also consider the following alternative text describing the intended interaction portrayed in Figure 2:

$$The\ user\ moves\ the\ dish\ in\ front\ of\ the\ end\ chair. \quad (2)$$

3.2 Identify Functional Grammar Roles

Next, we need to identify the functional grammar roles of the text representing the intended interaction. Traditional types of functional roles in grammar include subjects, predicators, objects, complements, and adjuncts, which are often represented by the SPOCA acronym [46]. Due to the structural ambiguities of complements and adjuncts [19], we have adopted the modifier functional role instead of the complement and adjunct roles [10]. Hence, the ***SPOM*** acronym represents the functional grammar roles that we use to analyze the texts representing intended interactions:

- A **subject** (***S***) is a noun phrase that indicates an actor (i.e., the person or thing performing an action) [21]. In (1), *(You)* is the subject. In our alternative text in (2), *The user* is the subject. In both, the user meant to perform the interaction is the subject.

- A **predicator (*P*)** is a verb phrase indicating the action being performed by the subject [46]. In (1), *Place* is the predicator, whereas in (2), *moves* is the predicator. While the verbs *place* and *move* can have different meanings, the two verbs are functionally equivalent for this interaction.
- An **object (*O*)** is a noun phrase that indicates the entity that the subject acts upon [21]. In (1), *the bowl* is the object. On the other hand, in (2), *the dish* is the object. Again, while the nouns *bowl* and *dish* can refer to different objects, the two are equivalent for this interaction, with both referring to the physical white bowl.
- A **modifier (*M*)** is any word or phrase that affects the meaning of a sentence [25]. Modifiers can take the form of adjectives, adverbs, or prepositional phrases [37]. Hence, they can affect subjects, predicators, or objects. In (1), *on the table* is a prepositional phrase modifier. Similarly, in (2), *in front of the end chair* is also a prepositional phrase modifier. In both cases, the prepositional phrases are modifying their respective predicators by providing more information on how the action is to be performed, specifically how the predicator affects the object. For this interaction, the modifiers indicate where the bowl should be positioned.

Given these grammar roles, we can now label them in (1):

$$\underset{S_1}{\underline{(You)}}\ \underset{P_1}{\underline{Place}}\ \underset{O_1}{\underline{the\ bowl}}\ \underset{M_1}{\underline{on\ the\ table.}} \tag{3}$$

Note that in addition to labeling the roles, we also use subscripts to number the roles, in order of their appearance. We do this so that our method can handle complex intended user interactions that are best represented by texts with multiple modifiers and sometimes even predicators in the case of simultaneous interactions. Similarly, we can label the roles of the text presented in (2). However, for clarity, we use different subscripts to distinguish the two texts:

$$\underset{S_2}{\underline{The\ user}}\ \underset{P_2}{\underline{moves}}\ \underset{O_2}{\underline{the\ dish}}\ \underset{M_2}{\underline{in\ front\ of\ the\ end\ chair.}} \tag{4}$$

3.3 Map Perceptual Stimuli and Shared Interaction State

For our next step, we first consider all the perceptual stimuli afforded by the interaction cue. As seen in Figure 2, the interaction cue is a semitransparent green bowl that is displayed on the table, in front of the end chair. We need to map this perceptual information to the interaction text and identified functional grammar roles. To do so, we must first consider the shared interaction state.

The **shared interaction state** is a set of information available to both the user and the system about what led to the interaction cue and intended interaction. It includes information about recent interactions, their objects, and their modifiers. It also includes what inputs the user has recently provided to the system (e.g., button presses and gestures) and what outputs the system has provided in return (e.g., feedback and prior interaction cues). For the current example depicted in Figure 2, the shared interaction state includes the recent action of the user picking up the physical white bowl and the semitransparent green "ghost" bowl appearing afterwards.

Now let's map the perceptual stimuli to the functional grammar roles in (3). Considering the perceptual similarities of the two bowls and the shared interaction state that the "ghost" bowl appeared after the user picked up the physical bowl, we can reasonably map the semitransparent green bowl to the white bowl, which is the object O_1. Given the virtual bowl mapped to the physical bowl and considering that the user just moved the white bowl to pick it up, we can reasonably map the placement of the green bowl to indicate that the user should also move the white bowl to the cue (i.e., the predicator P_1). Furthermore, we can reasonably map the positioning of the green bowl to indicate how the white bowl should be placed on the table, which is the modifier M_1. Finally, because the cue appeared in response to the user picking up the physical bowl, we can reason that the cue is intended for the user (i.e., the subject S_1). The same deductions apply for mapping the semitransparent green bowl cue to the functional grammar roles in (4).

To track the mappings of perceptual stimuli and interaction state information to functional grammar roles, we employ notations like:

$$\underset{S_1}{\underline{(You)}}\ \underset{P_1}{\underline{Place}}\ \underset{O_1}{\underline{the\ bowl}}\ \underset{M_1}{\underline{on\ the\ table.}} \tag{5}$$

$S_1 \rightarrow$ *The user picked up the white bowl.*
$P_1 \rightarrow$ *The positions of the white and green bowls are different.*
$O_1 \rightarrow$ *The semitransparent green bowl appeared.*
$M_1 \rightarrow$ *The position of the green bowl on the table.*

Now, let's consider an alternative interaction cue: *a green arrow pointing down in place of the previous green bowl*. Due to the lack of perceptual similarities between the white bowl and green arrow, we can no longer reasonably map the cue to the object O_1. While some users are likely to connect the appearance of the green arrow to picking up the white bowl, many users may not make that connection due to the perceptual dissimilarities. In turn, we can no longer map the placement of the green arrow to indicate that the white bowl should be placed on the table (i.e., the predicator P_1). For example, the system might have displayed the arrow to convey that the user should simply look at the position on the table, not involving any action with the white bowl at all. Alternatively, the system might have displayed the downward arrow to convey that the user should place the bowl upside down on the table. We believe more users are likely to infer that the downward arrow is conveying for the bowl to be placed *down* on the table, as opposed to the bowl being placed *downwards* on the table. Hence, the alternative green arrow cue would be perceptually ambiguous with regards to how the bowl should be placed (i.e., the modifier M_1):

$$\underset{S_1}{\underline{(You)}}\ \underset{P_1}{\underline{Place}}\ \underset{O_1}{\underline{the\ bowl}}\ \underset{M_1}{\underline{on\ the\ table.}} \tag{6}$$

$S_1 \rightarrow$ *The user picked up the white bowl.*
$P_1 \rightarrow$ *Ambiguous.*
$O_1 \rightarrow$ *Ambiguous.*
$M_1 \rightarrow$ *Ambiguous.*

It is also important to note how critical the shared interaction state is to this mapping process. Reconsider the green bowl cue depicted in Figure 2. Now, consider an interaction scenario in which the system expects the user to pick up the white bowl and preemptively displays the semitransparent green bowl. First, it is no longer clear that the user is the intended subject S_1 because the system displayed the cue without any interactions or input from the user. Second, we can no longer reasonably map the green virtual bowl to the white physical bowl (i.e., the object O_1). The system may actually be indicating that the user should find and place a physical green bowl at the indicated position. As such, we can no longer map the cue to the predicator P_1 of just placing a bowl, as it would also imply searching for the bowl. Given the lack of shared interaction state, the only reasonable perceptual stimuli to map is the position of the green bowl to how *a* bowl should be placed on the table:

$$\underset{S_1}{\underline{(You)}}\ \underset{P_1}{\underline{Place}}\ \underset{O_1}{\underline{the\ bowl}}\ \underset{M_1}{\underline{on\ the\ table.}} \tag{7}$$

$S_1 \rightarrow$ *Ambiguous.*
$P_1 \rightarrow$ *Ambiguous.*
$O_1 \rightarrow$ *Ambiguous.*
$M_1 \rightarrow$ *The position of the green bowl on the table.*

3.4 Assess Incomplete Properties

After mapping the perceptual stimuli and shared interaction state, we assess which functional grammar roles of the intended interaction were not sufficiently addressed. For example, in (5), we can see that all of the grammar roles were addressed by the semitransparent green bowl cue depicted in Figure 2. On the other hand, in (6), we note that only the subject is addressed by the alternative green arrow pointing down. Similarly, in (7), we find that only the modifier is sufficiently addressed due to the lack of shared interaction state information. From these examples, it becomes obvious that MAGIC inherently categorizes four types of incomplete cues, one for each functional grammar role.

3.4.1 Subject Incomplete

A **subject-incomplete** cue is any interaction cue that does not explicitly convey the subject (i.e., the user) for the intended interaction. In single-user applications, the user inherently serves as the subject for any given interaction cue. Hence, subject-incomplete cues are not problematic for such single-user experiences. However, in multi-user applications, whether co-located or remote [8], interaction cues displayed to multiple users can be problematic. If only one user is intended to execute the interaction, a subject-incomplete interaction cue displayed to multiple users could result in the wrong user executing the action. On the other hand, if anyone can execute the interaction, users may hesitate to act upon the intended action due to uncertainty about who the subject-incomplete cue is intended for. While most multi-user applications can likely show an interaction cue to only the intended user, it is important to note that team-based VR training or AR guidance applications may need to show such cues to everyone to promote teamwork [31].

3.4.2 Predicator Incomplete

A **predicator-incomplete** cue is any interaction cue that does not sufficiently convey the predicator (i.e., the action) for the intended interaction. Users may deduce what the intended interaction of a predicator-incomplete cue is in some cases. This is evident in our example above with the green arrow pointing down, which could signal to the user that the action is to place the bowl. However, it is not guaranteed that every user will be able to make such a deduction, which could result in the user simply looking at the position on the table. When interaction cues are ambiguous, this can result in critical incidents [18], in which the user does not know how to progress within the application.

3.4.3 Object Incomplete

An **object-incomplete** cue is any interaction cue that does not explicitly convey the object of the intended interaction. Object-incomplete cues are most often the result of an application relying on verbal or textual instructions to convey intended interactions to users [14], which can make the object of the interaction ambiguous (e.g., "Grab the screw" when multiple screws are present).

3.4.4 Modifier Incomplete

A **modifier-incomplete** cue is any interaction cue that does not explicitly convey any additional information required for the intended interaction. For the interaction conveyed in Figure 2, information about where to place the bowl is necessary to successfully complete the interaction. Imagine if the virtual bowl cue was replaced with a text prompting the user to "Place the bowl." While some users may use shared interaction state information (e.g., holding the physical bowl while facing the table) or prior experiences (e.g., bowls are often placed on tables) to infer that the physical bowl should be placed on the table, it is unlikely that many users would know to place the bowl in front of the end chair. Hence, depending on the required accuracy of the intended interaction (e.g., exactly where the virtual bowl is shown in Figure 2 compared to anywhere on the table), modifier-incomplete cues can be problematic for users and cause critical incidents.

3.4.5 Complete

A **complete** cue is any interaction cue that conveys all the information required for the intended interaction. Hence, we consider any interaction cue that is not subject incomplete, predicator incomplete, object incomplete, or modifier incomplete to be a complete interaction cue. However, it is important to note that our definition of a complete cue regards all the necessary information required. Hence, subject-incomplete cues that are complete in all other regards would be considered complete for single-user experiences, but not multi-user experiences. Similarly, it is important to note that not all intended user interactions involve objects and/or modifiers. For example, travel tasks often require modifiers (e.g., a direction or a location [26]) but often do not involve objects. In contrast, selection tasks always involve one or more objects [2] but rarely require modifiers.

4 Analyzing Preexisting Cues with MAGIC

In this section, we provide an example of using MAGIC to analyze the completeness of interaction cues empirically compared within the literature. First, consider the *Image* and *SWAVE* (Spherical Wave) interaction cues investigated by Renner and Pfeiffer [41] and the intended interaction depicted in Figure 3. In their research, Renner and Pfeiffer [41] found that the SWAVE cue afforded significantly faster completion times and less head movement than Image cues. Hence, they determined that SWAVE is a significantly better interaction cue than Image cues.

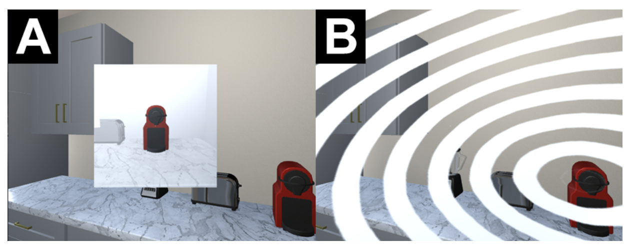

Figure 3: Recreations of interaction cues investigated in [41]: A) an *Image* cue indicates what object to pick, and B) a *SWAVE* cue indicates what object to pick and its direction through concentric circles.

Using MAGIC, we can see that the Image cue is predicator and modifier incomplete when presented to the user:

$$\underset{S_1}{\underline{(You)}}\ \underset{P_1}{\underline{Pick}}\ \underset{O_1}{\underline{the\ coffee\ maker}}\ \underset{M_1}{\underline{on\ the\ right.}} \tag{8}$$

$S_1 \rightarrow$ *The image is presented in the user's field of view.*
$P_1 \rightarrow$ *Ambiguous.*
$O_1 \rightarrow$ *The image of the coffee maker.*
$M_1 \rightarrow$ *Ambiguous.*

On the other hand, by using MAGIC, we can see that the SWAVE cue is only predicator incomplete.:

$$\underset{S_1}{\underline{(You)}}\ \underset{P_1}{\underline{Pick}}\ \underset{O_1}{\underline{the\ coffee\ maker}}\ \underset{M_1}{\underline{on\ the\ right.}} \tag{9}$$

$S_1 \rightarrow$ *SWAVE is presented in the user's field of view..*
$P_1 \rightarrow$ *None.*
$O_1 \rightarrow$ *The circled coffee maker.*
$M_1 \rightarrow$ *The direction of the concentric circles.*

By using MAGIC, we were able to analytically deduce the same conclusion that Renner and Pfeiffer [41] empirically found indicating that the SWAVE interaction cue is better than the Image cue. Furthermore, our MAGIC analysis provides a theoretical foundation as to why. More specifically, it indicates that SWAVE conveys the direction of the object (i.e., the modifier of the intended interaction) while the Image cue does not (i.e., it is modifier incomplete). This example demonstrates the potential usefulness of MAGIC as an analytical tool for interaction designers.

5 User Study of Incomplete and Complete Cues

We now present an empirical evaluation of MAGIC's natural categories to investigate their theoretical implications and to further demonstrate the usefulness and validity of employing MAGIC. We were specifically interested in the following research questions:

Q1 How do users respond when presented with **subject-incomplete** cues in the presence of multiple avatars?

Q2 How do users respond when presented with **predicator-incomplete** cues that clearly indicate objects to act upon but not what action to take?

Q3 How do users respond when presented with **object-incomplete** cues that clearly indicate an action to take but not what object to act upon?

Q4 How do users respond when presented with **modifier-incomplete** cues that clearly indicate an action to take upon an object but not in what manner?

Q5 How do users respond when presented with **complete** cues that clearly indicate an action to take upon an object and how?

The following study was approved by the University of Central Florida's Institutional Review Board (IRB).

5.1 Empirical Online Method

Interaction cues convey intended actions to users via perceptual stimuli and therefore should primarily affect perception and cognition but not motor actions [9]. Hence, we designed an experiment that primarily involved perception and cognition but not motor actions to avoid potential confounds due to motor skill differences among our participants. We developed an experimental method for evaluating interaction cues that consists of presenting a cue to the participant in the form of a video and then presenting multiple videos of potential interactions for the participant to choose from (see Figure 4). This allowed us to investigate the perceptual and cognitive effects of the different interaction cues without confounding the motor capabilities of our participants. Furthermore, it enabled us to conduct our study online, beyond our own university, to make our research more inclusive and accessible [43]. Hence, we chose to use the Prolific platform, which has a more-reliable pool of participants than Amazon's Mechanical Turk [39] and has important features like balancing male and female participants [38].

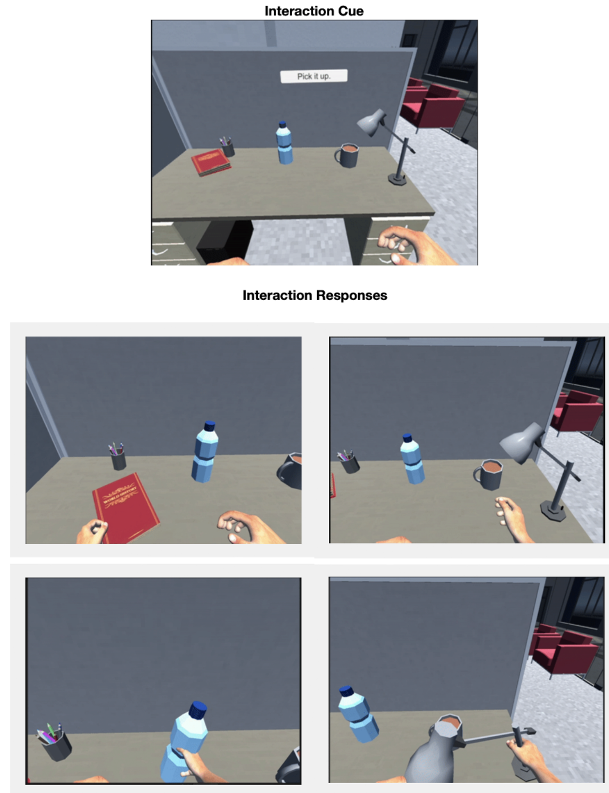

Figure 4: Our empirical online method involved presenting a cue to the participant in the form of a video and then presenting multiple videos of potential interactions for the participant to choose from.

5.2 Experimental Design

With our evaluation method established, we needed to determine the experimental design for each of our cue categories. We developed a total of 60 trials with 12 trials per category of cue (i.e., subject-incomplete, predicator-incomplete, object-incomplete, modifier-incomplete, and complete cues). We discuss the details of these trials for each category below.

5.2.1 Subject-Incomplete Trials

Each of our subject-incomplete trials involved the presentation of a cue within a multi-user scenario, and participants were required to choose one of four possible reactions: 1) *User*—Executing the intended interaction as the user, 2) *Male*—Allowing a male virtual agent to execute the interaction, 3) *Female1*—Allowing a female virtual agent to execute it, or 4) *Female2*—Allowing a second female virtual agent to execute the interaction. In order to make the cues subject-incomplete, we continuously rotated each cue over the interaction's target object such that it faced each of the four users for approximately the same stimulus duration (see Figure 5). Note that the male was not as prominent in every subject-incomplete trials (see Figure 1A). In addition to these reaction types, we decided to investigate three cue types based on the visual assets taxonomy of Gattullo et al. [2022]: 1) *Text*, 2) *Icon*, and 3) *Video*.

Figure 5: Our subject-incomplete cues were continuously rotated to face each person in our multi-user scenarios.

5.2.2 Predicator-Incomplete Trials

Each of our predicator-incomplete trials involved the presentation of a cue clearly indicating an object to act upon with no information about what action to take (see Figure 1B). Again, the participants were required to choose one of four possible reactions: 1) *View*—Looking directly at the object, 2) *Grab*—Reaching out and grasping the object, 3) *Move*—Reaching out and moving the object, and 4) *Use*—Interacting with the object based on its purpose (e.g., turning on a lamp). We again investigated three types of visual cues: 1) *Lighting* the object, 2) an *Outline* shader effect on the object, and 3) an *Arrow* pointing at the object. We chose these cues as we considered them to be inherently predicator incomplete.

5.2.3 Object-Incomplete Trials

Each of our object-incomplete trials involved the presentation of a cue indicating an action, such as "Pick it up." (see Figure 4). However, we carefully positioned each cue ambiguously above four potential objects with respect to the user's view: 1) *FarLeft*, 2) *Left*, 3) *Right*, and 4) *FarRight*. Again, we investigated three types of visual cues: 1) *Text*, 2) *Icon*, and 3) *Video*.

5.2.4 Modifier-Incomplete Trials

Each of our modifier-incomplete trials involved the presentation of a cue clearly indicating an object and the action to perform upon it. However, these cues lacked information about how or to what extent to perform the action, such as what direction to move an object or how bright to turn on a lamp (see Figure 6). For each modifier-incomplete cue, we created four possible reactions for participants to choose from based on the nature of the intended interaction. For our move-oriented cues, participants chose from moving an object backward, forward, left, or right. For our use-oriented cues, participants chose from four increasing degrees of use, such as how bright to turn on a light or how far to open a book to. Again, we investigated three types of visual cues: 1) *Text*, 2) *Icon*, and 3) *Video*.

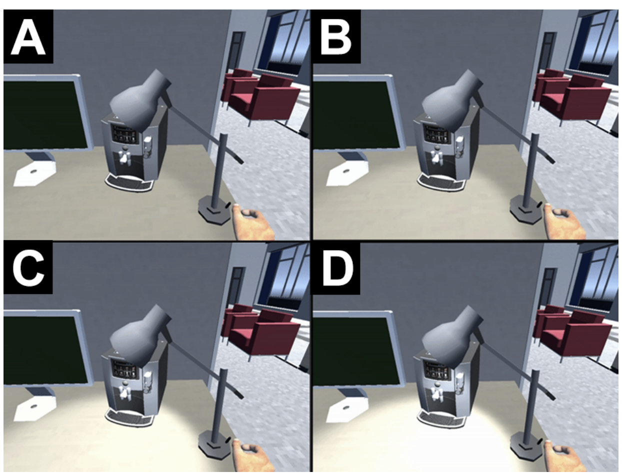

Figure 6: Our modifier-incomplete trials involved choosing different versions of the same action, such as setting the brightness of a lamp to: A) 50%, B) 75%, C) 125%, and D) 150%.

5.2.5 Complete Trials

Each of our complete trials involved the presentation of a cue indicating the subject (for multi-user scenarios), the action to perform, which object to act upon, and any action modifiers. Each of these trials were created by modifying one of the previous incomplete trials to be complete. We took this approach to avoid making our complete trials noticeably different from our incomplete trials, as we did not want participants to mentally distinguish between complete and incomplete trials when responding.

For completing the subject-incomplete trials, we no longer continuously rotated the cue over the target object to face everyone for approximately the same duration, but instead oriented the cue to face one of the non-user avatars (i.e., the intended subject for the interaction). For completing the predicator-incomplete trials, we added an additional *Text* cue directly above the preexisting *Lighting*, *Outline*, or *Arrow* cue. For example, we completed the predicator-incomplete cue shown in Figure 1B with a "Look at it." text prompt positioned immediately above the lighted book. For completing the object-incomplete trials, we used a variety of methods. For the *Text* cue, we changed the text from "Turn it on." to "Turn on computer." For the *Icon* cue, we positioned the icon directly above the water bottle. For the *Video* cue, we switched from showing a user picking up an object not present to picking up a coffee cup that was one of the four choices. Finally, for completing the modifier-incomplete trials, we used a similar variety of methods. We modified the *Text* cue to "Open the book cover." We modified the *Icon* cue to indicate what direction to move the object, and we modified the *Video* cue to show what brightness the lamp was turned on to.

For each complete trial, participants were presented with the same four options as the original incomplete trial. However, for analyses, we coded the intended reaction as the *Correct* option and the other three possible reactions as *Incorrect* options. This also enabled us to compare the frequency that the same option was selected when it was the only correct option for a complete trial to when it was one of four possible reactions for an incomplete trial.

5.3 Procedure

Once recruited through the Prolific platform, participants verified their eligibility to participate, provided their informed consent, and answered demographic questions. Participants were then presented the 60 trials in random order. We did this to avoid participants realizing the different categories of cues, which could influence their decision making. Additionally, the responses for each trial were randomly presented to avoid potential ordering effects. Participants took approximately 17 minutes on average to complete the online study and were compensated $5 USD via Prolific.

5.4 Participants

A total of 40 participants (20 females, 20 males) took part in this study. The average age of all participants was 38.3 ($SD = 15.2$). Average age was 39.8 for females and 36.8 for males. All participants were fluent in English and had no disabilities.

Participants took an average of 11.4 seconds to respond to each trial, which consisted of five videos (one for the cue and one for each of the four choices). For the complete trials, all of our participants correctly answered a number of trials greater than chance (25%), with an overall accuracy of 60.8%. Hence, we deemed each participant's data as acceptable in terms of quality research data.

5.5 Results

For each category of incomplete cues, we used a Chi-square goodness-of-fit test to determine whether the distribution of options chosen by participants was equally proportionate. If the four-action Chi-square test revealed that there were statistically significant differences among the chosen options, we performed post hoc Chi-square tests for each pairwise comparison with Bonferroni adjustments to correct for false positives due to multiple comparisons (i.e., p must be less than 0.0083).

5.5.1 Subject-Incomplete Trials

We found significant differences among the options chosen for our subject-incomplete trials ($\chi^2(3) = 218.957, p < 0.001$). The post-hoc tests revealed that the *User* option was chosen significantly more than the *Male* ($\chi^2(1) = 50.156, p < 0.001$), *Female1* ($\chi^2(1) = 126.295, p < 0.001$), and *Female2* ($\chi^2(1) = 128.030, p < 0.001$) options. We also found that the *Male* option was chosen significantly more than the *Female1* ($\chi^2(1) = 21.951, p < 0.001$)

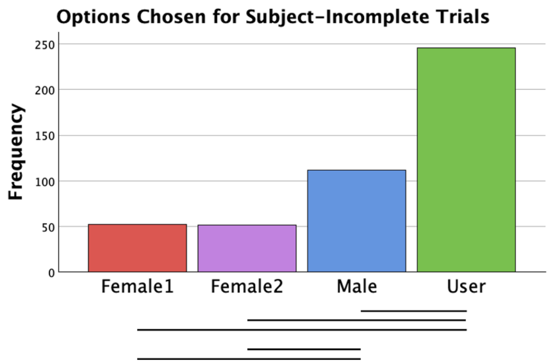

Figure 7: Frequencies that the reaction options for subject-incomplete trials were chosen. We found that the *User* option was chosen significantly more than the other three options. Lines at the bottom link significantly different conditions ($p < 0.05$).

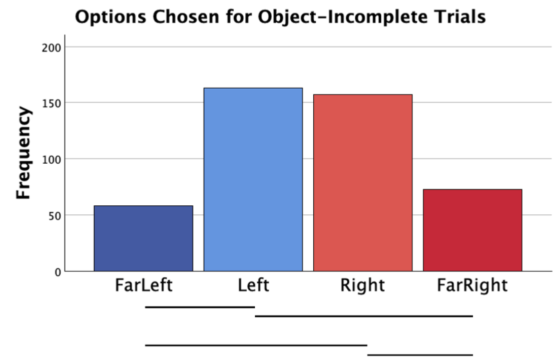

Figure 9: Frequencies that the reaction options for object-incomplete trials were chosen. We found that the *Left* and *Right* options were chosen significantly more than the *FarLeft* and *FarRight* options. Lines at the bottom link significantly different conditions ($p < 0.05$).

and *Female2* ($\chi^2(1) = 22.828, p < 0.001$) options. We did not find a significant difference between the *Female1* and *Female2* options ($\chi^2(1) = 0.010, p = 0.922$). See Figure 7 for these results.

5.5.2 Predicator-Incomplete Trials

We found significant differences among the options chosen for our predicator-incomplete trials ($\chi^2(3) = 87.201, p < 0.001$). The post-hoc tests revealed that the *Use* option was chosen significantly more than the *Grab* ($\chi^2(1) = 25.307, p < 0.001$), *Move* ($\chi^2(1) = 61.594, p < 0.001$), and *View* ($\chi^2(1) = 47.401, p < 0.001$) options. We also found that the *Grab* option was chosen significantly more than the *Move* option ($\chi^2(1) = 8.989, p = 0.003$) but not the *View* option ($\chi^2(1) = 3.817, p = 0.051$). We did not find a significant difference between the *Move* and *View* options ($\chi^2(1) = 1.119, p = 0.290$). See Figure 8 for these results.

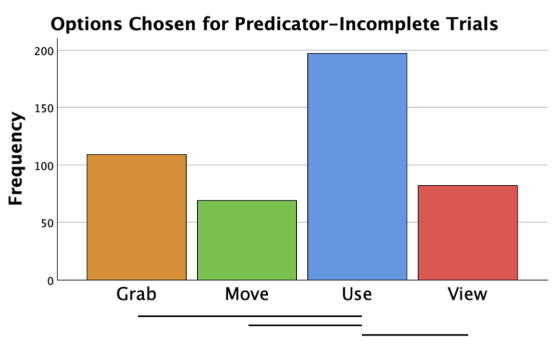

Figure 8: Frequencies that the reaction options for predicator-incomplete trials were chosen. We found that the *Use* option was chosen significantly more than the other three options. Lines at the bottom link significantly different conditions ($p < 0.05$).

5.5.3 Object-Incomplete Trials

We found significant differences among the options chosen for our object-incomplete trials ($\chi^2(3) = 80.361, p < 0.001$). The post-hoc tests revealed that the *Left* option was chosen significantly more than the *FarLeft* ($\chi^2(1) = 49.887, p < 0.001$) and *FarRight* ($\chi^2(1) = 34.322, p < 0.001$) options but not the *Right* option ($\chi^2(1) = 0.113, p = 0.737$). Similarly, we found that the *Right* option was chosen significantly more than the *FarLeft* ($\chi^2(1) = 45.586, p < 0.001$) and *FarRight* ($\chi^2(1) = 30.678, p < 0.001$) options. We did not find a significant difference between the *FarLeft* and *FarRight* options ($\chi^2(1) = 1.718, p = 0.190$). See Figure 9.

5.5.4 Modifier-Incomplete Trials

For our modifier-incomplete cues, we considered our move-oriented trials (with options for backward, forward, leftward, and rightward) separately from our use-oriented trials (with options for increasing levels of use, such as the brightness to turn a light on to).

For our move-oriented, modifier-incomplete trials, we found significant differences among the directional options chosen ($\chi^2(3) = 8.408, p = 0.038$). The post-hoc tests revealed that the *Forward* option was chosen significantly more than the *Backward* option ($\chi^2(1) = 7.895, p = 0.005$) but not the *Leftward* ($\chi^2(1) = 3.226, p = 0.072$) or *Rightward* ($\chi^2(1) = 1.744, p = 0.187$) options. However, we found no significant differences between the *Backward* and *Leftward* ($\chi^2(1) = 1.064, p = 0.302$), *Backward* and *Rightward* ($\chi^2(1) = 2.273, p = 0.132$), and *Leftward* and *Rightward* ($\chi^2(1) = 0.229, p = 0.632$) options. See Figure 10 above.

For our use-oriented, modifier-incomplete trials, we found significant differences among the increasing-level options chosen ($\chi^2(3) = 20.545, p < 0.001$). The post-hoc tests revealed that the *Level3* option was chosen significantly more than the *Level1* ($\chi^2(1) = 9.800, p = 0.002$) and *Level2* ($\chi^2(1) = 14.949, p < 0.001$)

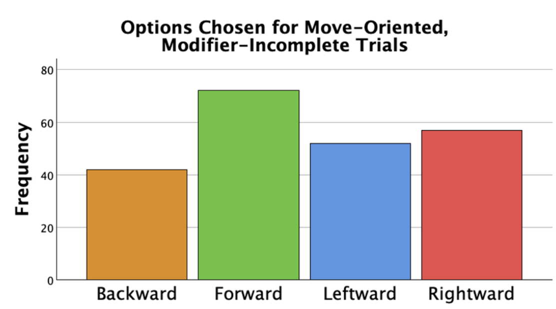

Figure 10: Frequencies that the reaction options for move-oriented, modifier-incomplete trials were chosen. We found that the *Forward* option was chosen significantly more than the *Backward* option. Lines at the bottom link significantly different conditions ($p < 0.05$).

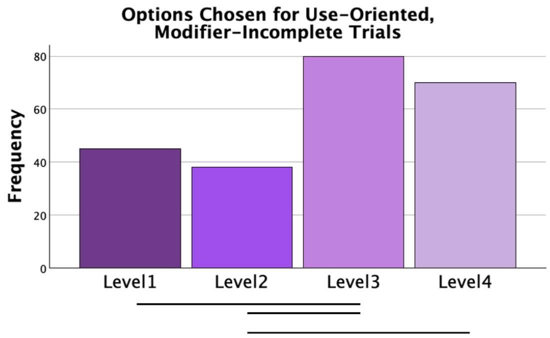

Figure 11: Frequencies that the reaction options for use-oriented, modifier-incomplete trials were chosen. We found that the *Level3* and *Level4* options were chosen significantly more than the *Level2* option and that the *Level3* option was chosen significantly more than the *Level1* option. Lines at the bottom link significantly different conditions ($p < 0.05$).

options but not the *Level4* option ($\chi^2(1) = 0.667, p = 0.414$). We also found that the *Level4* option was chosen significantly more than the *Level2* option ($\chi^2(1) = 9.481, p = 0.002$) but not the *Level1* option ($\chi^2(1) = 5.435, p = 0.020$, i.e., not less than the Bonferroni corrected 0.0083). We also did not find a significant difference between the *Level1* and *Level2* options ($\chi^2(1) = 0.590, p = 0.442$). See Figure 11 for these results.

5.5.5 Complete Trials

For our complete trials, we used a Chi-square goodness-of-fit test to determine whether the frequency of correct options chosen by participants was equal to the frequency of incorrect options. We found that the correct option for each complete trial ($N = 292$) was chosen significantly more ($\chi^2(1) = 34.041, p < 0.001$) than any of the three incorrect options ($N = 167$). We also tested the frequencies of correct and incorrect options for each type of complete trial based on its incomplete counterpart. Interestingly, we did not find any significant differences between the correct and incorrect options for the complete trials created from subject-incomplete trials ($\chi^2(1) = 0.009, p = 0.926$). However, we did find that the correct option was chosen significantly more than any of the incorrect options for the complete trials created from predicator-incomplete trials ($\chi^2(1) = 12.45, p < 0.001$), object-incomplete trials ($\chi^2(1) = 25.31, p < 0.001$), and modifier-incomplete trials ($\chi^2(1) = 9.47, p = 0.002$).

5.5.6 Complete vs. Incomplete Trials

As previously discussed in section 5.2.5, each of our complete trials was created by modifying the interaction cue of one of the previous incomplete trials to be complete. Hence, to further investigate how complete cues affected our results, we used a Chi-square goodness-of-fit test to determine whether the frequency that correct options for complete trials were chosen by participants was equal to the frequency that the same options for incomplete trials were chosen. We found that for each option the complete trial version ($N = 292$) was chosen significantly more ($\chi^2(1) = 78.039, p < 0.001$) than its incomplete trial counterpart ($N = 114$).

6 Discussion

In this section, we discuss why complete interaction cues are important, how MAGIC can be used to identify incomplete cues, that MAGIC requires design experience to correctly use, the limitations of our experimental design, and other limitations of this work.

6.1 Completeness Matters

Incomplete interaction cues do not provide sufficient information, which results in users making assumptions about the intended interaction. For subject-incomplete cues, we found that participants assume that they are the intended user of each cue. For predicator-incomplete cues, we found that participants were more likely to use objects based on their purposes (e.g., opening a book), as opposed to viewing, grabbing, or moving the object. For object-incomplete cues, we found that participants chose to interact with the objects spatially closest to the cue. For our modifier-incomplete cues, we found that participants were more likely to move objects forward than backward and more likely to use objects at increased degrees of interaction (e.g., turning a light brighter or opening a book to a later page). Finally, our results indicate that participants were more likely to choose the intended or correct interaction than another interaction in complete trials (except for those created from subject-incomplete trials) and that participants were more likely to choose that interaction in complete trials than their incomplete-trial counterparts. Altogether, these results indicate that complete interaction cues are important for prompting users to choose the intended interaction of a cue and to not make assumptions about the interaction, such as who is to perform it and on which object.

Outside of the presented study, there is evidence within the literature that incomplete cues are inferior to more-complete cues. Renner and Pfeiffer [41] compared several visual cues for a series of pick tasks. They found that the Image cue (see Figure 3A) was significantly worse than their SWAVE cue (see Figure 3B) in terms of completion times and excessive head movements. As analyzed in section 4, MAGIC can be used to show that the Image cue is modifier incomplete while the SWAVE cue is modifier complete. Similarly, Liu et al. [28] found that cues conveying objects and their spatial directions afforded significantly better performances than cues conveying only the objects for a path-following task. These results from the literature support the theory that completeness matters and that incomplete cues may be inferior to more-complete cues.

6.2 MAGIC is a Multipurpose Tool

In this paper, we have demonstrated that MAGIC can be used to analyze the functional grammar of interaction cues based on their intended interactions and perceptual stimuli. In section 3, we have provided a detailed explanation of how to use MAGIC as an analytic tool, including two text examples describing the same intended interaction. In section 4, we have demonstrated how MAGIC can be used to analyze preexisting interaction cues to denote important grammar differences, which further explain the empirical differences previously found [41].

However, it is important to point out that MAGIC can also be used as a tool for designing complete interaction cues. More specifically, an interaction cue designer can use MAGIC to determine if the perceptual stimuli of a cue design needs to be augmented to be more complete. For example, consider a simple arrow positioned near and pointing toward the object of an intended interaction, which is common within the literature [14]. By itself, this simple arrow is predicator incomplete. However, if the designer were to place a label with the text “Pick Up” at the base of the arrow, the cue would now be predicator complete. By using MAGIC in this manner, interaction designers can improve the designs of their interaction cues rapidly and iteratively prior to costly development or conducting usability testing, thereby saving time, effort, and resources. Hence, MAGIC may serve as a valuable analytic tool for designing more-effective interaction cues earlier in the life cycle.

6.3 MAGIC Requires Experience

One limitation of MAGIC is that interaction cue designers are likely not accustomed to writing texts that explicitly describe their intended interactions. While this may seem challenging at first, we want to stress the importance of focusing on the semantics and functional grammar of these texts, as opposed to their syntax and lexicon. As demonstrated in section 3, different texts that are functionally equivalent can be analyzed using MAGIC to produce the same inferences about an interaction cue and its completeness. Hence, interaction designers do not need to be grammar experts to use MAGIC. However, they do need to focus on the functional aspects of their intended interactions, perceptual cues, and shared interaction states, which will often require design experience.

An inexperienced or hurried designer may make three different types of errors when applying MAGIC. First, a designer may fail to fully describe an intended interaction when defining a text for it. For example, many designers may not specify the *on the table* modifier in (3) or a similar modifier, even if the intended interaction requires a specific position for the bowl. Second, a designer may assume that a perceptual stimulus clearly maps to an identified grammar role when it is in fact ambiguous. For example, reconsider the ambiguity of the green arrow in (6). Finally, a designer may make a mistake when considering (or not considering) the shared interaction state between the user and the system. For example, the only difference between the green bowl interaction cues mapped in (5) and (7) is the lack of a shared interaction state for the latter due to the system preemptively displaying the cue regardless of the user's interactions or input. However, as a result, we find the interaction cue in (5) to be complete while the cue in (7) is found to be predicator incomplete, object incomplete, and modifier incomplete.

6.4 Limitations of Our Experimental Design

One limitation of this work is that our empirical evaluation of incomplete cues was conducted online via a web browser as opposed to in an AR or VR headset. Furthermore, all of the investigated trials were VR-based, which did not evaluate how users would react in AR scenarios with both real and virtual stimuli. While this approach afforded less ecological validity than an in-person AR or VR study, it also eliminated the confounds introduced by differing motor skills of participants while focusing on the perceptual and cognitive aspects of the interaction cues. By using the Prolific platform, we were able to recruit a gender-balanced sample of participants outside of our own university population. In particular, the average age of all the participants was 38 years, which is more representative of the world population than common college samples.

Another concern related to our experimental design is whether the results are valid given that participants selected interactions from four predefined options. On one hand, this approach precluded participants from choosing interactions beyond those presented to them, such as viewing an out-of-view object or grabbing an object further away from the interaction cue. One way to address this would be through an in-person elicitation study like some recent gesture-elicitation studies [1, 11, 13]. However, we do not believe that such an elicitation study would result in more-consistent responses for our incomplete trials due to the broader range of possible interactions. It is possible that such a broader range of possibilities would decrease how often participants choose the correct interaction for complete cues. However, we believe we would still find results indicating that complete cues result in the same interaction being chosen significantly more than when the cue is incomplete.

Another possible concern about our experiment is the quality of our participant responses (e.g., whether participants randomly clicked options to progress through our study). Our statistically significant results indicate that poor participant data and random behaviors likely did not affect our results, particularly because we randomized the order of both the questions and the responses for each question. This is further supported by our results indicating that complete cues resulted in the same interaction option being chosen significantly more than with incomplete cues.

A final concern about our study is the variation and quality of our cue implementations. We originally designed different types of visual cues (e.g., *Text*, *Icon*, *Video*) to investigate each type of incomplete cue (i.e., subject, predicator, object, and modifier) to ensure that our results were not skewed by a single perceptual stimulus choice. However, it is possible that these variations created confounds within our results. Furthermore, our choice to create complete trials by modifying incomplete trials to be complete likely negatively impacted the quality of their completeness and our results. This is particularly true for the complete trials created from subject-incomplete trials, as we found no significant difference between the correct and incorrect options. Hence, future work should focus on designing more-effective complete trials that are not noticeably different from the incomplete trials.

6.5 Other Limitations and Future Work

Another limitation of the current work is that we have focused on visual cues, particularly given our experimental design. However, we believe MAGIC is applicable to any perceptual stimuli, including auditory and haptic cues (e.g., [30]). For example, consider the Image cue depicted in Figure 3A. If it were accompanied by a 3D sound spatially co-located with the coffee maker (e.g., [49]), it would become modifier complete and may have similar performance to the SWAVE cue depicted in Figure 3B. However, future work is necessary to verify that MAGIC is also applicable to other perceptual stimuli, and we encourage other researchers to explore such applications of MAGIC.

Another consideration about MAGIC is whether it can be applied to languages other than English. The concept of functional grammar has been applied to a number of non-English languages, such as Spanish and Chinese [35]. Hence, we believe that MAGIC can likely be applied with other languages, as it focuses on the functional grammar roles of the underlying texts, as opposed to their syntax and lexicon. Still, further research is necessary to demonstrate how MAGIC can be applied to languages other than English.

7 Conclusion

In this paper, we have presented a Method for Analyzing the Grammar of Incomplete Cues (MAGIC), which provides an analytical approach for assessing interaction cue designs and hypothesizing their efficacy based on the completeness or incompleteness of the information that they convey. We have detailed and demonstrated how MAGIC can be used to analyze preexisting interaction cues. We have also presented an empirical evaluation of incomplete and complete interaction cues that demonstrates completeness matters and incomplete cues force users to make assumptions about the intended interaction. Furthermore, we have discussed how MAGIC can also be used as a design tool, in addition to being an analysis tool for identifying potential usability issues.

Supplemental Materials

All supplemental materials are available at https://osf.io/62mbg/?view_only=8f53f89973ac4190b0b9af9f4fec7d2b, released under a CC BY 4.0 license. They include (1) the materials used for the user study, (2) the results obtained from the user study, and (3) complete analyses of the results obtained from IBM SPSS Statistics version 30.0.

Acknowledgments

This material is based on work partially supported by the National Science Foundation award #2021607 - "CAREER: Leveraging the Virtualness of Virtual Reality for More-Effective Training".

References

[1] A. Ali, M. Ringel Morris, and J. O. Wobbrock. "I Am Iron Man": Priming improves the learnability and memorability of user-elicited

gestures. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21. Association for Computing Machinery, New York, NY, USA, 2021. doi: 10.1145/3411764.3445758 9

[2] F. Argelaguet and C. Andujar. A survey of 3D object selection techniques for virtual environments. *Computers & Graphics*, 37(3):121–136, 2013. doi: 10.1016/j.cag.2012.12.003 4

[3] R. P. W. Arne Seeliger and S. Feuerriegel. Context-adaptive visual cues for safe navigation in augmented reality using machine learning. *International Journal of Human–Computer Interaction*, 0(0):1–21, 2022. doi: 10.1080/10447318.2022.2122114 1

[4] H. Bai, P. Sasikumar, J. Yang, and M. Billinghurst. A user study on mixed reality remote collaboration with eye gaze and hand gesture sharing. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, p. 1–13. Association for Computing Machinery, New York, NY, USA, 2020. doi: 10.1145/3313831.3376550 2

[5] S. C. Barathi, D. J. Finnegan, M. Farrow, A. Whaley, P. Heath, J. Buckley, P. W. Dowrick, B. C. Wuensche, J. L. J. Bilzon, E. O'Neill, and C. Lutteroth. Interactive feedforward for improving performance and maintaining intrinsic motivation in VR exergaming. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, p. 1–14. Association for Computing Machinery, New York, NY, USA, 2018. doi: 10.1145/3173574.3173982 2

[6] J. Bertrand, D. Brickler, S. Babu, K. Madathil, M. Zelaya, T. Wang, J. Wagner, A. Gramopadhye, and J. Luo. The role of dimensional symmetry on bimanual psychomotor skills education in immersive virtual environments. In *2015 IEEE Virtual Reality (VR)*, pp. 3–10, 2015. doi: 10.1109/VR.2015.7223317 1

[7] F. Bork, C. Schnelzer, U. Eck, and N. Navab. Towards efficient visual guidance in limited field-of-view head-mounted displays. *IEEE Transactions on Visualization and Computer Graphics*, 24(11):2983–2992, 2018. doi: 10.1109/TVCG.2018.2868584 1, 2

[8] F. Born, P. Sykownik, and M. Masuch. Co-located vs. remote gameplay: The role of physical co-presence in multiplayer room-scale VR. In *2019 IEEE Conference on Games (CoG)*, pp. 1–8, 2019. doi: 10.1109/CIG.2019.8848001 4

[9] S. Card, T. MORAN, and A. Newell. The model human processor- an engineering model of human performance. *Handbook of perception and human performance.*, 2(45–1), 1986. 5

[10] M. Dalrymple and L. Mycock. *The Oxford reference guide to lexical functional grammar*. Oxford University Press, USA, 2019. 2

[11] A. Danielescu and D. Piorkowski. Iterative design of gestures during elicitation: Understanding the role of increased production. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, CHI '22. Association for Computing Machinery, New York, NY, USA, 2022. doi: 10.1145/3491102.3501962 9

[12] K. R. Dillman, T. T. H. Mok, A. Tang, L. Oehlberg, and A. Mitchell. A visual interaction cue framework from video game environments for augmented reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, p. 1–12. Association for Computing Machinery, New York, NY, USA, 2018. doi: 10.1145/3173574.3173714 1, 2

[13] Y. Felberbaum and J. Lanir. Better understanding of foot gestures: An elicitation study. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, p. 1–12. Association for Computing Machinery, New York, NY, USA, 2018. doi: 10.1145/3173574.3173908 9

[14] M. Gattullo, A. Evangelista, A. E. Uva, M. Fiorentino, and J. L. Gabbard. What, How, and Why are Visual Assets Used in Industrial Augmented Reality? A systematic review and classification in maintenance, assembly, and training (from 1997 to 2019). *IEEE Transactions on Visualization and Computer Graphics*, 28(2):1443–1456, 2022. doi: 10.1109/TVCG.2020.3014614 1, 2, 4, 8

[15] R. K. Ghamandi, R. K. Kattoju, Y. Hmaiti, M. Maslych, E. M. Taranta, R. P. McMahan, and J. LaViola. Unlocking understanding: An investigation of multimodal communication in virtual reality collaboration. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, CHI '24. Association for Computing Machinery, New York, NY, USA, 2024. doi: 10.1145/3613904.3642491 2

[16] U. Gruenefeld, D. Ennenga, A. E. Ali, W. Heuten, and S. Boll. EyeSee360: designing a visualization technique for out-of-view objects in head-mounted augmented reality. In *Proceedings of the 5th Symposium on Spatial User Interaction*, SUI '17, p. 109–118. Association for Computing Machinery, New York, NY, USA, 2017. doi: 10.1145/3131277.3132175 1, 2

[17] M. A. K. Halliday and C. M. Matthiessen. *Halliday's introduction to functional grammar*. Routledge, 2013. 1

[18] R. Hartson and P. S. Pyla. *The UX Book: Process and guidelines for ensuring a quality user experience*. Elsevier, 2012. 4

[19] D. Hindle and M. Rooth. Structural ambiguity and lexical relations. In *Proceedings of the 29th Annual Meeting on Association for Computational Linguistics*, ACL '91, p. 229–236. Association for Computational Linguistics, USA, 1991. doi: 10.3115/981344.981374 2

[20] X. Hu, A. Moore, J. Coleman Eubanks, A. Aiyaz, and R. P. McMahan. Evaluating interaction cue purpose and timing for learning and retaining virtual reality training. In *Proceedings of the 2020 ACM Symposium on Spatial User Interaction*, SUI '20. Association for Computing Machinery, New York, NY, USA, 2020. doi: 10.1145/3385959.3418448 1

[21] R. Huddleston and G. K. Pullum. *The Cambridge Grammar of the English Language*. Cambridge University Press, 2002. doi: 10.1017/9781316423530 2, 3

[22] D. Kao, A. J. Magana, and C. Mousas. Evaluating tutorial-based instructions for controllers in virtual reality games. *Proc. ACM Hum.-Comput. Interact.*, 5(CHI PLAY), oct 2021. doi: 10.1145/3474661 1

[23] V. Kaptelinin and B. Nardi. Affordances in HCI: toward a mediated action perspective. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, p. 967–976. Association for Computing Machinery, New York, NY, USA, 2012. doi: 10.1145/2207676.2208541 2

[24] B. M. Khuong, K. Kiyokawa, A. Miller, J. J. La Viola, T. Mashita, and H. Takemura. The effectiveness of an AR-based context-aware assembly support system in object assembly. In *2014 IEEE Virtual Reality (VR)*, pp. 57–62, 2014. doi: 10.1109/VR.2014.6802051 1, 2

[25] M. Kolln. Rhetorical grammar: A modification lesson. *The English Journal*, 85(7):25–31, 1996. 3

[26] C. Lai, X. Hu, A. A. Aiyaz, A. Segismundo, A. Phadke, and R. P. McMahan. The cognitive loads and usability of target-based and steering-based travel techniques. *IEEE Transactions on Visualization and Computer Graphics*, 27(11):4289–4299, 2021. doi: 10.1109/TVCG.2021.3106507 4

[27] D. Lange, T. C. Stratmann, U. Gruenefeld, and S. Boll. HiveFive: immersion preserving attention guidance in virtual reality. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, p. 1–13. Association for Computing Machinery, New York, NY, USA, 2020. doi: 10.1145/3313831.3376803 1, 2

[28] J.-S. Liu, C. Elvezio, B. Tversky, and S. Feiner. Using multi-level precueing to improve performance in path-following tasks in virtual reality. *IEEE Transactions on Visualization and Computer Graphics*, 27(11):4311–4320, 2021. doi: 10.1109/TVCG.2021.3106476 1, 2, 8

[29] J.-S. Liu, B. Tversky, and S. Feiner. Precueing object placement and orientation for manual tasks in augmented reality. *IEEE Transactions on Visualization and Computer Graphics*, 28(11):3799–3809, 2022. doi: 10.1109/TVCG.2022.3203111 1, 2

[30] A. Marquardt, C. Trepkowski, T. D. Eibich, J. Maiero, E. Kruijff, and J. Schöning. Comparing non-visual and visual guidance methods for narrow field of view augmented reality displays. *IEEE Transactions on Visualization and Computer Graphics*, 26(12):3389–3401, 2020. doi: 10.1109/TVCG.2020.3023605 9

[31] M. M. Montoya, A. P. Massey, and N. S. Lockwood. 3D collaborative virtual environments: Exploring the link between collaborative behaviors and team performance. *Decision Sciences*, 42(2):451–476, 2011. doi: 10.1111/j.1540-5915.2011.00318.x 4

[32] A. G. Moore, X. Hu, J. C. Eubanks, A. A. Aiyaz, and R. P. McMahan. A formative evaluation methodology for VR training simulations. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pp. 125–132, 2020. doi: 10.1109/VRW50115.2020.00027 1

[33] M. Murcia-López and A. Steed. A comparison of virtual and physical

training transfer of bimanual assembly tasks. *IEEE Transactions on Visualization and Computer Graphics*, 24(4):1574–1583, 2018. doi: 10.1109/TVCG.2018.2793638 1, 2

[34] A. Muresan, J. Mcintosh, and K. Hornbæk. Using feedforward to reveal interaction possibilities in virtual reality. *ACM Trans. Comput.-Hum. Interact.*, 30(6), sep 2023. doi: 10.1145/3603623 1, 2

[35] I. N. Mwinlaaru and W. W. Xuan. A survey of studies in systemic functional language description and typology. *Functional Linguistics*, 3:1–41, 2016. 9

[36] D. A. Norman. Affordance, conventions, and design. *Interactions*, 6(3):38–43, May 1999. doi: 10.1145/301153.301168 2

[37] T. Parsons. Modifiers and quantifiers in natural language. *Canadian Journal of Philosophy Supplementary Volume*, 6:29–60, 1980. doi: 10.1080/00455091.1980.10715756 3

[38] T. C. Peck, L. E. Sockol, and S. M. Hancock. Mind the gap: The underrepresentation of female participants and authors in virtual reality research. *IEEE Transactions on Visualization and Computer Graphics*, 26(5):1945–1954, 2020. doi: 10.1109/TVCG.2020.2973498 5

[39] E. Peer, L. Brandimarte, S. Samat, and A. Acquisti. Beyond the turk: Alternative platforms for crowdsourcing behavioral research. *Journal of Experimental Social Psychology*, 70:153–163, 2017. doi: 10.1016/j.jesp.2017.01.006 5

[40] K. Quinn and J. L. Gabbard. Augmented reality visualization techniques for attention guidance to out-of-view objects: A systematic review. In *2024 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 826–835, 2024. doi: 10.1109/ISMAR62088.2024.00098 2

[41] P. Renner and T. Pfeiffer. Attention guiding techniques using peripheral vision and eye tracking for feedback in augmented-reality-based assistance systems. In *2017 IEEE Symposium on 3D User Interfaces (3DUI)*, pp. 186–194, 2017. doi: 10.1109/3DUI.2017.7893338 1, 2, 4, 5, 8

[42] S. Rothe, D. Buschek, and H. Hußmann. Guidance in cinematic virtual reality-taxonomy, research status and challenges. *Multimodal Technologies and Interaction*, 3(1), 2019. doi: 10.3390/mti3010019 1, 2

[43] T. Schmelter, L. Kruse, S. Karaosmanoglu, S. Rings, F. Steinicke, and K. Hildebrand. Towards more inclusive and accessible virtual reality: Conducting large-scale studies in the wild. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI EA '23. Association for Computing Machinery, New York, NY, USA, 2023. doi: 10.1145/3544549.3583888 5

[44] B. Schwerdtfeger and G. Klinker. Supporting order picking with augmented reality. In *2008 7th IEEE/ACM International Symposium on Mixed and Augmented Reality*, pp. 91–94, 2008. doi: 10.1109/ISMAR.2008.4637331 1

[45] A. Seeliger, G. Merz, C. Holz, and S. Feuerriegel. Exploring the effect of visual cues on eye gaze during AR-guided picking and assembly tasks. In *2021 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, pp. 159–164, 2021. doi: 10.1109/ISMAR-Adjunct54149.2021.00041 1, 2

[46] G. Thompson. *Introducing functional grammar*. Routledge, 2013. 1, 2, 3

[47] B. Volmer, J. Baumeister, R. Matthews, L. Grosser, S. Von Itzstein, S. Banks, and B. H. Thomas. A comparison of spatial augmented reality predictive cues and their effects on sleep deprived users. In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 589–598, 2022. doi: 10.1109/VR51125.2022.00079 1, 2

[48] J. Wuertz, S. A. Alharthi, W. A. Hamilton, S. Bateman, C. Gutwin, A. Tang, Z. Toups, and J. Hammer. A design framework for awareness cues in distributed multiplayer games. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, p. 1–14. Association for Computing Machinery, New York, NY, USA, 2018. doi: 10.1145/3173574.3173817 1, 2

[49] S. Yong and H.-C. Wang. Using spatialized audio to improve human spatial knowledge acquisition in virtual reality. In *Companion Proceedings of the 23rd International Conference on Intelligent User Interfaces*, IUI '18 Companion. Association for Computing Machinery, New York, NY, USA, 2018. doi: 10.1145/3180308.3180360 9