Advance Access publication on 22 December 2014
doi:10.1093/iwc/iwu046

Designing Intuitive Interactions: Exploring Performance and Reflection Measures

Jeremiah D. Still[1,*], Mary L. Still[1] and Joseph Grgic[2]

[1]*Department of Psychology, San Jose State University, One Washington Square, San Jose, CA 95192-0120, USA*
[2]*User Experience, Havas Life, New York, NY, USA*
Corresponding author: Jeremiah.Still@sjsu.edu

Intuitive interactions are supported by users' implicit and explicit learning experiences. But, determining user knowledge can be difficult. With many options available for eliciting that knowledge, we tested the effectiveness of two methods—performance and reflection. Users were presented with simple interactions that had varying levels of intuitiveness (affordance, convention, bias). They were asked to perform the interaction or to describe how the interaction should be designed. These methods of knowledge elicitation produced inconsistent results; sometimes they produced the same result (affordance-based interactions), sometimes the opposite (convention-based interactions). Furthermore, when both methods were used, results obtained from the second measure were often contaminated by completion of the first measure. Carryover effects were present regardless of which measure was completed first. These results indicate that the method used to elicit knowledge should be selected based on the type of interaction that is being investigated and multiple measures should be used with caution.

RESEARCH HIGHLIGHTS

- Results can be contaminated by employing multiple methods of knowledge elicitation.
- Interactions vary on a continuum of intuitiveness.
- User self-reflection can provide misleading information about conventional interactions.
- Self-reflection and performance measures can be used to investigate affordances.
- Intuitive interactions stem from perceptual processing and previous experience.

Keywords: human computer interaction; HCI theory, concepts and models; empirical studies in HCI; user studies; user interface design; interaction design theory, concepts and models; methodology

Received 17 March 2014; Revised 1 November 2014; Accepted 13 November 2014

1. INTRODUCTION

Designers are often faced with arbitrary choices. Even a simple task, such as determining which of two vertically oriented buttons should be used to instigate a rightward movement on a screen, has no correct answer. In these instances of uncertainty, how should a designer decide which mapping to adopt for an intuitive interaction? While there are ultimately many ways to make design decisions, this article is focused on the use of performance and reflection as means of eliciting user knowledge. Therefore, in the aforementioned scenario, one option is to provide users with the vertically oriented buttons and ask them to 'move' right. A second option is to simply survey users, asking them how they would expect the action to be mapped (cf. Bergum and Bergum, 1981). Would both methods lead to the same design decision?

Regardless of the method a designer uses to solve the hypothetical vertical-button scenario, the goal is to create an interface that will facilitate the user's ability to complete a task (Simon, 1969/1981). In addition, designers often aim to minimize the cognitive effort required to use an interface by making the interactions with it seem obvious, direct (Hutchins *et al*., 1986), or intuitive to the user. Even though use of the term *intuitive* is pervasive in both academic and professional literatures, it can be difficult to predict *a priori* what interactions will be intuitive. One of the reasons for this may stem from the very nature of intuition. In the cognitive sciences, intuition is described as a signal that is inherently vague; it comes about automatically, without insight, such that an individual may have no direct understanding of what events or what evidence led to the intuition (Betsch, 2008a,b; Bolte *et al*., 2003; Epstein, 2010; Wippich, 1994). Even though the information is vague, it is used as a source for decision-making (e.g. Betsch, 2008a,b; De Vries *et al*., 2008; Epstein, 2010). Because intuition arises automatically, it allows decisions to be made quickly with little effort (Topolinski and Strack, 2009). The implicit aspect of intuition is one of the challenges to predicting what designs will be intuitive; the user may not be able to verbalize why an interface is intuitive and, similarly, the designer may not be able to use his or her own experience to gain insight into what is intuitive.

In order to make predictions, it is necessary to delineate the conditions that facilitate intuitive experiences. There is a broad agreement that intuitive interactions emerge from the level of congruence between mental representations of past experiences and the current interaction. In particular, an intuitive experience emerges when pre-existing knowledge can be applied to the current situation with minimal effort (e.g. Allen and Buie, 2002; Hurtienne and Isreal, 2007; Naumann *et al*., 2007; Naumann and Hurtienne, 2010; Spool, 2005; Wippish, 1994). Therefore, familiarity with an interaction facilitates intuitive interactions (Raskin, 1994). Importantly, familiarity with an interaction is not tied to a specific system or product (Blackler *et al*., 2003, 2004, 2010; Naumann *et al*., 2007). For example, Blackler *et al*. (2003, 2004, 2010) identified users with varying experience with an interface (e.g. digital cameras and remote controls) and then presented them with a novel product containing some interaction (e.g. using a button to take a picture, using the zoom function) that they would have experienced on their own when using different products. In each of these studies, it was demonstrated that some interactions with a novel device can be intuitive as long as those interactions were familiar.

1.1. Origins of intuitive interactions

The constructs of *intuition* and *intuitive design* can be used to help define what constitutes an intuitive interaction. Intuitive interactions result from direct communication between a design and user. This directness is facilitated through either innate perceptual processing or automatic processing of the interactions, which results from many consistent learning instances. Both sources of information influence how an interface is perceived. Perceptual affordances provide one example of how innate perceptual processes can produce an intuitive interaction. Perceptual affordances are a reflection of our ability to unconsciously process an overwhelming amount of sensory data into actionable information (e.g. distance calculations, object recognition, spatial grouping). Gibson (1979) coined the term *affordances* to describe the emergent potential interactions between actors and their physical artifacts. Though Norman (1988) popularized the concept of perceived affordance within the Human–Computer Interaction design community, he expanded it beyond physical structures to include mental influences (i.e. cultural, logical and semantic constraints). We believe that from a designer's perspective, the only useful affordances are those that are perceived by the user. Therefore, like other researchers we place Gibson's affordances subordinate to perceived affordances (cf. Kannengiesser and Gero, 2012; Still and Dark, 2013). Interestingly, affordances are unique in that they describe our ability to perceive available actions within complex scenes with minimal cognitive effort (Still and Dark, 2013). According to You and Chen (2007), 'affordances challenge designers to avoid the reliance on symbols and cultural conventions in design. Instead, it encourages them to utilize possible intuitive actions that can serve a function in the process-product interaction' (p. 29). Thus, designers benefit from having a deeper understanding of how our sensation and perception system operates (e.g. visual grouping principles or depth cues). In addition, we have convention-based interactions (Norman, 1999) that stem from experience rather than an innate perceptual processing ability. This experience-based understanding of our environment is also critical for understanding the roots of intuitive interactions.

The user experiences a variety of interactions, storing some representation of those interactions in memory. When the user experiences a similar design in the future (e.g. same interaction in a different interface), it is possible that the interaction will be experienced as intuitive. To have an intuitive interaction, the current interaction must be similar to a previous interaction and the mental representation of the previous interaction must be accessed and applied quickly and with little effort. Although these definitions describe what constitutes an intuitive interaction, the formation of mental representations that support these interactions is underspecified. For example, familiarity alone does not ensure an intuitive interaction; the representation must also be accessed automatically. Automaticity is achieved over many consistent interactions (Shiffrin and Schneider, 1977). Because the representations serving automaticity are dynamic and they develop incrementally, it is more accurate to conceptualize them as being on a continuum between *effortful* and *automatic* rather than being either effortful or automatic (Moors and De Houwer, 2006).

We believe a similar conceptualization can be applied to intuitive interactions. At one end of the continuum are novel interactions. At the other end of the continuum are interactions that can be completed automatically with little effort. For instance, the interaction might be one that the user has experienced many times and those interactions have been consistent (e.g. there is no ambiguity about the actions associated with the interaction). Under these conditions, the designer can expect the user to have an intuitive interaction even if the interaction is experienced in a new context (e.g. embedded in a new interface). For example, experience with a physical slide lock can foster intuitive interactions with a phone's virtual slide lock. Although it might seem logical to always try to use an interaction that will be intuitive to the user on sight, there are circumstances in which that might be impossible or undesirable (e.g. copyrights, branding). For these reasons, many interactions will fall somewhere in the middle of the continuum.

An interaction might be in the middle of the continuum if the user has too little experience with the interaction or if a competing mental representation is in place (e.g. the user may have encountered a similar design element in the past, but the interaction differed). In either case, the mental representation supporting the interaction will be weaker than those supporting intuitive interactions. Of these familiar interactions, those that are experienced repeatedly and have consistent mapping have the potential to develop into intuitive interactions. That is, in the future, they could support intuitive interactions in novel situations. Other familiar interactions that are not encountered repeatedly or are associated with too much ambiguity, will not develop into intuitive interactions. The scrollbar is a good example of this distinction. The function of a scrollbar is to move the content in the central frame up or down. Designers may decide which way the interaction should work based on their personal knowledge and usage metaphors (e.g. content should move up as the bar moves down like one would scan down a page with his eyes or finger). However, the original decision is somewhat arbitrary. After the user gains experience with a particular scrollbar mapping, though, the interaction is no longer arbitrary. At this point, the interaction is familiar to the user. If he continues to have the same type of interactions with scrollbars, the interaction will be intuitive at a later point in time—this is an example of a potentially intuitive interaction. This consistent experience with the scrollbar would strengthen the mental representation associated with the interaction making it seem more intuitive (e.g. Betsch, 2008a,b). In addition, if the user encounters a scrollbar with the opposite mapping (e.g. scrolling down moves content down instead of up) the interaction will not be intuitive even though he has had experience with scrollbars in the past. Interestingly, the experience with the 'inconsistent' scrollbar creates ambiguity as to which action should be associated with the design element. With enough inconsistent experiences, it becomes less likely that an interaction will be experienced as intuitive.

Intuitive interactions are supported either by the perceptual processing system or by our previous experience being automatically applied. Our cognitive system is continuously attempting to resolve ambiguity within an overwhelming environment. It is usually successful in its perceptual calculations within natural environments (with the exception of visual illusions, for instance). Although users have these abilities, designed environments carry much more ambiguity, which requires users to rely on both explicit and implicit experience. Additionally, the amount of experience a user has with a given design varies. It is a significant challenge, then, to understand users' expectations of a design and to facilitate intuitive interactions.

1.2. Varying degrees of intuitive interactions: affordances, conventions and biases

One of the first steps in understanding user expectations is to see if different user experiences result in functionally different responses. In this vein, Still and Dark (2008) investigated whether or not empirical differences exist between three different types of intuitive interactions—perceptual affordance, cultural convention and bias. In their study, users interacted with a simple two-button design; the buttons were arranged horizontally, vertically or diagonally and users were asked to use the buttons to 'move' up, down, left or right. The combination of required 'movements' and button-configurations allowed the use of affordance, convention and bias interactions (see Table 1). Affordance-based interactions were those in which the prescribed movement corresponded directly with the relative button positions (e.g. using the button on the right to move right). Because affordances guide user interactions immediately and effortlessly, they may be the most intuitive type of interaction. Affordances clearly support user-product interactions (You and Chen, 2007) and they may be particularly useful for facilitating intuitive interactions with novel interfaces (Blackler, 2008).

Table 1. Mapping of button configurations, direction task, and type of intuitive interaction.

	Button configuration		
Direction	Horizontal	Vertical	Diagonal
Up	Right; Bias	Top; Affordance	Top; Affordance
Down	Left; Bias	Bottom; Affordance	Bottom; Affordance
Left	Left; Affordance	Bottom; Convention	Left; Affordance
Right	Right; Affordance	Top; Convention	Right; Affordance

Note: Expected button-to-action mappings and their categorizations based on button configuration.

Although affordances are naturally intuitive, it is not always possible to integrate them into an interface. Scrollbar functionality and the key placement on a *qwerty* keyboard are prime examples of this kind of 'arbitrary' interaction mapping. In light of the prevalence of non-affordance interactions, Still and Dark (2008) simulated these conditions by including non-affordance button-to-action mappings in their study (see Table 1). Mappings with a high level of agreement between users (i.e. 80% of participants selected the same mapping) were classified as conventions (e.g. using the top vertical button to move right). Mappings that occurred above chance level, but below the convention cutoff were classified as biases (e.g. using the left horizontal button to move down). Similar distinctions have been made in the human factors stimulus–response compatibility literature. In that context, conventions are referred to as population stereotypes (cf. Proctor and Vu, 2006).

According to Still and Dark (2008, 2010), the distinction between conventions and biases is important as it reflects the state of the memory structures underlying the interaction mapping and how those structures are accessed. A convention reflects a learned, standardized interaction that has likely become automatized. User interactions with a *qwerty* keyboard, for example, are best characterized as conventions. When an individual is learning to type, there seems to be no rhyme or reason guiding letter placement. Although some letters that are 'near' each other in the alphabet are loosely clustered together (e.g. *C*, *D*, *E*, and *L, M, N, O, P*), one cannot use alphabetization to predict where a letter is located. But, after years of consistent experience with the *qwerty* keyboard (e.g. keys remain in the same relative positions), learning occurs and the ambiguity is removed from the interaction (Logan, 2002). Neisser (1976) referred to this learning as a perceptual cycle in which past experience influences current perception. The combination of repeated interaction and reduced ambiguity strengthens the internal representation associated with the interaction, thereby making it more intuitive (e.g. Betsch, 2008a,b; Raskin, 1994). Still and Dark (2010) proposed that even interactions with other, seemingly unrelated, interfaces shape the user's representation of basic interactions. For instance, a user might move a lever 'up' in their vehicle to activate the right turn signal. This interaction knowledge could be used in any similar task, even one as simple as moving to the 'right' when only two vertical buttons are available.

Whereas a convention reflects a standardized interaction, a bias reflects some lack of familiarity with or consistency within the interaction. Because of this, a bias requires some level of conscious or controlled processing in order to be used. If the biased interaction has a consistent mapping, it could become a convention with sufficient experience. In turn, the interaction could be intuitive to the user in the future as experience with the interaction increases. We refer to these interactions as being consistent biased interactions that are *potentially intuitive*. If, in contrast, the biased interaction includes a mapping that is inconsistent, it is unlikely that it will ever develop into a convention. Inconsistent mappings are easy to find, for example, in Fig. 1, the right button on the remote control is associated with an increase in volume, but the right button on the monitor is associated with a decrease in brightness. This ambiguity prevents the interaction from being accessed automatically. Therefore, an inconsistent biased interaction may have some level of familiarity to users, but it will not be intuitive when compared with a convention or affordance.

Figure 1. These two interfaces represent opposite interaction mappings.

In terms of intuitive interactions, it is logical that affordances and conventions are more intuitive than biases. But it is less clear whether or not there is a functional difference between affordances and conventions. Empirical investigations by Still and Dark (2008, 2010) began to address this issue. The 2008 study revealed no significant difference in the number of users adopting affordance or conventional interactions and no significant difference in the speed with which the interactions were completed. In a more sensitive test of the potential differences between affordances and conventions, Still and Dark (2010) used the same button configurations but included some situations with inverted mappings (upon pressing a button, the cursor moved in the opposite direction as the affordance or convention) and included a working memory manipulation. The trials with inverted mapping were intended to determine if violating a convention would disrupt users as much as violating an affordance. The working memory manipulation was intended to determine if affordances and conventions are equally effective under cognitive load. The results revealed that it was equally disruptive to violate conventions and affordances, suggesting that the internal representations supporting conventions are well-defined and support automatic processes. Interestingly, the working memory manipulation revealed that while conventions were sensitive to working memory load—response times for conventional interactions were slower in a high load condition compared with a low load condition— affordances were not. This finding suggests that although conventions may be well practiced, they may not be 'hard wired' in the same way as affordances. Based on the fact that conventional interactions can be behaviorally indistinguishable from affordance-based interactions, it follows that conventional interactions are intuitive. But, because there may be a slight working memory cost associated with conventions, we propose that conventional interactions might be slightly less intuitive than affordances.

1.3. Performance and reflective measures of intuitive interactions

If conventions provide an intuitive interaction that is comparable with affordances, but biases do not, it becomes important from a practical standpoint to have an accurate method for identifying conventions and biases. While Still and Dark (2008) used performance measures to determine *post hoc* which interactions were conventions and which were biases, there are a variety of techniques available for assessing intuitive interactions. They can be broadly categorized by whether the user is asked to perform an action using the interaction or the user is asked to think about or describe the interaction. Although there are exceptions, performance measures tend to be more objective and are also more likely to be based on implicit knowledge while reflective measures tend to be more subjective and more likely to be based on explicit knowledge.

It has been shown on numerous occasions that users' subjective reflections of their interactions do not match performance measures (e.g. Hornbaek and Law, 2007). For instance, users might believe adding color to a display improves their specific task performance, but in reality it does not (Jeffrey and Beck, 1972). Therefore, making design decisions based on subjective data may led to poorer performing systems (Bailey, 1993). According to Kissel (1995), subjective ratings and objective measures do not tightly correspond, but users with more experience with a system have a better connection. Thus, in some cases, there appears to be a positive correlation between subjective reflection and actual performance, but many contradictions remain apparent (Nielson and Levy, 1994). This has led researchers to recommend employing both subjective and objective measures (Andre and Wickens, 1995).

Some measures assume that the user has reflective insight into the intuitive experience. For example, the Questionnaire for the Subjective Consequences of Intuitive Use (QUESI; Naumann and Hurtienne, 2010) asks users to consider the degree to which the interface was easy to learn, easy to use (effort required during use and number of errors made), helpful for completing tasks and familiar to them. Although the questionnaire may not explicitly ask if the interaction was intuitive, it assumes that this information can be gathered indirectly via user's self-report. In contrast to the subjective components of the QUESI, the number of errors a user makes when using an interface, is an objective measure of the ease that should accompany an intuitive interaction. Similar indicators of intuitiveness include disfluencies or 'gaps' during concurrent verbal protocol, and the time it takes to complete a task (e.g. Blackler *et al.*, 2004).

Blackler *et al.* (2003) also used a combination of subjective and objective measures to assess and predict the presence of intuitive interactions. Participants were first asked to complete two tasks with a digital camera they were unfamiliar with, under think aloud conditions. Their interactions with the camera were recorded and coded. After finishing the two tasks, participants completed the Technology Familiarity Questionnaire—a survey that assesses user experience with similar devices and interactions—and participated in a structured interview during which they were asked to rate how familiar they were with each feature on the interface, what similar interaction experiences they have had, and how the interactions fit with their expectations. The results of the Technology Familiarity Questionnaire and the structured interview provide clear opportunities to discover how past experiences affect the user's representation of the interaction. Participant errors, when analyzed in the context of the responses to the structured interview, also have the potential to reveal cases of negative transfer. A prime example of this is reported in Blacker *et al.* (2003, p. 500) when several individuals used the wrong camera button to complete a task. During the structured interview, it was revealed that although they had pressed the wrong button, their decision to press that button was not baseless given that the button had been used in a similar capacity before. In this case, the error revealed something more than an interaction that is difficult to use, it revealed something about the user's representation of the interaction. This information emerged from the combination of objective and subjective data.

1.4. Memory types and elicitation methods

With the variety of methods used to assess intuitive interactions, we wondered if both performance and reflective measures would be equally effective in assessing affordances, conventions and biases. In cognitive psychology, it has been recognized that learning can be implicit and explicit (Frensch and Rünger, 2003; Mathews *et al.*, 1989) and that the manner in which knowledge is learned affects how that knowledge is retrieved. Typically, implicit knowledge is gained from repeated exposure to some consistency in the environment. The consistency can be learned even if the individual is unable to consciously identify the pattern (Reber, 1967). After the consistency is learned, the individual is often unable to verbalize how he or she acquired the information (Broadbent *et al.*, 1986). For example, a person can learn to balance on a bicycle after gaining sufficient experience. He may be fully aware that he can balance on a bicycle and can tell you when he has successfully balanced on a bicycle, but that does not mean he can accurately describe how he has accomplished that feat. In contrast, explicit knowledge can be gained in a single exposure and the individual is typically aware both of how he gained the information and how he has access to that information. To illustrate these two types of knowledge, consider how one knows the locations of keys on a keyboard. When expert users are asked, 'where is "h" located on the keyboard?' they might pause and place their hands out in front of them on an imaginary home row and move their right index fingers to the left. This is a clever strategy; the user cannot immediately verbalize his knowledge, so he observes his own actions to provide an answer. However, if a user were asked, 'where is "w" located on the keyboard?' without hesitation he might reply, 'next to the "q" key'. What is the difference between the first and second question? The user must access procedural knowledge, which is based on implicit memory, to answer the first question. In contrast, the user can access declarative knowledge, which is based on explicit memory, to answer the second question—the American (*qwerty*) keyboard is named by the top left hand row of keys Q, W, E, R, T and Y. An interaction could rely on implicit knowledge, explicit knowledge or both types of knowledge. Therefore, discovery of successful interactions is complicated by the fact that users may not be able to describe or accurately explain some of their actions.

1.5. Empirical exploration of knowledge elicitation methods

This experiment was designed to test whether or not varying levels of intuitive interactions—affordances, conventions and biases—necessitate different methods for eliciting *accurate* user knowledge. To accomplish this, we used the same interactions as Still and Dark (2008); users performed a simple button-pressing task acting on either affordance or non-affordance conditions to 'move' up, down, left or right. We contend that when users complete the button-pressing task, they utilize previous experience. If they have extensive and consistent experience with an interaction, their mental representation of the 'rules' guiding the interaction is based on multiple types of inputs and interaction episodes that have been integrated into a shared representation (Logan, 2002). In this case, the user may be able to complete the interaction automatically, but will not be able to verbally describe the knowledge guiding his behavior (Reingold and Merikle, 1988). Therefore, we predict that performance measures (e.g. accuracy, response time) would be more accurate than reflective measures (e.g. self-report of knowledge, familiarity rating) for determining user knowledge of conventional interactions. It is more difficult to make predictions about biased interactions; as a class of interactions, they can vary greatly in terms of their familiarity and automaticity. If the biased interaction is associated with a high level of ambiguity, it is possible that either performance or reflective measures would accurately capture user knowledge. The ambiguity in the interaction essentially prevents the mental representation from being applied implicitly; therefore, users might have some insight into the knowledge supporting their actions. Finally, although affordances are acted on automatically, we predict that the interaction may be so direct either type of measure would suffice.

In addition to investigating methods of knowledge elicitation for affordances, conventions and biases, we investigated whether or not these methods could contaminate one another. This possibility was examined by having half of the users complete a button-pressing task (objective, performance-based method of knowledge elicitation) before completing a survey (subjective, user reflection on button-to-action mappings), while having the remaining users complete the survey before the button-pressing task. There were experimental and practical reasons to examine order effects. From an experimental perspective, we wanted to use a within-subjects design and were concerned that there could be carryover from the different methods of knowledge elicitation. If left unaccounted, it would decrease the internal validity of the study. From a practical perspective, we wanted to have data that would inform best practices in knowledge elicitation. If there were no order effects, then designers could feel free to use the methods, however, they choose. However, if order effects were obtained, designers would have to consider the cost associated with using both performance and reflective measures. Would they need to abandon one of the methods? Should they have different groups of users for each measure? Based on previous research, it was unclear whether or not order effects would emerge in this experiment.

2. METHOD

2.1. Participants

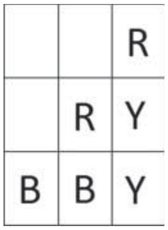
The university institutional review board approved all experimental procedures. Thirty undergraduate volunteers (27 right-handed) were recruited to participate in exchange for course credit in an introductory psychology course.

2.2. Stimuli and apparatus

Every participant completed both the computer interaction (i.e. button-pressing task) and the survey. The computer component required participants to respond by key presses when presented with directional cues; three two-button key configurations were tested for each participant. The survey provided participants with a scenario that required explicit descriptions of how they would map certain buttons to directional movements given three pairs of key configurations.

This study was created and executed within E-prime experimental presentation software (Psychology Software Tools, Inc., www.pstnet.com). On each trial, one of four directional cue words (*Up*, *Down*, *Left* or *Right*) was presented centrally in 48 point Arial font. Responses were collected through a PS/2 keyboard's numeric keypad. Only three pairs of keys (six buttons) were used and those keys were covered with colored stickers. Red stickers—covering keys 5 and 9—indicated the diagonal configuration. Blue stickers—covering keys 1 and 2 —indicated the horizontal configuration. Yellow stickers—coving keys 3 and 6—indicated the vertical configuration (see Fig. 2).

The survey was printed on paper and completed using a pencil. It provided a scenario in which participants explicitly described their button-to-action mappings (see Appendix: Survey). The scenario created a real-world context for the user's button-to-action mapping task. The scenario included the



		R
	R	Y
B	B	Y

Figure 2. Visual representation of the two-button pairs.

three key configurations used in the computer portion of the experiment. The survey instructed participants to put an arrow next to each button indicating which direction it ought to be used; they were asked to map up and down and then for left and right. Additionally, the survey asked the user to justify why they assigned the given directions to the buttons.

2.3. Procedure

A mixed experimental design was used with knowledge elicitation (computer interaction, survey response) and key configuration (diagonal, horizontal, vertical) manipulated within-subjects and order of knowledge elicitation tasks (survey first, computer interaction first) manipulated between subjects. The order in which participants completed the knowledge elicitation tasks was randomly assigned such that half of the participants took the survey first, while the other half interacted directly with the computer first. All participants completed both the survey and computer components of the experiment. When completing the survey, participants first mapped the button to its directional movement and then justified that design.

During the computerized portion of the experiment, the button-pressing task was manipulated across blocks within subjects. There were three button configurations (vertical, horizontal, diagonal) and four directional cues (up, down, left, right). Only two buttons were authorized for response within each of the three blocks: vertical (yellow buttons), horizontal (blue buttons) and diagonal (red buttons). Block order was counterbalanced across participants such that there were three possible block orders (yellow, blue, red; blue, red, yellow; red, yellow, blue) and each participant was assigned to one order. Each block contained 80 trials, 20 with each directional cue. Participants were asked to complete a large number of trials in order to collect reliable response time data. The order of cues (up, down, left, right) within a block was randomized for each participant.

Participants were instructed at the beginning of each block to place their fingers on the appropriate keys using the same hand.[1] Further, they were told to use those keys to move in the cued direction to the best of their ability given the current key configuration. Therefore, participants often faced an ambiguous situation in which they were not instructed *how to act* on the given button configuration, nor was any feedback given. They were also encouraged to respond as accurately and as quickly as possible. These instructions were intended to encourage participants to make speeded responses and intended to indicate that one button press was in fact correct for each directional cue. The instructions also ensured that participants' fingers were on the correct buttons before the onset of a block of trials.

After the instructions, each trial proceeded as follows: 'Get Ready' was presented for 2000 ms followed immediately by the presentation of a randomly selected directional cue (Up, Down, Left or Right). The next trial began when the participant pressed one of the buttons (only the two buttons assigned for the block would advance the experiment). New instructions were given at the beginning of each block. Participants completed the experiment in ∼15 min.

2.4. Coding and measures

Button response choice was analyzed in the context of expected button-to-action mappings (see Table 1). These expected mappings and their categorizations (affordance, convention or bias) were based on the results of previous experiments (Still and Dark, 2008, 2010). The affordance mappings for the button configurations were matched with their spatial arrangement (e.g. vertical buttons: top button for moving up and bottom button for moving down). The vertical top button was mapped with moving to the right and the bottom button for moving to the left (previously identified convention). The horizontal right button was mapped to moving up and the left button was mapped to moving down (previously identified bias). Each button configuration and direction combination along with the predicted mappings are presented in Table 1. Before analyzing the data, participants' button-pressing behavior and selected survey mappings (e.g. top button 'moves' right) were categorized as affordances, conventions or biases. Therefore, all of the data are presented in this context.

The survey captured participants' reflective button-to-directional movement mappings. These data are coded as a single response choice (the mapping reflects an affordance, a convention or a bias). In addition, the survey captured participants' justifications for their button-to-action mappings. The goal was to gain insight into their conscious, explicit, representations of the interactions. However, the justification for why a particular button mapping was chosen, required subjective interruption on our part. After collecting the data, similar user justifications were clustered together. This led to six, operationally defined categories: spatial location (e.g. upper, lower, above, below), movement (e.g. forward,

---

[1] Our sample of left-handed participants was too small to examine handedness effects. When we informally look across multiple experiments that use this task there are no apparent differences between left- and right-hand dominant users in preferred spatial button mappings.

backward, reverse), non-defined pairing (e.g. left is down, right is up), real-world objects (e.g. how elevators or escalators move or display directional cues), relative finger placement (e.g. vertical buttons: the middle finger is used for moving up and the index finger for moving down) and no response. Any justifications that appeared to fall between categories were resolved through a short discussion between the research assistants. The inter-observer rating accuracy between the research assistants was 79%. To conduct the statistical analysis, we combined all the spatial mappings justifications (i.e. spatial location, movement, non-defined pairings and relative finger placement) into one category and kept the real-world object justifications as another category.

3. RESULTS

All statistical tests used an alpha level of 0.05. Error bars in the figures represent the mean standard error. Both the button that was pressed and response times were recorded on each computerized trial. Responses on specific trials were excluded from the analysis if response times were <200 ms or >2000 ms; this filtering removed only 1% (68 outliers) of the total data. Three dependent variables were examined: button response choice, button-pressing response time and verbal justification for survey mappings. Condition means for participant button response choices appear in Fig. 3 (button-pressing task) and Fig. 4 (survey).

3.1. Button response choice

Participants displayed consistent button-to-action mapping either with our accepted scheme or with an inverted mapping, that is, once they selected a mapping, they retained it. This consistency probably arises from the existence of a logical constraint (i.e. if one button is left, the other must be right; one key should not reflect opposite directions). We defined ‘consistent’ responding as making the same response to a given cue 80–100% of the time. The following data reflect the proportion of individuals who adhered to the expected mapping scheme. Based on the condition means, participants generally adopted the button-action mappings that were established by Still and Dark (2008, 2010).

To further examine these mappings, a 3 × 2 × 2 mixed analysis of variance (ANOVA) with the variables interaction type (affordance, convention or bias), method of elicitation (performance or reflection) and order of knowledge elicitation tasks (computer or survey first), was conducted. The ANOVA revealed a main effect of interaction type, $F(2, 27) = 22.19$, $P < 0.001$, $\eta_p^2 = 0.622$ and a main effect of order of knowledge elicitation tasks, $F(1, 28) = 13.94$, $P = 0.001$, $\eta_p^2 = 0.332$. These main effects were qualified by a two-way interaction between interaction type and order, $F(2, 27) = 14.62$, $P < 0.001$, $\eta_p^2 = 0.096$. This interaction was explored through a series of independent samples t-tests for each interaction type. No other effects were statistically significant: method of elicitation, $F(1,28) = 0.48$, $P = 0.495$, $\eta_p^2 = 0.017$; method of elicitation X order, $F(1, 28) = 2.76$, $P = 0.108$, $\eta_p^2 = 0.090$;

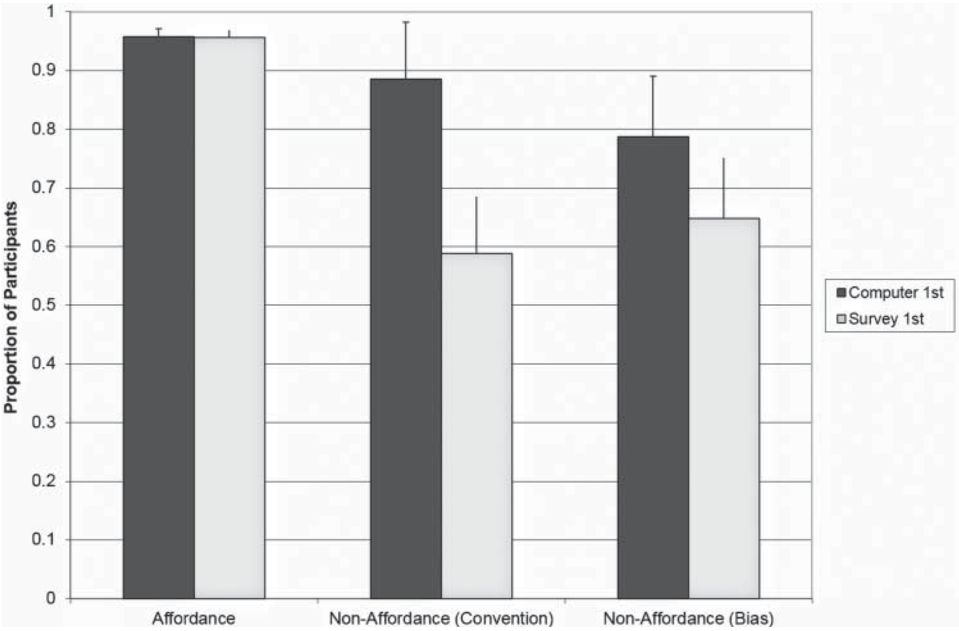


Figure 3. Proportion of individuals following predicted button-to-action mappings during the button-pressing task. Black bars represent button responses from participants who completed the button task first. Gray bars represent button responses from participants who completed the survey first.

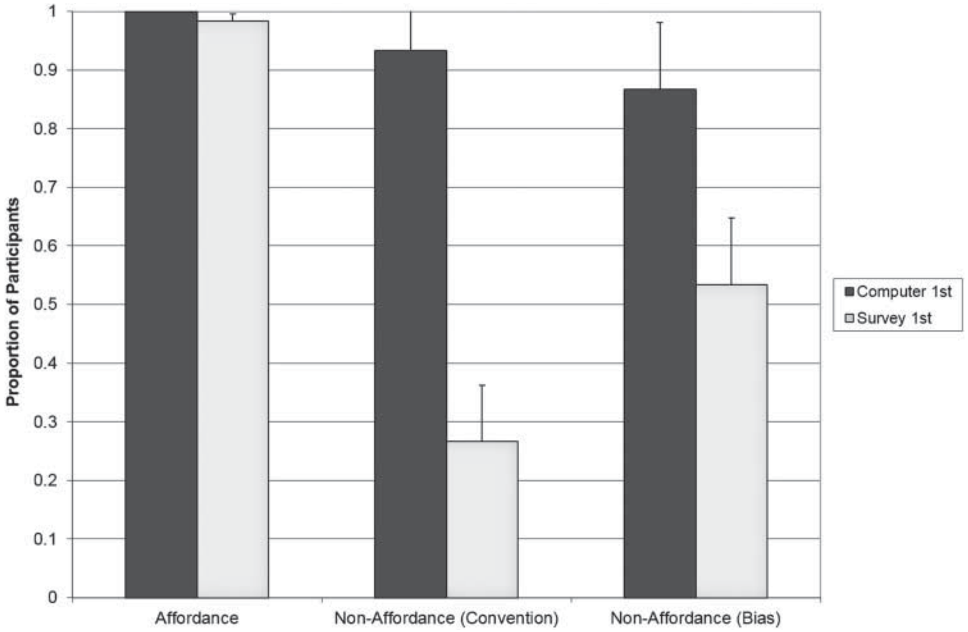


Figure 4. Proportion of individuals following predicted button-to-action mappings in the survey. Black bars represent survey responses when the button-pressing task was completed first. Gray bars represent survey responses when the survey was completed first.

interaction type X method of elicitation, $F(2, 27) = 1.44$, $P = 0.254$, $\eta_p^2 = 0.096$; interaction type X method of elicitation X order, $F(2, 27) = 1.53$, $P = 0.235$, $\eta_p^2 = 0.102$.

3.2. Affordances

An independent samples t-test confirmed that affordance-based button-to-action mappings were not affected by the order of knowledge elicitation tasks. Button presses elicited before taking the survey ($M = 0.96$, SEM $= 0.01$) and after the survey ($M = 0.96$, SEM $= 0.02$), were not statistically different, $t(28) = 0.10$, $P = 0.923$, $\eta_p^2 < 0.001$.

Similar results were obtained for survey data. When a mapping that was consistent with an affordance (e.g. mapping up and down on a vertical button configuration) was presented, participants' hypothetical button-to-action mappings were no different whether the survey was administered first ($M = 0.98$, SEM $= 0.06$) or was administered after completing the button-pressing task ($M = 1$, SEM $= 0$), $t(28) = 1$, $P = 0.326$, $\eta_p^2 = 0.034$. Together these findings provide evidence that affordances can be assessed equally well using a button-pressing task or a survey. In addition, both tasks were used with no carryover effects.

3.3. Conventions

In contrast to affordances, there were carryover effects for convention-based interactions (non-affordance interactions with a vertical key configuration). A t-test confirmed that button-to-action mappings differed depending on whether they were performed before ($M = 0.89$, SEM $= 0.06$) or after the survey ($M = 0.59$, SEM $= 0.12$), $t(28) = 2.18$, $P = 0.038$, $\eta_p^2 = 0.145$. When participants performed the button-pressing task first, a larger proportion of them adhered to the expected button-to-action mapping than when they completed the survey first. This provides evidence that just having users reflect on their interactions may change their performance on other related tasks. Importantly, when participants took the survey before completing the button-pressing task, their button-pressing behavior changed so much that this 'convention' no longer meets the standard for a convention established by Still and Dark (2008). It has turned into a bias!

Carryover effects were also obtained in the survey data. A t-test confirmed that users making convention judgments on the survey after completing the button-pressing task adhered more to the expected conventions ($M = 0.93$, SEM $= 0.07$) than those who completed the survey first ($M = 0.27$, SEM $= 0.12$), $t(28) = 4.91$, $P < 0.001$, $\eta_p^2 = 0.463$. These results suggest that most participants ($\sim$70%) who completed the survey first reported that they would use a non-affordance button mapping that is the *opposite* of how most respond to the button task. Interestingly, participants who completed the survey after the button-pressing task showed the opposite pattern of results. It is possible that participants who acted on the button configurations first, simply referred to that experience when they later completed their surveys, much like a typist observing the movement of his fingers in order to figure out where 'h' is located on a keyboard. Additional evidence for this claim is found in the qualitative data (under 'Survey Justifications'); participants were more likely to use spatial mapping justifications if they acted on the computer before completing the survey. These results clearly suggest that the elicitation of conventions depends on the method being used to examine the interaction and the presence of other elicitation methods.

3.4. Bias

No carryover effects were observed for bias interactions in the button-pressing task. The t-test revealed no significant difference between button presses whether captured before the survey ($M = 0.79$, $\text{SEM} = 0.09$) or after the survey ($M = 0.65$, $\text{SEM} = 0.12$), $t(28) = 0.946$, $\eta_p^2 = 0.031$, $P = 0.352$. Although numerically the results are in the same direction as those obtained for conventions, the overall variability was much greater in the bias condition.

In contrast, a t-test revealed that biased judgments are affected by whether participants completed the survey first ($M = 0.53$, $\text{SEM} = 0.13$) or after the button-pressing task ($M = 0.87$, $\text{SEM} = 0.09$), $t(28) = 2.07$, $P = 0.048$, $\eta_p^2 = 0.132$. Similar to conventions, it appears that participants may use their experience in biased interactions during the button-pressing task to inform their key designations within the survey. This could explain why the bias appears stronger after the button-pressing task.

3.5. Computer button-pressing response times

A 3 × 2 mixed ANOVA with the variables, interaction type (affordance, convention or bias) and order of knowledge elicitation tasks (computer or survey first) was conducted on button choice response times. The ANOVA revealed a significant main effect of interaction type, $F(2, 27) = 3.65$, $P = 0.040$, $\eta_p^2 = 0.213$. *Post hoc* tests using the Bonferroni correction revealed that affordance-based interactions ($M = 563\,\text{ms}$, $\text{SEM} = 16$) were only marginally faster than convention-based interactions ($M = 602\,\text{ms}$, $\text{SEM} = 18$), $P = 0.054$. Bias-based interactions ($M = 597\,\text{ms}$, $SEM = 21$) were not statistically different than affordances ($P = 0.210$) or conventions ($P = 1.00$) given their larger variability. No other effects were statistically significant: order, $F(1, 28) = 0.31$, $P = 0.581$, $\eta_p^2 = .011$; interaction type X order, $F(2, 27) = 1.00$, $P = 0.382$, $\eta_p^2 = 0.069$.

3.6. Survey justifications

One participant failed to provide any explanation for the survey mapping and, therefore, was excluded from the survey justification analyses. The coded data (proportion of spatial mapping and real-world responses) reveal that button-to-action mappings associated with affordances were always described using spatial location (100%) regardless of task order. This perfect consistency precluded additional statistical analyses for affordances and reinforced the idea that affordances can be directly mapped on to an interaction.

Chi-square tests were used to examine differences in the proportions of justifications given (spatial mapping or real-world) to interaction types (convention or bias) based on the order of knowledge elicitation (survey first or second). When the survey was completed first, the same proportions of justifications were obtained for conventions and biases (Real-World 71%; Spatial Mapping 29%). Interestingly, when the survey followed the button-pressing task, real-world justifications decreased for both conventions and biases. This numerical trend was not significant for conventions (53%), $\chi^2(1, N = 29) = 1$, $P = 0.316$, Cramer's $V = 0.186$. However, the decreased reliance on real-world justifications was significant for biases with the majority of participants (67%) using spatial mapping justifications instead, $\chi^2(1, N = 29) = 4.21$, $P = 0.04$, Cramer's $V = 0.381$.

We suggest the observed trend, from affordances being unaffected to biases being heavily affected, reflected the varying amount of stability in the mental representation. Bias interactions are more arbitrary, which makes them more susceptible to influence from recent interactions. Conventions stem from more consistent interactions that are based on implicit representations. These results suggest that when participants are asked to justify an action that is based on implicit knowledge, they are susceptible to giving incorrect descriptions of what their actual behavior would have been. This outcome is more likely to occur with implicit knowledge because the participants do not have conscious access to the processes that support the behavior. Paradoxically, these 'incorrect' justifications may appear more valid because the justifications for the selected button-action mappings are more informative—they seem to reveal other interactions from the user's history that support the convention.

4. DISCUSSION

It is often the case that designers need to know a user's current conventional knowledge in order to determine whether or not an interaction will be deemed intuitive. One way to obtain this information is to ask the users to perform the interaction task and observe their actions with no guidance. Another way to obtain this information is to ask the users to verbalize their predicted actions and reflect upon their interactions. Previous research suggests that methods of knowledge elicitation are most effective when they 'match' the method in which the knowledge was acquired (Ericsson and Simon, 1993). Therefore, if it is the case that the representations underlying intuitive interactions are implicit, it is possible that implicit (e.g. performance-based button-pressing task) and explicit (e.g. reflective survey) forms of knowledge elicitation would yield different results. In the event that these methods of knowledge elicitation do produce different results, it is also important to know whether or not one method would contaminate the other.

The present experiment was designed to investigate the accuracy and effectiveness of performance-based and reflective methods when examining three levels of intuitive interactions. To vary the intuitiveness of an interaction, participants were asked to 'move' in one of four directions using one of three different key configurations; the three key configurations promoted affordance and non-affordance (convention and bias) interaction conditions. It was assumed that affordances represent naturally intuitive interactions, conventions represent learned intuitive interactions and biases represent interactions with which the user lacks either consistency or experience, making them unintuitive interactions in their current state. To examine effective knowledge elicitation, participants completed a survey—asking them to assign and justify button-to-action mappings—either before or after completing a button-pressing task.

4.1. Effective methods of knowledge elicitation

The results of this experiment reveal that the implicit and explicit methods of knowledge elicitation are equivalent when examining affordance-based interactions, but are not equivalent when examining non-affordance-based interactions. This difference is clear when comparing the first task participants completed. Participants who completed the computer-based button-pressing task first, adopted the expected mapping for conventions (89% agreement) and the expected mapping for biases (79% agreement). In stark contrast, when participants completed the survey first, only 27% selected the expected conventional mapping in the survey and 53% adopted the expected bias mapping. Based on these data, the bias interaction would still be classified as a bias, but it appears much weaker compared with the button-press data. More disturbing is the fact that these survey data indicate that the convention is not a convention or even a bias! If these data were used to make decisions about button-to-action mappings, the convention would be mapped opposite of the preferred interaction that was observed in the button-pressing task.

The knowledge a designer elicits from user non-affordance interactions is clearly sensitive to the elicitation method. Even though both conventions and biases are learned interactions, conventions are more sensitive to the elicitation method. On the surface, this finding is unexpected. Mental representations of conventions are purportedly stronger and less ambiguous than those of biases, should not this protect them from distortion? We propose that those very features are what lead to disparate results. Conventions are learned interactions that have been automatized. To achieve automation, the user must have repeated and consistent interactions. As the interactions can occur in a variety of contexts (e.g. different interfaces), the representation supporting a convention may be a compilation of experiences that cannot be accessed individually. Thus, the implicit representation of how an interaction should unfold may not be available for accurate self-report. When the user is overtly asked to describe the interaction, he may be able to generate an example of the interaction based on previous experience, but there is no guarantee that the example will be representative of the most common mapping he has encountered. In contrast, the mental representation for a biased interaction is associated with more ambiguity. While the users might usually experience one implementation of the interaction (bottom button for moving left), they also have experiences with alternative implementations (bottom button for moving right). In these cases, the user cannot automatically know which action to apply. This is true whether he is completing the button-pressing task or providing a button-action mapping in the survey. Thus, the mappings elicited from either method are more likely to be congruent compared with conventions.

Our explanation of the results depends on the assumption that conventions are automatically accessed and that this type of representation is most accurately assessed using methods that rely on implicit knowledge. We have also proposed that the representations supporting convention- and bias-based interactions differ primarily on the ambiguity associated with the representation. As our experiment was primarily intended to test contamination between knowledge-elicitation measures, we did not do an exhaustive test of these two assumptions. We instead based our assumptions on established theories and principles. Both of these assumptions could be tested, though, by introducing participants to a novel interaction and then tracking their performance over an extended period of time (e.g. months) using methods that rely on explicit and implicit knowledge. If a between-subjects design was used, the same novel interaction could be presented to both groups of participants with one group always having consistent interactions while the other group could experience inconsistent interactions (e.g. 60% mapped one way, 40% mapped another). This manipulation lays the groundwork for establishing a convention in one group and a bias in the other. The second manipulation would be to periodically elicit user knowledge about the interaction using methods sensitive to explicit or implicit knowledge. Because of the possibility of contamination, this manipulation should also be performed between subjects. If our assumptions are correct, we would expect to see the data obtained from the explicit and implicit measures diverge as the novel interaction becomes a convention. This shift should coincide with the increased automaticity with which the representation is accessed. A similar, but less pronounced, pattern would be expected in the bias condition.

4.2. Interactions between methods of knowledge elicitation

Based on the evidence that methods eliciting implicit and explicit knowledge are not equally effective in assessing varying levels of intuitive interactions, one might recommend simply using both methods and then comparing the results. That recommendation has, in essence, been made before. Andre and Wickens (1995) note that there can be a dissociation between users' subjective assessment of a system and their actual performance with a system. They propose that designers could measure user performance and preference, then, if the results differ, decide whether it is more important to ensure effective performance or user preference. Intuitive interaction researchers have also tacitly endorsed this strategy by using a variety of measures, some implicit (e.g. accurate usage), some explicit (e.g. asking users to explain or talk through their actions). The results of our experiment suggest that this strategy may inadvertently contaminate both types of data.

The most obvious carryover effects occurred when participants took the survey first and then completed the button-pressing task. There were no effects for affordance-based interactions, but there were clear order effects for non-affordance-based interactions. When participants complete the implicit button-pressing task first, they select similar button-to-action mappings whether they are acting on the buttons or are assigning mappings in the survey. In contrast, when participants complete the explicit survey first, they are less consistent in their button-to-action mappings. This inconsistency is greatest for convention-based interactions where only 27% of participants select the expected conventional interaction in the survey, but 59% of participants proceed to use the expected conventional interaction during the button-pressing task.

Although the outcome is more subtle, there were also carryover effects in participants' subjective explanations for the mappings they selected in the survey for non-affordance-based interactions. When participants completed the survey first, they tended to explain their convention- and bias-based interactions using examples of interactions with real objects (e.g. escalator, the left side goes up and the right goes down). On the surface, these explanations are valuable to designers because they provide concrete examples of exactly what the designer is looking for—i.e. the reason why an interaction is intuitive. When participants completed the button-pressing task first, they were less likely to justify the mappings using interactions with real objects. This shift was most pronounced for bias-based interactions where participants were much more likely to explain their survey selection using basic spatial justifications (e.g. right is for going forward). Although these justifications may be accurate, it is possible that they seem less informative as they essentially mimic the information already provided when they specified which button should be matched to a specific direction.

Carryover effects are extremely important to consider. If multiple methods are used without testing for order effects, it is impossible for a designer to know whether or not their data are contaminated. Not only that, our experiment shows that carryover effects can be large. A convention—an interaction that over 80% of users typically agree upon—can disappear. Similarly, carryover effects can lead to qualitatively different explanations for why an interaction should act in a specified way.

4.3. User history and intuitive design

We have proposed that repeated and consistent experiences are required for an interaction to be intuitive, further, those interactions must have some degree of automaticity. From a design perspective, it is important, therefore, to consider what factors contribute to consistency and automaticity. Not surprisingly, user history plays a key role in determining what interactions are intuitive.

Interactions do not occur in a vacuum. Instead, every user carries representations of past interactions in long-term memory and spontaneously accesses those representations when similar interactions are encountered. When previous experience aligns with the current interaction, positive transfer can occur, making the interaction seem intuitive as the interaction proceeds as the user would expect. When previous experience does not align with the current interaction, negative transfer can occur (Besnard and Cacitti, 2005). Negative transfer, at minimum, impedes the user. Imagine a skilled *qwerty* keyboard typist attempting to type using a keyboard that has a different key configuration. Every time he attempts to type an A, a T appears. Although he may explicitly know where the T key is located and, ultimately, be able to select the appropriate key, he must stop to think about the key location. His expertise hinders his performance. In this case, cognitive resources must be diverted to letter selection, leaving fewer resources for higher level cognitive tasks (e.g. effectively communicating his ideas). Negative transfer can also lead to unsuccessful interactions (Still and Dark, 2013). Imagine a user visits a webpage and clicks on an underlined word. Nothing happens. Past experience dictates that the underlined word should be a hyperlink. This incorrect mapping of visual element to the user's expectation leads to an unintuitive interaction.

Not only does negative transfer affect users during their initial interaction with an interface, it can affect their ability to learn the new interaction (e.g. Keppel and Underwood, 1962). For example, underlined words in webpages often represent hyperlinks and are therefore strongly associated with clicking behavior. If a user encounters an underlined word that is not a hyperlink, he or she might try to click the word, thereby disrupting that specific interaction. This single encounter does not negate previous experience with hyperlinks. Therefore, it will take more time and effort to learn the new response to underlined words (do not click) than it would take to learn a response to a novel stimulus. New interaction experiences do not simply overwrite old experiences stored in long-term memory. Through repeated interaction, the representation associated with the interaction may be modified, even to the point that the new response becomes prepotent, but the previous experiences are not erased. From a design perspective, therefore, it is important not only to identify interactions users have encountered in the past (e.g. as in the Technology Familiarity Questionnaire by Blackler *et al.*, 2003), but also to actively sample interactions that may foster positive *and* negative transfer.

Generally speaking, an interaction must be familiar to the user in order for positive transfer to occur, but it is unclear what level of familiarity is required to facilitate an intuitive interaction. Based on the cognitive literature, only a partial match between a stimulus and a representation stored in memory is needed to elicit a familiarity response (Cleary *et al*., 2004, 2007). Cleary and Langley (2007) demonstrated this by asking participants to remember nonsense sentences (e.g. *efficient dreams write better umbrellas*) and then later testing them using studied sentences and novel nonsense sentences that either shared or did not share the same grammatical structure (e.g. *energetic trees demand silly frameworks* and *grimy inert stones stumble awkwardly*, respectively). Participants were more likely to misjudge novel sentences as being familiar when they shared syntax with the sentences they had studied. Thus, a feature of the structure of the stimuli (the syntax) was inadvertently stored in participants' memories and was later used to make memory judgments. Bolte and Goschke (2008) obtained similar results using fragmented images of common objects. After exposure to the images, participants were faster to recognize the name of the object in a lexical decision task compared with the names of objects that had not been seen during the experiment. These results suggest that being exposed to only one component of a larger representation (e.g. syntax, object features) can lead to the activation of the larger representation and may induce a feeling of familiarity.

An alternative explanation for these familiarity findings is that a signal emerges based on the ease, or fluency, with which a stimulus is processed and then that signal is interpreted based on the context (e.g. Jacoby and Dallas, 1981; Jacoby and Whitehouse, 1989). When the participant is exposed to a stimulus (or stimulus feature) for the second time, the stimulus will be processed more fluently. That ease of processing must then be attributed to some factor. The attribution process is sensitive to context. For example, if the participant remembers seeing a stimulus twice, she would likely feel that the second stimulus was easy to process because she had seen it before. If the participant does not know that she encountered the stimulus before, then she may misattribute the ease of processing to something else, like familiarity (cf. Jacoby and Whitehouse, 1989). Although it can be challenging to translate these findings into interaction design recommendations (e.g. what component of an interaction would be comparable to a stimulus feature like sentence syntax), the results suggest that it is worthwhile to continue parsing interface designs into their component interactions and those interactions into their component parts (cf. Blackler *et al*. 2003, 2004, 2010).

It is important to note that users may not be able to accurately describe why a stimulus is familiar—just as they are not always able to accurately describe how an interaction should work. User judgments can be based on vague signals. For instance, a reliable familiarity signal can be generated even when an individual cannot identify the stimulus; this phenomenon is referred to as recognition without identification (RWI; Cleary and Greene, 2000). An RWI experience can be elicited by asking participants to study a long list of words, and then instead of clearly showing the words during the recognition test, the words are briefly displayed so that participants often cannot report what they have seen. Interestingly, when asked to rate the familiarity of the 'unseen' words, participants reliably indicate that the studied words are more familiar than the non-studied words, thereby demonstrating successful use of the familiarity signal (Cleary and Greene, 2005). This effect is not limited to word stimuli, as similar findings have been obtained for famous faces, famous scenes and threatening stimuli (Cleary *et al.*, 2013). These findings demonstrate that a familiarity signal can be spontaneously generated and applied; as the participants could not even report what stimulus they had encountered, it follows that one cannot expect that they should be able to give an accurate report of the features that drove their response. In the context of intuitive interactions, this means that techniques beyond self-report or surveys are necessary to uncover the primitives of a familiar, and by proxy, intuitive design.

4.4. Affective component of intuitive interaction

One aspect of intuitive interactions that we have not addressed is the affective experience of the user. According to some researchers (e.g. Betsch, 2008a,b), 'feelings' play an integral part in an intuitive experience. Although the subjective experience of an intuitive interaction may not always reflect a truly intuitive experience (i.e. it can be affected by other factors), its contribution to intuition merits further research as affect has been clearly linked to both familiarity and fluency. We believe that further examination of these relationships can be used to inform intuitive design.

While familiarity is often colloquially referred to as a feeling, empirical evidence also supports this assertion. Morris *et al.* (2008) found a relationship between familiarity and autonomic arousal (measured via skin conductance response) within the standard RWI paradigm. Goldinger and Hansen (2005) also report a link between 'feelings' and familiarity. In their study participants were more likely to indicate that an item was familiar if it had been paired with a low-level tone—creating an undetectable vibration in participants' chairs—than if it had been presented alone. The vibrations served to trigger an autonomic response that was misattributed to familiarity with the stimulus.

In a related line of research, Topolinski and Strack (2009) examined the contributions of processing fluency and affect in intuitive judgments. They proposed that increased processing fluency serves to briefly increase positive affect; that positive affect is then interpreted as an indicator of intuitiveness. Across 11 experiments, Topolinski and Strack manipulated processing fluency (e.g. visual contrast) and participant affect (e.g. affective priming). In doing so, they were able to manipulate participants' intuitive judgments, thereby providing support for their theoretical framework. Although familiarity and fluency researchers posit different explanations for the judgments participants make, both have identified a clear relationship with affect. Not only can the results of these studies be taken as evidence that affect also plays a role in intuition, we suggest that the techniques and measures used in these lines of research would be valuable additions to intuitive design research.

5. CONCLUSION

There are obvious merits to designing an intuitive interaction. At minimum, intuitive interactions are valuable because they can be completed with little effort, allowing users to focus limited cognitive resources on higher level tasks. The challenge is determining whether or not interactions are intuitive; this is not a trivial task. Users are continuously interacting with their environments and updating their representations, making ‘intuitiveness’ a moving target. To further complicate the issue, our study shows that different methods of knowledge elicitation can produce conflicting results for some types of intuitive interactions (i.e. conventions, but not affordances). In addition, we found that the implicit and explicit tasks used in the study contaminated each other. Therefore, it is not advisable to simply use multiple measures and then compare the results. Instead, we recommend that in order to get an accurate assessment of user knowledge, the elicitation task must be congruent with the mental representations supporting the interaction.

REFERENCES

Allen, B.G. and Buie, E. (2002) What’s in a word? The semantics of usability.Interactions, 9, 17–21.

Andre, A.D. and Wickens, C.D. (1995) When users want what’s not best for them. Ergon. Des., 3, 10–13.

Bailey, R.W. (1993) Performance vs. Preference. In: Proc. Human Factors & Ergonomics Society 37th Annual Meeting, Seattle, WA, pp. 282–286. Santa Monica, CA.

Bergum, B.O. and Bergum, J.E. (1981) Population Stereotypes: An Attempt to Measure and Define. In: Proc. Human Factors Society 25th Annual Meeting, Rochester, NY, pp. 662–665. Santa Monica, CA.

Besnard, D. and Cacitti, L. (2005) Interface changes causing accidents. An empirical study of negative transfer. Int. J. Hum.-Comput. Stud., 62, 105–125.

Betsch, C. (2008a) Chronic Preferences for Intuition and Deliberation in Decision-Making: Lessons Learned about Intuition from an Individual Differences Approach. In Plessner, H., Betsch, C. and Betsch, T. (eds) Intuition in Judgment and Decision Making, pp. 231–248. Lawrence Erlbaum, New York.

Betsch, T. (2008b) The Nature of Intuition and its Neglect in Research on Judgment and Decision-Making. In Plessner, H., Betsch, C. and Betsch, T. (eds) Intuition in Judgment and Decision Making, pp. 3–22. Lawrence Erlbaum, New York.

Blackler, A. (2008) Intuitive Interaction with Complex Artifacts: Empirically-Based Research. VDM Verlag, Saarbrucken, Germany.

Blackler, A., Popovic, V. and Mahar, D. (2003) The nature of intuitive use of products: an experimental approach. Design Stud., 24, 491–506.

Blackler, A., Popovic, V. and Mahar, D. (2004) Studies of Intuitive Use Employing Observation and Concurrent Protocol. In Marjanovic, D. (ed.) Proc. Design 2004-The 8th Int. Design Conf., pp. 135–142.Dubrovnik, Croatia.

Blackler, A. Popovic, V. and Mahar, D. (2010) Investigating users’ intuitive interaction with complex artifacts. Appl. Ergon., 41, 72–92.

Bolte, A. and Goschke, T. (2008) Intuition in the context of object perception: intuitive gestalt judgments rest on the unconscious activation of semantic representations. Cognition, 108, 608–616.

Bolte, A. Goschke, T. and Kuhl, J. (2003) Emotion and intuition: effects of positive and negative mood on implicit judgments of semantic coherence. Psychol. Sci., 14, 416–424.

Broadbent, D.E., Fitzgerald, P. and Broadbent, M.H.P. (1986) Implicit and explicit knowledge in the context of complex systems. Brit. J. Psychol., 77, 33–50.

Cleary, A.M. and Greene, R.L. (2000) Recognition without identification. J. Exp. Psychol. Learn., 26, 1063–1069.

Cleary, A.M. and Greene, R.L. (2005) Recognition without perceptual identification: a measure of familiarity? Q. J. Exp. Psychol., 58A, 1143–1152.

Cleary, A.M. and Langley, M.M. (2007) Retention of the structure underlying sentences. Lang. Cognitive Proc., 22, 614–628.

Cleary, A.M., Langley, M.M. and Seiler, K.R. (2004) Recognition without picture identification: Geons as components of the pictorial memory trace. Psychon. Bull. Rev., 11, 903–908.

Cleary, A.M., Morris, A.L. and Langley, M.M. (2007) Recognition memory for novel stimuli: the structural regularity hypothesis. J. Exp. Psychol. Learn., 33, 379–393.

Cleary, A.M., Ryals, A.J. and Nomi, J.S. (2013) Intuitively detecting what is hidden within a visual mask: familiar-novel discrimination and threat detection for unidentified stimuli. Mem. Cognit., 41, 989–999.

De Vries, M., Holland, R.W. and Witteman, C.L.M. (2008) Fitting decisions: mood and intuitive versus deliberative decision strategies. Cogn. Emot, 22, 931–943.

Epstein, S. (2010) Demystifying intuition: what it is, what it does, and how it does it. J. Psychol. Inq., 21, 295–312.

Ericsson, K.A. and Simon, H.A. (1993) Protocol Analysis: Verbal Reports as Data. MIT Press, Cambridge, MA.

Frensch, P.A. and Rünger, D. (2003) Implicit learning. Curr. Dir. Psychol. Sci., 12, 13–18.

Gibson, J.J. (1979) The Ecological Approach to Visual Perception. Houghton Mifflin, Boston.

Goldinger, S.D. and Hansen, W.A. (2005) Remembering by the seat of your pants. Psychol. Sci., 16, 525–529.

Hornbaek, K. and Law, E.L.-C. (2007) Meta-Analysis of Correlations Among Usability Measures. In: Proc. SIGCHI Conf. Human Factors in Computing Systems, San Jose, CA, 617–626.

Hurtienne, J. and Israel, J.H. (2007). Image Schemas and their Metaphorical Extensions—Intuitive Patterns for Tangible Interaction. In proceedings of the 1st international conference on tangible and embedded interaction (TEI), pp. 127–134.

Hutchins, E.L., Hollan, J.D. and Norman, D.A. (1986) Direct Manipulation Interfaces. In: Norman D.A. and Draper, S.W. (eds) User Centered System Design: New Perspectives on Human–Computer Interaction, pp. 87–124. Lawrence Erlbaum, Hillsdale, NJ.

Jacoby, L.L. and Dallas, M. (1981) On the relationship between autobiographical memory and perceptual learning. J. Exp. Psychol. Gen., 110, 306–340.

Jacoby, L.L. and Whitehouse, K. (1989) An illusion of memory: false recognition influenced by unconscious perception. J. Exp. Psychol. Gen., 118, 126–135.

Jeffrey, T.E. and Beck, F.J. (1972) Intelligence Information from Total Optical Color Imagery (U.S. Army Behavior and Systems Research Lab, Research Memorandum 72-4). http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=ADA079386

Kannengiesser, U. and Gero, J.S. (2012) A process framework of affordances in design. J. Des. Issues, 28, 50–62.

Keppel, G. and Underwood, B.J. (1962) Proactive inhibition in short-term retention of single items. J. Verbal Learning Verbal Behav., 1, 153–161.

Kissel, G.V. (1995) The Effect of Computer Experience on Subjective and Objective Software Usability Measures. In Katz, I., Mack, R. and Marks, L. (eds) Conf. Companion on Human Factors in Computing Systems, pp. 284–285. ACM, NY.

Logan, G.D. (2002) An instance theory of attention and memory. Psychol. Rev., 109, 376–400.

Mathews, R.C., Buss, R.R., Stanley, W.B., Blanchard-Fields, F., Cho, J.R. and Druhan, B. (1989). Role of implicit and explicit processes in learning from examples: a synergistic effect. J. Exp. Psychol. Learn., 15, 1083–1100.

Moors, A. and De Houwer, J. (2006) Automaticity: a theoretical and conceptual analysis. Psychol. Bull., 132, 297–326.

Morris, A.L., Cleary, A.M. and Still, M.L. (2008) The role of autonomic arousal in feelings of familiarity. Conscious. Cogn., 17, 1378–1385.

Naumann, A. and Hurtienne, J. (2010) Benchmarks for Intuitive Interaction with Mobile Devices. In proceedings of the 12th international conference on Human Computer Interaction with mobile devices and services. Libson, Portugal, pp. 401–402.

Naumann, A., Hurtienne, J., Israel, J.H., Mohs, C., Kindsmuller, M.C., Meyer, H.A. and Hublein, S. (2007) Intuitive Use of User Interfaces: Defining a Vague Concept. In: Harris, D. (ed.), Engineering Psychology and Cognitive Ergonomics, pp. 128–136. Springer, Heidelberg.

Neisser, U. (1976) Cognition and Reality. W.H. Freeman, San Francisco.

Nielson, J. and Levy, J. (1994) Measuring usability: preference vs. performance. Commun. ACM, 37, 66–75.

Norman, D.A. (1988) The Psychology of Everyday Things. Basic Books, New York.

Norman, D.A. (1999) Affordance, conventions, and design. Interactions, 6, 38–42.

Proctor, R. W. and Vu, K.-P. L. (2006) Stimulus-response Compatibility Principles: Data, Theory, and Application. CRC Press, Boca Raton, FL.

Raskin, J. (1994) Intuitive equals familiar. Commun. ACM, 37, 17.

Reber, A.S. (1967) Implicit learning of artificial grammars. J. Verbal Learn. Verbal Behav., 6, 855–863.

Reingold, E.M. and Merikle, P.M. (1988) Using direct and indirect measures to study perception without awareness. Percept. Psychophys., 44, 563–575.

Shiffrin, R.N. and Schneider, W. (1977) Controlled and automatic human information processing: II. Perceptual learning, automatic, attending, and a general theory. Psychol. Rev., 84, 127–190.

Simon, H.A. (1969/1981) The Sciences of the Artificial (2nd edn). MIT Press, Cambridge, Massachusetts.

Spool, J.M. (2005) What makes a design seem 'intuitive'? User Interface Engineering. http://www.uie.com/articles/design_intuitive/ (accessed November 20, 2013).

Still, J.D. and Dark, V.J. (2008) An Empirical Investigation of Affordances and Conventions. In: Gero, J. S. and Goel, A. K. (eds) Design Computing and Cognition, pp. 457–472. Springer, Netherlands.

Still, J.D. and Dark, V.J. (2010) Examining working memory load and congruency effects on affordances and conventions. Int. J. Hum.-Comput. Stud., 68, 561–571.

Still, J.D. and Dark, V.J. (2013) Cognitively describing and designing affordances. J. Design Stud., 34, 285–301.

Topolinski, S. and Strack, F. (2009) The architecture of intuition: fluency and affect determine intuitive judgments of semantic and visual coherence and judgments of grammaticality in artificial grammar learning. J. Exp. Psychol. Gen., 138, 39–63.

Wippish, W. (1994) Intuition in the context of implicit memory. Psychol. Res., 56, 104–109.

You, H. and Chen, K. (2007) Applications of affordances and semantics in product design. Design Stud., 28, 23–28.

APPENDIX

SURVEY

Imagine you were just hired by J.S. Innovations to help design their new portable Mp3 player. The company wants

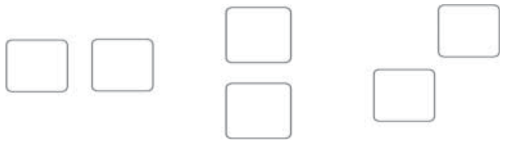


Figure A1. Hypothetical two-button configurations.

to improve the player's efficiency (i.e. faster to navigate through the music menus). Their old Mp3 devices had four buttons for navigating the music menus with each button either corresponding to up, down, left and right directional movement. In order to meet their efficiency goal, the design was limited to only two buttons. The company has provided you with three different two-button spatial configurations (see Fig. A1). Your job is to determine which two buttons should be used for up, down, left and right. After you have decided how the buttons and movement should be mapped, further testing by the company will be completed to determine whether the two-button configuration is actually more efficient.

Please, consider each of the two-button spatial configurations.

INSTRUCTIONS

For each button configuration, you will be asked to assign the directions: Up, Down, Left and Right to the given buttons by putting an arrow next to the button indicating which direction it will be used for. After you have assigned directions to the buttons, it is important that you explain, in detail, why you chose that button to move in that direction.

Please, provide justification for why you assigned the given directions to these buttons.