The ATB Framework: Quantifying and Classifying Epistemic Strategies in Tangible Problem-Solving Tasks

Augusto Esteves[1], Saskia Bakker[2], Alissa N. Antle[3], Aaron May[3], Jillian Warren[3], Ian Oakley[4]
[1]Exact Sciences and Engineering Center, University of Madeira, Funchal, Portugal
[2]Department of Industrial Design, Eindhoven University of Technology, Eindhoven, the Netherlands
[3]School of Interactive Arts and Technology, Simon Fraser University, Surrey, B.C., Canada
[4]Department of Human and Systems Engineering, UNIST, Ulsan, Republic of Korea
augustoeae@gmail.com, s.bakker@tue.nl, {aantle, amay, jlw29}@sfu.ca, ian.r.oakley@gmail.com

ABSTRACT

In task performance, pragmatic actions refer to behaviors that make direct progress, while epistemic actions involve altering the world so that cognitive processes are faster, more reliable or less taxing. Epistemic actions are frequently presented as a beneficial consequence of interacting with tangible systems. However, we currently lack tools to measure epistemic behaviors, making substantiating such claims highly challenging. This paper addresses this problem by presenting ATB, a video-coding framework that enables the identification and measurement of different epistemic actions during problem-solving tasks. The framework was developed through a systematic literature review of 78 papers, and analyzed through a study involving a jigsaw puzzle – a classical spatial problem – involving 60 participants. In order to assess the framework's value as a metric, we analyze the study with respect to its reliability, validity and predictive power. The broadly supportive results lead us to conclude that the ATB framework enables the use of observed epistemic behaviors as a performance metric for tangible systems. We believe that the development of metrics focused explicitly on the properties of tangible interaction are currently required to gain insight into the genuine and unique benefits of tangible interaction. The ATB framework is a step towards this goal.

Author Keywords

Epistemic actions; tangible interaction; video-coding.

ACM Classification Keywords

H5.2. User Interfaces: Theory & Methods.

INTRODUCTION AND RELATED WORK

Tangible interaction [24], an interface paradigm based on manipulating physical artifacts that both represent and control digital information, provides a compelling directness and physicality that has long made it a topic of study in research labs. As the field matures, successful products in areas such as musical performance (e.g. Reactable [14]) or play (e.g. Siftables [23]) illustrate how notions of users literally grasping information with their hands can be converted into rich, expressive and viable commercial systems. However, while users and designers continue to be drawn by the allure of physically handling digital data, it remains challenging to understand and quantify the genuine benefits of tangible interaction [30]. Indeed, we argue there is currently no systemic account of the underlying properties or qualities of tangible systems that can explain or justify their enduring appeal. There is no comprehensive way to answer questions regarding the true value provided by tangible systems to their users.

However, steps are being taken to develop such answers and explanations. One fertile source is the cognitive science literature that focuses on embodied (or situated) cognition. Work on this topic typically contends that the body and its interactions with the world play a central role in human thought and experience [25, 29]. More concretely, numerous authors assert that the worldly, physical representations of a problem, situation or task radically impact the strategies people employ to tackle it, their reasoning abilities and ultimately their overall performance [18, 21]. One aspect of this literature that is highly relevant to tangible interaction deals with how objects and their inherent properties (e.g., their ability to be stacked, ordered, annotated) can be leveraged by people to simplify or aid performance of information processing tasks [1].

Work has elaborated on this claim in a number of ways. For example, authors have described how users can manipulate objects to not only conduct pragmatic, goal-driven actions, but also to employ complementary actions [3] that involve exploring, testing, annotating or re-structuring a system state. In the field of Human-Computer Interaction (HCI), these kinds of activities are comprehensively documented in, for example, work analysis of air-traffic control – an activity in which operators rely heavily on physical paper strips to mediate their complex and safety-critical work tasks [22]. Other authors have depicted how users leverage external (non-mental) structures as tools to simplify

TEI '15, January 15–19, 2015, Stanford, California, USA..

ACM 978-1-4503-3305-4/15/01...$15.00.
http://dx.doi.org/10.1145/2677199.2680546

cognitive work [19]. For example, experienced jigsaw puzzle-solvers often cluster physical pieces together (e.g. by color) to simplify subsequent processes of visual search and recognition [2]. These kinds of account are important as they provide a basis for explaining the appeal and value of tangible systems – they cast light on the ways in which problem-centric tangible systems, such as the Senseboard [13] or Urp [27], really provide benefits to their users, potentially steering future design and development efforts.

However, applying principles from cognitive science to the design of tangible systems remains a challenging and intricate task. Although work on this topic remains embryonic, several distinct approaches exist. One key thread, instantiated as illustrative design frameworks [e.g. 10, 12, 21], aims to provide high level guidance and recommendations for how tangible systems can best be created. A second strand, more directly related to the literature on embodied cognition, seeks to expand our repertoire of metrics for understanding and assessing performance with tangible systems. Basically, it argues that theoretically grounded techniques that enable us to rigorously and empirically examine the mechanisms by which users rely on physical objects to aid their cognitive work will help us assess (and ultimately learn how best to design) such systems.

This latter approach has been explicitly explored in the context of Kirsh et al.'s [20] categorization of behaviors into either *pragmatic* or *epistemic* actions – a contrast between activities that users conduct to directly move towards a desired goal state versus those that are intended to alter the world to make the cognitive processes faster, more reliable or less taxing. While Kirsh originally explored this idea in the context of a purely virtual computer game (Tetris), Antle et al. [2] were arguably the first to apply it to a tangible system. Their work compares tangible and touch interfaces to a classic problem-solving task – a jigsaw puzzle – and they use video-coding analysis to classify user behaviors as either pragmatic or epistemic. Their results suggest that the tangible representation afforded more natural and efficient epistemic strategies such as clustering pieces to improve subsequent visual search, or relying on the elevated edges of the table to help structure the puzzle. In broadly similar work, Esteves et al. [6] report on user performance when playing a modified game of Four-in-a-row with three different interfaces (tangible, touch and mouse). Using video coding, specific epistemic actions relating to gesturing and receiving feedback on the game-board were measured and the results suggested that while epistemic activity was a significant component of all interfaces, users were more efficient in performing them with the tangible system.

While this work provides evidence for the conjecture that tangible interaction aids performance of epistemic actions, one major weakness is in terms of the granularity with which the activities are recorded. Basically, work that defines and discusses epistemic activity is typically highly specific and contextual – particular epistemic behaviors are described as being used to achieve particular ends in particular situations. However, the evaluation frameworks in HCI are broad and general [e.g. 2], often reducing the diversity of epistemic activity to a single categorical label. This paper aims to address this issue by expanding an existing video-coding framework [1, 2] that categorizes hand actions to include a detailed classification scheme for epistemic activity – the ATB (Artifact, Tool and Body) framework. This paper argues that this framework will contribute to our understanding of how epistemic actions are used in human problem-solving tasks, providing researchers with a tool to more systematically assess this complex type of behavior in tangible interaction.

In terms of HCI, this tool has two objectives. Firstly, it is intended as a mechanism to evaluate tangible systems in terms of the type, diversity and appropriateness of the epistemic actions they support, and in terms of the impact these actions can have on more traditional metrics such as performance time or errors. Secondly, in the long term, we argue that a series of such evaluations will result in a corpus of knowledge describing the use of epistemic actions in real tasks. This data can be used as the basis for grounded, practical design knowledge on how to create novel systems that truly support epistemic actions, and thus, improve our ability to design tangible interaction that is natural and meets the real needs of the user.

As such, this paper makes two contributions. Firstly, we describe a detailed framework of epistemic activity based on a systematic literature review of 78 papers (an early version of the framework was previously introduced in [5]). Although loosely related prior classifications exist [e.g. 18, 20], the framework presented in this paper is the first to be based on a systematic review, the first to aim for a focused, fine-grained description of epistemic behavior, and the first to be specifically directed towards the development of an actionable empirical tool for capturing and expressing observed epistemic actions. Secondly, this paper presents an initial experiment to explore the framework's reliability, validity and predictive power. This substantial lab study involved 60 participants across three countries completing a physical problem-solving task – a jigsaw puzzle. Three raters analyzed the data to support commentary on reliability. Validity is explored by contrasting the video-coding results among our purposely diverse participant group with other measures such as spatial ability tests and task completion rates and times. The outcomes of this study provide insights into epistemic activity and demonstrate the usefulness of the framework as an analytic tool that other researchers can apply in their own design and evaluation activities in the field of tangible interaction.

THE ATB VIDEO-CODING FRAMEWORK

The work in this paper builds on the action classification framework presented by Antle et al. [1, 2]. In their work, an action can be classified as either a *direct placement* (DP), an *indirect placement* (IP), or as *exploratory* (EXP). In the puzzle task they studied, a *DP* action corresponded to those situations where users already know where to place a piece before picking it up, leading to a fast and direct transition between acquiring a piece, moving to the final destination and correctly placing it. *IP* represented similar outcomes but described situations in which users are not initially certain of where to position the pieces they pick up. As such, they translated or rotated the piece while searching for its correct destination. Finally, *EXP* represented those actions where pieces do not end in their final and correct position. As with IP, if these intermediary actions make the task easier for the user they are considered epistemic (e.g., if a user organizes pieces into different piles for subsequent identification and retrieval).

Antle's framework [1, 2] provides basic features that enables the study of epistemic actions. It allows researchers to measure epistemic activity levels within a task, reporting on the frequency, duration, and moments at which epistemic actions occur. While valuable, we argue that this classification is too broad to fully articulate and explain the role of epistemic action in problem solving. For example, different types of epistemic action may be used in different tasks [6], and recording such variations in detail will better characterize the role and importance of epistemic activity. This paper argues that only by considering epistemic actions at a fine-grained level of detail will we be able to understand not just how many epistemic actions are performed during a task, but also which epistemic actions are chosen and to which purpose. In terms of tangible interfaces, a detailed classification scheme will support investigations of what interface elements facilitate what epistemic actions, quantifying the differences between novel systems, and allowing designers to tailor interaction that better supports users' natural, epistemic behaviors.

ATB Framework Development and Use

To develop the ATB framework we conducted an extensive literature review with the goal of capturing a wide range of epistemic activity descriptions [5]. A set of keywords was used to conduct a literature search on both Google Scholar and Science Direct. The search terms were 'epistemic action(s)', 'complementary action(s)' and 'complementary strategies'. The first 60 results from each of these searches were kept for further inspection. Additionally, papers referencing seminal work in the area (specifically [16] and [20]) and including the keywords defined above were also retained. Ultimately, 78 papers were obtained through this process – a typical number for meta-analysis papers in the area of HCI [e.g. 9]. Each paper was then inspected for any mention of actions that could be interpreted as epistemic, or were directly treated as epistemic, and quotes such as: "(...) *preparing the workplace, for example, by partially sorting nuts and bolts before beginning an assembly task in order to reduce later search* (...)" [26, p. 515] were extracted. These represented concrete examples of epistemic actions from research literature in a range of fields (such as mathematics, cognitive science, HCI and design) from the last three decades. A complete list of references for all 78 publications can be found at http://www.mysecondplace.org/ATB/ full-reference-list.txt.

In total, 335 quotes were compiled. Two of the authors then worked collaboratively to create an affinity diagram that identified different clusters of epistemic actions. Quotes judged to depict actions with unclear epistemic value were discarded. This process led to the identification of 20 types of epistemic action based on a subset of 225 of the original quotes. These were then grouped by actions performed with 1) task artifacts (e.g. objects marked with fiducials), 2) tools (e.g. a pencil that can be used for annotations) or 3) the users own bodies (summarized in Table 1). A full scheme

#	Epistemic actions
A1	**Manipulation of an artifact**
A2	Spatial arrangement of artifacts in relation to one another, the task environment, or the users
2.1	Cluster or group artifacts together
2.2	Divide workspace into several stations in which only a subset of actions are afforded
2.3	Place an artifact in a contrasting environment
2.4	Rearrange a representation
2.5	Clear and clean clutter
A3	Parallel use of two artifacts, two representations, or an artifact and a representation
A4	Artifact trial-and-error positioning
A5	Shuffle artifacts
A6	Compare an artifact with a possible destination or other artifacts
A7	Mark an artifact
A8	Test the state or response of a system, model or other user
	Manipulation of a tool
T9	Tag or annotate an artifact
T10	General notes and annotations
T11	Use of a tool to physically constraint the user or the use of other artifacts and tools
T12	Build a model or external representation
	Bodily action
B13	Use the body to externalize an internal process
B14	Talk or gesture to guide and direct attention
B15	Move the body, problem space, or representation

Table 1. A list of all the 20 types of epistemic actions present in the ATB video-coding framework, which groups actions by those performed with an artifact (an epistemic action in itself), a tool, or the user's body. Five types of epistemic action are grouped under A2, a broad type of action.

describing these 20 types of epistemic action can be found at http://www.mysecondplace.org/ATB/atb-framework.pdf.

These categories are then used as the basis for classifying behaviors through video-analysis, according to the following procedure. Firstly, raters should categorize actions as being either DP, IP, or EXP, as in Antle et al's framework [1, 2]. After this process is completed, raters should review each action that can contain epistemic activity (i.e. those coded as either IP or EXP) and match these to one (or more) of the 20 types of epistemic actions identified. For a graphical workflow of how to video-code with using the ATB framework, please consult http://www.mysecondplace.org/ATB/coding-flowchart.pdf. A coding scheme file was created to facilitate the process of video coding with Anvil, a free and popular video-coding tool. This file can be downloaded from http://www.mysecondplace.org/ATB/atb.xml.

APPLYING THE ATB FRAMEWORK: AN INITIAL STUDY

To determine the usefulness of the ATB framework in capturing and distinguishing among different epistemic actions, and the fundamental value and worth of this kind of information, we conducted an observational study of users performing a classical problem-solving task – a jigsaw puzzle. This task was selected as there is a large body of work on epistemic actions using puzzles in both HCI [e.g. 1, 2], and cognitive science [e.g. 15, 16, 17]. Furthermore, puzzle metaphors are commonplace in the design of tangible systems [e.g. 8]. The goal of this initial study was to explore the reliability and sensitivity of the framework, and to assess its internal and external validity, and predictive power. To meet these classic methodological objectives, a diverse participant pool was recruited (from Korea, Canada and the Netherlands) and a range of spatial and subjective workload tests were performed to establish the main results in a theoretical context.

Experimental Design and Participants

All participants in this study completed a single condition. In total, there were 60 participants, 20 of whom were Korean, 20 Dutch, and 20 Canadian – 10 male and 10 female participants from each nationality. The study also took place at three sites, one in each of these countries, with all participants residing in their respective countries of origin. Participant ages ranged from 20 to 76 ($M = 27.33$, $SD = 10.72$) and occupied a wide range of professions, from undergraduate to postgraduate students, to sailors, game designers, artists, drivers, and writers. Before the study, all participants filled in a brief online questionnaire to exclude puzzle hobbyists. To motivate participants to perform to the best of their capabilities, a prize of $25 (or equivalent) was awarded to the fastest participant to solve the puzzle in each of the three countries. In addition to this prize, Korean and Canadian participants received a $10 compensation for participating. Dutch participation was not compensated due to different funding policies in the three research groups.

Procedure

Each session involved a single participant solving two puzzles, and performing an additional test at the beginning and end of the study. All tasks were performed in small and otherwise empty offices, and all sessions followed the same structure. Sessions commenced with a brief introduction to the first task, a paper folding spatial ability test [4]. Participants were then introduced to the first jigsaw puzzle, an unmeasured practice task, which they were asked to solve in a maximum time of 10 minutes. This was followed by the main task of the study, a second jigsaw puzzle which participants were asked to solve in a maximum time of 15 minutes. Though the two puzzles presented different images, each consisted of 70 pieces and was 38x26cm in size. The order in which the puzzles were presented was the same for each participant. Before starting the main puzzle, participants were reminded of the monetary prize. During both puzzle tasks participants were left alone to ensure that their epistemic actions were unmediated and private. Both tasks were recorded on video for later analysis. At the end of the main puzzle, participants completed a subjective test to measure the perceived workload of the main task.

Measures

In addition to recording the time that it took participants to finish the puzzles, the following metrics were used:

Spatial ability (paper folding test): Upon starting the study, participants were required to solve two sets of spatial tests [4], each in under three minutes.

Video-coding framework: Several metrics were derived from the data obtained through the video-coding framework being examined. These include the mean number of pragmatic (coded as DP) and epistemic actions (coded as either IP or EXP) performed; and, the individual mean frequency of each of the 20 types of epistemic action in the framework. These frequency metrics include both aggregate and running means, on a minute-by-minute basis.

Subjective Workload: Each participant completed the NASA TLX [7] after the main puzzle task.

Participant group		**Spatial ability**	**Overall Workload**	**Overall time to finish**
Puzzle completion	Finished	15.11 (4.14)	8.74 (2.31)	9:38.10
	Incomplete	11.67 (6.27)	11.36 (2.75)	-
Gender	Male	15.61 (3.59)	9.36 (2.69)	10:49.22
	Female	13.13 (5.62)	9.19 (2.48)	10:46.26
Cultural background	Korean	17.48 (2.76)	8.79 (2.47)	12:44.00
	Dutch	15.06 (3.50)	10.29 (1.57)	08:58.54
	Canadian	10.56 (5.19)	9.00 (3.24)	10:40.48
Rater	Rater 1	15.15 (4.70)	9.18 (2.56)	10:52.03
	Rater 2	13.94 (5.07)	9.96 (2.27)	11:34.39
	Rater 3	14.01 (4.89)	8.67 (2.76)	09:57.00

Table 2. Mean scores for the spatial ability paper folding test (higher is better), the NASA TLX (lower is better), and the overall time to finish the task (mm:ss.ms). Std. dev in brackets.

RESULTS AND DISCUSSION

The study presented in this paper has two goals. Firstly, to assess the usefulness and correctness of the ATB framework. Secondly, to introduce epistemic actions as a comprehensive new performance metric for systems incorporating tangible interaction. To facilitate this discussion, the results of the study are divided into four sections: *spatial ability*; *framework reliability*; *framework validity*; and, *framework predictive power*. Workload test results are described when relevant throughout this section.

Spatial Ability

The paper folding test [4] was performed to provide meaningful data on participant's inherent spatial ability, as it could have a significant impact in how they perform during a jigsaw puzzle task (and how they might ultimately rely on epistemic actions). As expected, participants that were able to finish the main task in less than 15 minutes obtained higher scores than those who did not (see Table 2). An independent-samples t-test revealed that this difference was not statistically significant ($p = 0.062$), which may be attributable to the disparity between the two group sizes – of 60 participants, only 13 failed to complete the task. Additionally, each group of twenty participants coded by the three individual raters exhibited statistically similar spatial scores (one-way ANOVA, $p = 0.744$), demonstrating the equivalence of these groupings. These results will be further discussed in the following sections.

Framework Reliability

The first step when introducing a new measurement instrument, such as a video-coding framework, is to establish its reliability. To do this, three independent raters applied the video-coding framework proposed in this paper to the videos of the 60 participants performing the main puzzle task. Each rater coded 20 videos: 10 male and 10 female participants, with either six or seven of these from each of the three cultural backgrounds sampled. Events were classified with a timestamp, Antle's classification (DP, IP or EXP), and one or more types of epistemic action (if epistemic activity was observed). Additionally, two of these raters acted as second-coders for eight of the 60 videos, providing a selection of double-coded content. These two raters obtained a substantial agreement in terms of Antle et al.'s three categories of action (85.6%), and in the main three categories of the ATB framework – artifact, tool, and body (75.7%). These high level results illustrate how the ATB framework can inform how future tangible systems are designed and implemented. Basically, these results demonstrate that the framework enables researchers to reliably identify the most common source of epistemic activity during interaction with a system, be it through artifact, tool or body. Understanding where the focus of epistemic activity lies will help direct research and design efforts in the most appropriate directions. For instance, if users depend on task artifacts (as we observed in our task, see e.g., Figure 2), attention can be directed to improving these. Alternatively, if users look at tools to alleviate cognitive burdens, these should serve as inspiration for new artifacts or systems that effortlessly accommodate tool use. Finally, if a task is most suited for bodily actions, designers can focus on how best to sense and enable these.

Beyond this point, the strength of the ATB framework is in the granularity of the epistemic actions recorded. In regards to its 20 different types of action, coders reached a moderate level of agreement of 60.9%. A confusion matrix (not shown) revealed that 87.86% of misclassified actions belonged to two groups: actions in which one rater observes

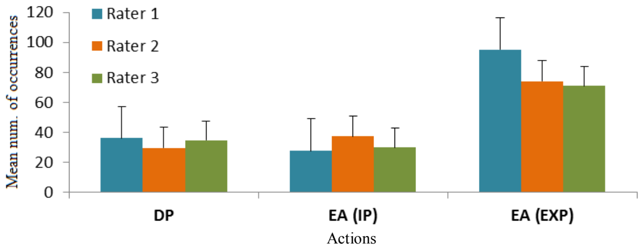

Figure 1. Mean number of pragmatic (DP) and epistemic actions (EA, grouped by actions classified as IP or EXP). Overall data grouped by rater, standard deviation in bars.

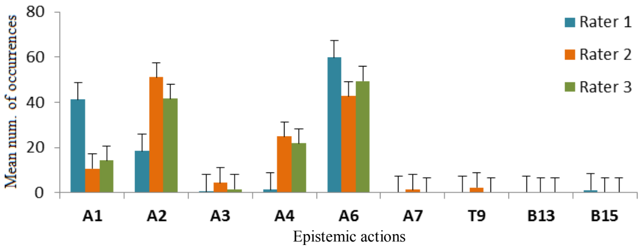

Figure 2. Mean number of individual types of epistemic action, with A2 grouping actions classified from 2.1 to 2.5. Data clustered by rater, standard deviation in bars.

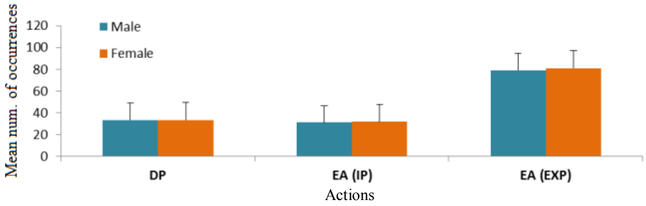

Figure 3. Mean number of pragmatic (DP) and epistemic actions (EA, grouped by actions classified as IP or EXP). Overall data grouped by gender, standard deviation in bars.

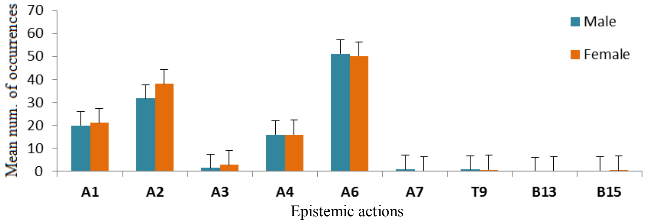

Figure 4. Mean number of individual types of epistemic action, with A2 grouping actions classified from 2.1 to 2.5. Data clustered by gender, standard deviation in bars.

some epistemic activity, but the other does not (61.15% of all misclassifications) and 'vertical' misclassifications, where both raters use nested epistemic actions to classify a particular event (e.g. between IP: A2 and IP: A2.2, 26.7% of all misclassifications). The most common of these misclassifications occurred with action A1, under which all epistemic actions performed with an artifact fit.

Given the limited number of videos double-coded by two raters, additional statistical tests were performed on data from all the 60 sessions to further explore and qualify agreement levels between the three individual raters (subsequently termed *R1*, *R2*, and *R3*) in terms of mean categorical response rates across all rated data. This data is summarized in Figures 1 and 2. A one-way ANOVA was used to study the differences between these datasets – Welsch's F and Games-Howell post-hoc tests were used when the assumption of homogeneity was violated. All three raters reported similar mean numbers of epistemic actions coded as **IP** ($p = 0.090$), and R2 and R3 reported similar mean numbers of actions coded as **EXP** ($p = 0.952$), **A1** ($p = 0.343$), **A2** ($p = 0.573$), **A4** ($p = 0.699$), and **A6** ($p = 0.448$) – the four most common epistemic actions coded by these raters. While these four epistemic actions were also the most commonly reported by R1, most of their occurrence rates significantly varied from the other two raters ($p < 0.015$). Close examination of this pattern suggests it can be explained by the confusion matrix described earlier – R1 tended to classify events with the broad A1 action (Figure 2).

Taken together, these are promising results that vouch for the reliability of the ATB framework as an instrument to record epistemic work. Substantial to moderate agreement levels were attained on different levels of the framework and examination of the raw data shows clear parallels between rater performance. Observed misclassifications fall in a limited number of acceptable types. We believe this data effective illustrates the reliability of the framework.

Framework Validity

In this section we assess the *external validity* of the ATB framework – the generalizability of the framework to a broad participant group. Two different methods are used to achieve this: (1) contrasting the obtained results with current theory on gender differences in spatial ability; and (2), comparing the results of participants from different cultural backgrounds.

There is a long tradition of studying of cognitive differences between the genders, with predictably conflicting and controversial results. While men are often regarded as having higher spatial ability, some studies suggest gender differences are small [e.g. 11] and, indeed, diminishing [28]. Our own results show that male participants obtained higher spatial scores than female participants ($p = 0.029$), but both reported a similar perceived workload when performing the task ($p = 0.797$) and finished with statistically similar mean times ($p = 0.893$) – see Table 2. We argue that this can be explained by examining the results obtained with the ATB framework. These show no statistically significant differences between the mean number of actions performed between the genders (see Figure 3 and 4): **DP** ($p = 0.908$), **IP** ($p = 0.728$), and **EXP** ($p = 0.840$); **A1** ($p = 0.801$), **A2** ($p = 0.244$), **A4** ($p = 0.985$), and **A6** ($p = 0.842$). We suggest that by performing the same number of epistemic actions as their male counterparts, female participants were able to make up for any differences in spatial ability, and indeed that adopting appropriate epistemic behaviors may be more important in this task than high spatial ability.

Furthermore, the experimental data also supports the framework's ability to generate coherent results across broad and varied participant groups. Specifically, the spatial tests recorded a lower spatial score from Canadian participants when compared to both Korean and Dutch participants (one-way ANOVA, Games-Howell post-hoc tests: $p < 0.001$ and $p = 0.003$, respectively). As with the gender groups, however, Canadian participants reported similar workload levels as participants from the other two cultural backgrounds (one-way ANOVA, $p = 0.055$) – see Table 2. We again argue this is attributable to the similar mean number of epistemic actions performed by participants from each cultural group: **A1** ($p = 0.145$), **A2** ($p = 0.064$), **A4** ($p = 0.561$), and **A6** ($p = 0.329$). More so, Canadian participants completed the main task in an average of 10 minutes and 40 seconds, two minutes and four seconds quicker than Korean participants (independent samples t-test, $p = 0.037$). The reason for this result may be in the first minute of the task [2, 19], where Canadian participants performed almost twice the number of epistemic actions than Korean participants. This idea will be explored in more detail in the next section.

Framework Predictive Power

This section explores whether the ATB framework records data that meaningfully relates to other performance metrics (extending the discussion of validity), and whether or not it offers novel explanatory insights into participants' epistemic work. It does this by contrasting the results of participants who successfully finished the main puzzle task with those who did not. Specifically, informed by prior suggestions that early performance of epistemic actions is important in successful performance of spatial problems such as puzzle tasks [e.g. 2, 19], we firstly examined the impact of the rate of epistemic action performance in the first moments of interaction (see Figure 7 and 8). A linear regression showed that the aggregate number of epistemic actions performed in the first minute of the task were statistically significant in predicting participant's required time to finish the task (adjusted $R^2 = 0.177$, $F(1, 58) = 12.501$, $p = 0.001$). These results reinforce current suggestions [e.g. 2, 19] that successful use of epistemic

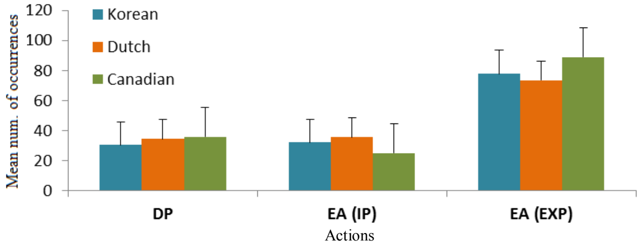

Figure 5. Mean number of pragmatic (DP) and epistemic actions (EA, grouped by actions classified as IP or EXP). Overall data grouped by country, standard deviation in bars.

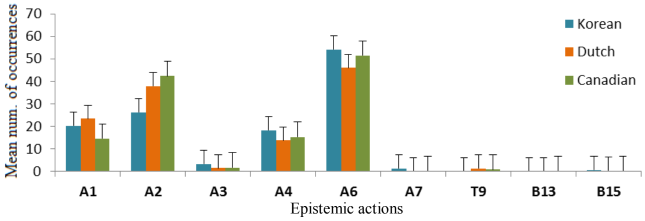

Figure 6. Mean number of individual types of epistemic action (A2 groups actions from 2.1 to 2.5). Data clustered by country, standard deviation in bars.

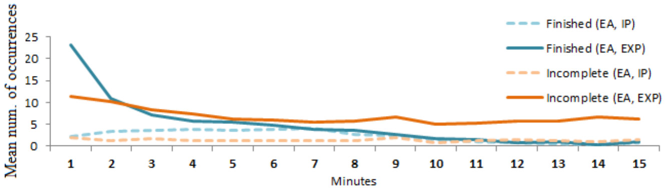

Figure 7. Running mean number of epistemic actions (EA, grouped by actions classified as IP or EXP). Overall data grouped by puzzle completion.

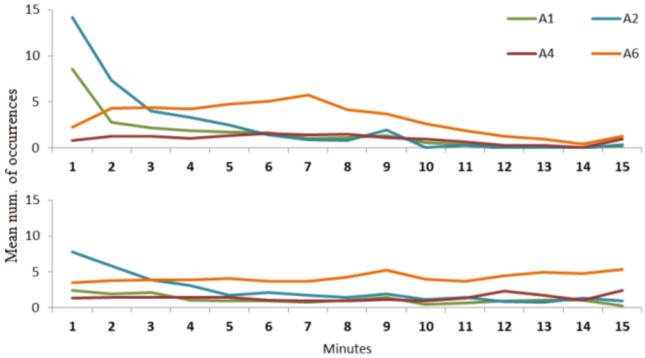

Figure 8. Running mean number of four types of epistemic action (A2 groups actions from 2.1 to 2.5). Data clustered by puzzle completion: finished (top) and incomplete (bottom).

actions relates, not simply to frequency of activity, but in knowing when specifically it is worth performing them.

We then examined this data at the level of individual epistemic action as performed over the entire experimental task. Figure 8 shows this data for the most common epistemic actions performed (Table 1: A1, A2, A4, and A6, accounting for 96.48% of all actions). A multiple regression revealed that the ocurrence of these actions in the first minute of the task was a significant predictor of the time it took participants to complete the puzzle (adjusted $R^2 = 0.232$, $F(4, 55) = 4.158$, $p = 0.005$). Interestingly, not all of these variables added statistically significantly to the prediction (see Table 3). As such, we suggest the ATB framework was successful in identifying which epistemic actions are more relevant and helpful for the user in the context of the early stages of a puzzle task (A1 and A2), which were not particularly helpful (A6), and which had (non-significantly) detrimental effects on user performance (A4). These findings provide evidence that ATB framework is a useful tool that can highlight what kinds of epistemic work are suitable for what kinds of problem and, we argue, this kind of knowledge is valuable for the both the design and assessment of tangible systems.

Variable	*B*	*p*	*Beta*
Intercept	12.663	-	-
A1	-0.077	0.008	-0.378
A2	-0.104	0.001	-0.470
A4	0.139	0.645	0.058
A6	-0.101	0.393	-0.110

Table 3. Regression coefficients and *p* values of a multiple regression that successfully predicts the time it takes participants to finish the main puzzle by measuring how many EA (A1, A2, A4, and A6) occur in the first minute of the task.

FUTURE WORK AND CONCLUSIONS

This paper provides the basis for studies capturing granular data about epistemic activity. It does so by presenting and validating a novel framework that enables the measurement and detailed categorization of epistemic actions. Not only did the presented study test methodological aspects of the framework, it also took concrete steps towards developing evaluation metrics specifically targeted towards user experience in tangible systems. Limitations of this work include that, while it presents a study that looks at a substantial and diverse group of participants, it focuses on a single spatial task. While the framework was informed by epistemic actions collected from various fields, and thus should be applicable to most problems with physical properties, the most pressing future work lies on applying the framework to additional problem-solving tasks. This work will allow us to reinforce, extend and generalize the findings currently presented. These include understanding the importance of actions A1 (*Manipulation of an artifact*) and A2 (*Spatial arrangement of artifacts in relation to one another, the task environment, or the users*) in other spatial problems and developing knowledge about which epistemic actions are relevant for non-spatial, problem-solving tasks. These understandings will ultimately lead to design guidelines specific to tangible interaction, allowing for the development of new interactive systems that support not only goal-directed, pragmatic actions, but epistemic strategies that enable users to apply natural, real-world knowledge to interaction with digital information.

ACKNOWLEDGEMENTS

This research was supported by: the Portuguese Foundation of Science and Technology (FCT, SFRH/BD70203/2010); the SSHRC and GRAND NCE (Canada); and the Basic Science Research Program through the National Research Foundation of Korea (NRF), funded by the Ministry of Science, ICT and Future Planning (2014R1A1A1002223).

REFERENCES

1. Antle, A. N. Exploring how children use their hands to think: an embodied interactional analysis. *Behaviour & Information Technology* (2012).
2. Antle, A. N. and Wang, S. Comparing motor-cognitive strategies for spatial problem solving with tangible and multi-touch interfaces. In *Procs of TEI'13*, ACM, 65-72.
3. Clark, A. *Being There: Putting Brain, Body, and World Together Again*. MIT Press, 1998.
4. Ekstrom, R. B., French, J. W., Harman, H. H. and Dermen, D. *Manual for Kit of Factor-Referenced Cognitive Tests*. Educational Testing Service, 1976.
5. Esteves, A., Bakker, S., Antle, A. N., May, A., Warren, J. and Oakley, I. Classifying physical strategies in tangible tasks: a video-coding framework for epistemic actions. In *Extended Abs. of CHI'14*, ACM, 1843-1848.
6. Esteves, A., Hoven, E. v. d. and Oakley, I. Physical games or digital games?: comparing support for mental projection in tangible and virtual representations of a problem-solving task. In *Procs of TEI'13*, ACM (2013).
7. Hart, S. G. and Staveland, L. E. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In A. H. Peter and M. Najmedin (eds.) *Advances in Psychology*, *52*, (1988), 139-183.
8. Horn, M. S. and Jacob, R. J. K. Tangible programming in the classroom with tern. In *Extended Abs. of CHI'07*, ACM (2007), 1965-1970.
9. Hornbæk, K. and Law, E. L.-C. Meta-analysis of correlations among usability measures. In *Procs of CHI'07*, ACM (2007), 617-626.
10. Hornecker, E. and Buur, J. Getting a grip on tangible interaction: a framework on physical space and social interaction. In *Procs of CHI'06*, ACM (2006), 437-446.
11. Hyde, J. S. How large are cognitive gender differences? A meta-analysis using !w^2 and d. *American Psychologist*, *36*, 8 (1981), 892-901.
12. Jacob, R. J. K., Girouard, A., Hirshfield, L. M., Horn, M. S., Shaer, O., Solovey, E. T. and Zigelbaum, J. Reality-based interaction: a framework for post-WIMP interfaces. In *Procs of CHI'08*, ACM (2008), 201-210.
13. Jacob, R. J. K., Ishii, H., Pangaro, G. and Patten, J. A tangible interface for organizing information using a grid. In *Procs of CHI'02*, ACM (2002), 339-346.
14. Jordà, S., Geiger, G., Alonso, M. and Kaltenbrunner, M. The reacTable: exploring the synergy between live music performance and tabletop tangible interfaces. In *Procs of TEI'07*, ACM (2007), 139-146.
15. Kirsh, D. Adapting the Environment Instead of Oneself. *Adaptive Behavior*, *4*, 3/4 (1996), 415-452.
16. Kirsh, D. Complementary strategies: Why we use our hands when we think. In *Procs of CogSci'85*, 212-217.
17. Kirsh, D. Distributed Cognition, Coordination and Environment Design. In *Procs of EuroCog'99*, 1-11.
18. Kirsh, D. The intelligent use of space. *Artificial Intelligence*, *73*, 1-2 (1995), 31-68.
19. Kirsh, D. Interactivity and multimedia interfaces. *Instructional Science*, *25*, 2 (1997), 79-96.
20. Kirsh, D. and Maglio, P. On Distinguishing Epistemic from Pragmatic Actions. *Cognitive Science*, *18*, 4 (1994), 513–549.
21. Klemmer, S. R., Hartmann, B. and Takayama, L. How bodies matter: five themes for interaction design. In *Procs of DIS'06*, ACM (2006), 140-149.
22. MacKay, W. E. Is paper safer? The role of paper flight strips in air traffic control. *ACM Trans. Comput.-Hum. Interact.*, *6*, 4 (1999), 311-340.
23. Merrill, D., Kalanithi, J. and Maes, P. Siftables: towards sensor network user interfaces. In *Procs of TEI'07*, ACM (2007), 75-78.
24. Shaer, O. and Hornecker, E. Tangible User Interfaces: Past, Present, and Future Directions. *Foundations and Trends in Human-Computer Interaction*, *3*, 1-137.
25. Shapiro, L. *Embodied Cognition (New Problems of Philosophy)*. Routledge, 2011.
26. Tarjan, R. E. Amortized Computational Complexity. *SIAM. J. on Algebraic and Discrete Methods*, *6*, 2 (1985), 306-318.
27. Underkoffler, J. and Ishii, H. Urp: a luminous-tangible workbench for urban planning and design. In *Procs of CHI'99*, ACM (1999), 386-393.
28. Voyer, D., Voyer, S. and Bryden, P. M. Magnitude of sex differences in spatial abilities: A meta-analysis and consideration of critical variables. *Psychological Bulletin*, *117*, 2 (1995), 250-270.
29. Wilson, M. Six views of embodied cognition. *Psychonomic Bulletin & Review*, *9*, 4 (2002).
30. Zaman, B., Abeele, V. V., Markopoulos, P. and Marshall, P. Editorial: the evolving field of tangible interaction for children: the challenge of empirical validation. *Personal Ubi. Comput.*, *16*, 4, 367-378.