Sense of Agency and User Experience: Is There a Link?

Bergström, Joanna; Knibbe, Jarrod; Pohl, Henning; Hornbæk, Kasper

Published in:
ACM Transactions on Computer-Human Interaction

DOI (link to publication from Publisher):
10.1145/3490493

Publication date:
2022

Document Version
Publisher's PDF, also known as Version of record

Sense of Agency and User Experience: Is There a Link?

JOANNA BERGSTRÖM, University of Copenhagen
JARROD KNIBBE, University of Melbourne
HENNING POHL, Aalborg University
KASPER HORNBÆK, University of Copenhagen

Sense of control is increasingly used as a measure of quality in human-computer interaction. Control has been investigated mainly at a high level, using subjective questionnaire data, but also at a low level, using objective data on participants' sense of agency. However, it remains unclear how differences in higher level, experienced control reflect lower level sense of control. We study that link in two experiments. In the first one we measure the low-level sense of agency with button, touchpad, and on-skin input. The results show a higher sense of agency with on-skin input. In the second experiment, participants played a simple game controlled with the same three inputs. We find that on-skin input results in both increased sense and experience of control compared to touchpad input. However, the corresponding difference is not found between on-skin and button input, whereas the button performed better in the experiment task. These results suggest that other factors of user experience spill over to the experienced control at rates that overcome differences in the sense of control. We discuss the implications for using subjective measures about the sense of control in evaluating qualities of interaction.

CCS Concepts: • **Human-centered computing** → *User studies*; *Laboratory experiments*; *HCI theory, concepts and models*; *Empirical studies in HCI;*

Additional Key Words and Phrases: User experience, agency, on-skin input

ACM Reference format:
Joanna Bergström, Jarrod Knibbe, Henning Pohl, and Kasper Hornbæk. 2022. Sense of Agency and User Experience: Is There a Link?. *ACM Trans. Comput.-Hum. Interact.* 29, 4, Article 28 (March 2022), 22 pages.
https://doi.org/10.1145/3490493

1 INTRODUCTION

Sense of control is considered an important quality of human-computer interaction (HCI). For example, the classic user interface design principles, such as those of Shneiderman [45] and Nielsen [38], include a sense of control as one of the factors influencing usability and user experience. Over the past decade, then, HCI research has started adopting both objective and subjective measures

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement 648785).
Authors' addresses: J. Bergström and K. Hornbæk, Department of Computer Science, University of Copenhagen, Universitetsparken 5, 2100 Copenhagen, Denmark; emails: joanna@di.ku.dk, kash@di.ku.dk; J. Knibbe, School of Computing and Information Systems, University of Melbourne, 700 Swanston Street, Carlton, VIC 3053, Australia; email: jarrod.knibbe@unimelb.edu.au; H. Pohl, Aalborg University, Selma Lagerløfs Vej 300, 9220 Aalborg, Denmark; email: henning@cs.aau.dk.

1073-0516/2022/03-ART28 $15.00
https://doi.org/10.1145/3490493

of the sense of control in evaluating user interfaces and in assessing the quality of interaction. However, it is unclear whether the objective and subjective measures relate to control in the same sense.

Measuring a user's sense of control objectively in HCI is hard. Cognitive neuroscience offers objective paradigms [e.g., 17, 36, 37] that measure the immediate and pre-reflective experience of control, that is, the sense of agency. Since Coyle et al. [7] introduced the field of HCI to these measures, we have seen that sense of agency is greater when tapping the skin than when tapping a button [2, 7], that mid-air interactions and haptics increase our sense of agency [6], and that increased system automation and speech interfaces decrease our sense of agency [7]. However, a paradigm for objectively measuring SoA (called intentional binding) is hard to integrate with HCI experiments for two reasons. First, the task relies on low-level inputs, such as individual, binary button clicks, from which the participants estimate the time from the action to the outcome. As a result of this, the task cannot be incorporated into real-world interactive tasks relevant for HCI [e.g., 24] such as tasks with a sequence of button presses (e.g., typing), or other actions (e.g., pointing). Second, the task is demanding for participants, because it regularly requires over 20 minutes per individual condition [6], during which the participant must pay close attention and perform many repetitions of millisecond level time estimations from the button clicks to the position of the arms of an on-screen clock. Thereby, already performing the tasks for the measure takes a significant portion of the experiment duration, and makes adding another interactive task and other measures relevant for HCI on top challenging.

Subjective measures provide an attractive alternative to objective measures, as they are easy to use after any interactive task. The subjective sense of control can be measured with questionnaires [e.g., 18, 37]. These questionnaires were first adapted by psychologists to measure the embodiment of avatars in virtual reality [16, 32] combined with a questionnaire of body ownership [3]. Since then, they have been adopted more widely in HCI to evaluate the quality of interaction in virtual reality [e.g., 14, 15, 19, 31, 39, 41, 48]. However, it is unclear what these subjective measures show about the sense control in HCI. First, how do they relate to the objective measures? If the subjective measures reflect control in the same sense as the objective measures, then the objectively measured sense of agency should be reflected in the subjective experience of control. Second, how do they relate to other measures of user experience? If the sense of control is an important quality in interaction, it should be reflected in the overall measures of user experience.

To understand these links between the measures better, we study in two experiments whether interactions with known differences in objectively measured sense of control also differ (a) in experienced control, and (b) in other factors of user experience. We depart from our prior work on the sense of agency - the pre-reflective sense of control - with a button, a touchpad, and on-skin input [2]. Experiment 1 introduces these three input styles and their differences in the sense of agency. The results show that the sense of agency is significantly higher with on-skin input compared to either the touchpad or button input. In Experiment 2, we measure the user experience in interacting with the same button, touchpad, and on-skin devices as in Experiment 1. To do this, we use a simple game that introduces interaction in which the experiences can be measured, and does so with the same input style as in Experiment 1: binary input by tapping. We measure experience of control, physical effort, and pragmatic and hedonic quality by using AttrakDiff [22], and NASA-TLX questionnaires [20], and a set of questions related to control and ownership derived from the literature measuring subjective experience of control. We find that on-skin input results in an increased sense of control both with objective and subjective measures compared to touchpad input. However, the corresponding difference is not found between on-skin and button input. These results suggest that other factors of user experience spill over to the sense of control. We discuss the implications of this for evaluating HCI with the measures about the sense of control.

2 RELATED WORK

Control is an important design factor in HCI. Earlier design guidelines, such as those of Shneiderman [45] and Nielsen [38], emphasized allowing users to be free initiators and in full control of their actions with user interfaces. In contrast, current interaction techniques may directly influence control, such as to improve target acquisition by sharing the control of one's virtual body [19] or improving reaction time with electrical muscle stimulation [25]. As the user interface design influences the user's control over their actions, the user's sense of control has also been considered important to measure. Exploring sense of control has indeed gained popularity in HCI [e.g., 2, 5–7, 29], under the assumption that an increased sense of control is a desirable user interface quality and design goal.

The sense of agency refers to a users' feeling of *"I did that"* [12] or of *being the author or controller of one's actions* [4, 9]. Importantly, a distinction is drawn between two levels of agency [46]; a low-level, implicit sense of control, and a higher-level, explicit experience of control. However, we are yet to fully understand whether there is a correlation between the objective measures on the sense of agency and the subjective measures on user experiences. It may be that any implicit sense of agency is not manifested in any explicit experiential quality, whether about the experience of control or other factors of user experience. Furthering our understanding here will allow the HCI community to reflect on the utility of the sense of control as both a desirable quality of interaction and the usefulness of its measures. Next, we provide a background on measuring the sense of control in neuropsychology and in HCI, as well as on the possible links between different measures for it.

2.1 Objective Measures of Control

There are two popular techniques for measuring the implicit sense of control: temporal binding and sensory attenuation [10]. Sensory attenuation looks to identify the reduction in the intensity of self-initiated action effects [10], but has not gained much traction in HCI. Therefore, we focus here on the temporal binding methodologies.

The temporal binding paradigms build on the finding that voluntary actions and outcomes are perceived to occur closer together in time [17, 28]. There are two prominent methods to measure temporal binding: intentional binding and interval estimation [7]. With intentional binding, participants are asked to report either the timings of their actions (e.g., a button press) or their associated outcomes (e.g., beep sounds) across multiple repetitions and conditions. A popular version of this task involves showing participants a Libet Clock to assist with timing cues [17, 28]. Recently, both haptic and auditory stimuli have also been shown to provide suitable cues [6]. While intentional binding offers the most accurate results for temporal binding, this method requires the participants' focused attention and many repetitions, with even simple one-to-one input comparisons, easily summing to an hour of study time [e.g., 7]. The demands of intentional binding studies, therefore, leave little opportunity for additional data collection or incorporation into anything but the experimental task [7].

Through the intentional binding method, research in HCI has shown that speech interfaces can have reducing effects for sense of agency [29], while mid-air [5], haptic, and skin-based interfaces provide a heightened sense of agency compared to a button or touchpad [2, 7].

An alternate approach to measuring temporal binding is using interval estimation [12]. Here, participants are asked to guess an interval duration (in milliseconds) between input and output events [7]. This does not demand visual attention from the participant and, as such, can be used in a wider range of experimental tasks. However, it is generally considered less robust than intentional binding, and perhaps because of that has not gained much popularity in HCI.

2.2 Subjective Measures of Control

Explicit measures typically rely on asking participants how much agency they feel, using rating scales to indicate ownership (*"I did X"*) or causation of an event (*"I caused X"*) [10, 36]. For example, in Ebert and Wegner's studies, consistency between actions and consequences was manipulated during a push/pull task using a joystick as the input device. The authors used one question rated on a 7-point scale *"Indicate how much it felt like moving the joystick caused the object on the computer screen to move"* as a measure of explicit sense of agency [11].

Such control judgments have also been shown to relate to other measures of explicit sense of agency. For instance, Barlas and Kopp manipulated action and outcome congruencies and measured the explicit sense of agency with one item asking participants to indicate the degree of control they have felt over the changes on the screen [1]. Longo and Haggard [30] also developed items to measure senses of ownership and agency over a virtual hand. One example item for the sense of agency was *"It felt like I was in control of the hand I was looking at"*.

2.3 Correlations Between the Measures

The correlations between both temporal binding and agency, and between implicit and explicit measures of agency, are still broadly debated by the wider research community (see [10] for a discussion). For example, cognitive neuroscience has questioned the efficacy of the techniques used to measure sense of agency [e.g., 40]. Research has shown that intentional binding, the technique popular for measuring our sense of agency [17], can be modulated independently of whether the person is the author of action [35]. For example, Poonian et al. [40] and Moore et al. [35] demonstrated intentional binding for both self-created and observed actions. Other research has failed to find any correlation between implicit and explicit measures of control, both with traditional methods [10] and through neurophysiological correlates [26]. For example, Dewey and Knoblich [10] found no correlation between implicit and explicit measures of sense of agency. They also found no correlation between sensory attenuation and temporal binding - both measures of implicit sense of agency. Saito et al. [42] reported no correlation either between the implicit and explicit measures of agency. And Kuhn et al. [26] found different neuro correlates for implicit and explicit measures of agency in an electrophysiology study (N1 components for expected action effects and P3 components correlated with judgments of agency).

As the paradigms for measuring the sense of agency continue to gather interest in HCI, it is apparent that their links to each other and to other factors of user experience are still uncertain. Therefore, while control is an important design consideration for any interface [45], we so far have not been able to reflect upon the measures we use in HCI for assessing the sense of control. In the experiments that follow, we study whether user interfaces with known differences in the implicit sense of agency also differs in experienced control and in other factors of user experience.

3 EXPERIMENT 1

This first experiment is our prior work on the sense of agency with a button, a touchpad, and on-skin input [2]. The purpose of this experiment was to show the differences of these three input styles in the sense of agency with an objective measure.

We replicated an experiment on the sense of agency with on-skin and button input by Coyle et al. [7] with an extension of adding a touchpad as a third input. The motivation behind this was to pursue further insights into whether the difference Coyle et al. found in intentional binding (the objective measure for SoA) relate to using a body versus an external device for input, or to the input method, that is, taps versus presses. Therefore, our main hypothesis for Experiment 1 is that **H1: There is a difference in the sense of agency between on-skin and device inputs**. We

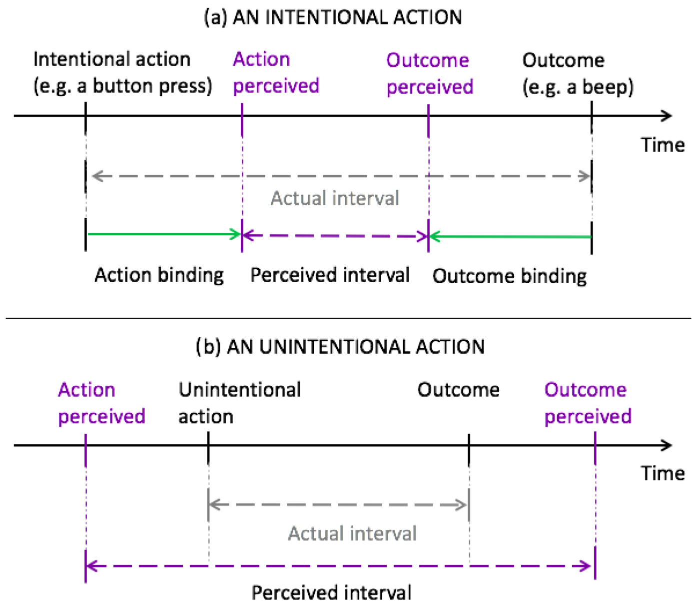

Fig. 1. Time perception in an intentional action (a) and an unintentional action (b). For intentional actions, the perceived interval between action and outcome is shorter than the actual interval, while for unintentional actions the perceived interval is longer than the actual.

hypothesised that the smaller intentional binding with button presses is caused by the intrinsic difference in the input modalities. The rationale behind this is that a button press involves three states (released, touched, and press) whereas a tap on the skin involve two (released and touched). Furthermore, the selection time with a button press can be ambiguous (i.e., whether the selection happens during a button press, at the bottom of it, or when releasing). Therefore, we added the touchpad condition to Coyle's experiment, which allows us to compare a two-state input method similar to tapping on the skin, but on an external device, and thus address whether an increased sense of agency is a result of skin input or touch input in general.

3.1 Libet Clock Method

The method we used for measuring the sense of agency is called the Libet Clock method. Coyle et al. [7] introduced the Libet Clock method to HCI a decade ago. The method was developed by Haggard [17], built it on the work of Libet et al. [28]. The method is used for empirically measuring the sense of agency from changes in perceived time. It builds on the phenomenon that the perceived time interval between intentional actions and their outcomes is shorter than the actual time interval (Figure 1). For example, when a person presses a button that causes a beep after a short time interval, it is likely that the action is perceived to have happened later and the beep earlier than they actually did. The first of these phenomena is called *action binding*, and the latter *outcome binding*. The sum of these factors describes the main dependent variable, called *intentional binding*, which correlates with a sense of agency.

3.2 Task

The Libet Clock (Figure 2) consists of a clock face with an arm that rotates clockwise through a full cycle once every 2,560 ms. The participant sits in front of the screen looking at the clock, and

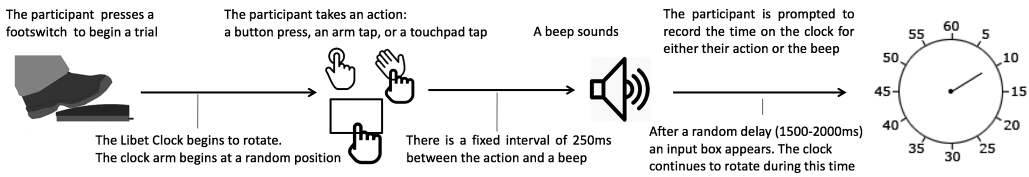

Fig. 2. The protocol for the active conditions in the experiment. Baseline conditions omit either the action or the beep. The Libet Clock on the right as shown on a display in the experiment.

is asked to report where the rotating arm on the clock was pointing when an action (input) was performed, or when an outcome (beep) was observed.

3.3 Design

The experiment followed a within-subjects design with the three input interfaces as levels of the independent variable (Button, Skin, and Touchpad). For each interface four measurements are collected: Action Baseline, Action Active, Outcome Baseline, and Outcome Active. Figure 3 illustrates these four measurements and the calculations for the intentional binding effect. The conditions for measurements are described below in the section on the procedure.

The order of the interfaces and the order of the measurements with each interface were balanced using Latin Squares. The input interfaces were used one at a time to collect all the measurements. We used 30 repetitions to maintain an experiment duration (approximately 80 minutes) which prevents excessive fatigue and loss of attention despite three input conditions.

3.4 Apparatus

The experimental setup is shown in Figure 4. The three interface conditions were set up as follows:

Button: We used an Enter -key of a USB -connected numpad for the Button condition.

Skin: For the Skin condition we used a piezo-electric contact microphone. The piezo sensor with a diameter of 2.7 mm was wrapped on the participant's forearm with a medical bandage. The piezo was located on the posterior side of the arm because that allowed the participants to maintain a natural and comfortable posture, and us to place each input interface to the same location on the table in front of the participant. The piezo sensor was connected to the computer running the experimental software via a high-fidelity soundcard (Focusrite Scarlett 6i6 2nd Gen).

Touchpad: The MacBook's touchpad was used for the Touchpad conditions. The tap could be performed anywhere on the touchpad's active area. The keyboard and the screen of the MacBook were not used during the study and were covered with black cardboard to prevent participant distraction. No audio feedback was used with the touchpad because the beep sound in this Libet Clock experiment is already used for measuring perception of outcome.

A designated area on the table, in front of the participant, was used as the input area for each interface condition (Figure 4). The touchpad and the numpad were stowed away behind the desktop display during other conditions, and the left arm was kept on the side of the table during Touchpad and Button conditions.

Coyle provided the experimental software used in their original Libet Clock study. We used the software for displaying the Libet Clock, running the trials, and for logging the actual times and the reported perceived times. In our experiment the clock size was the same 100 pixels in diameter, and it was displayed on a screen with 1,920 × 1,080 resolution. We used a MacBook Pro (2013) for running the software and collecting the data. The original software could process sensor

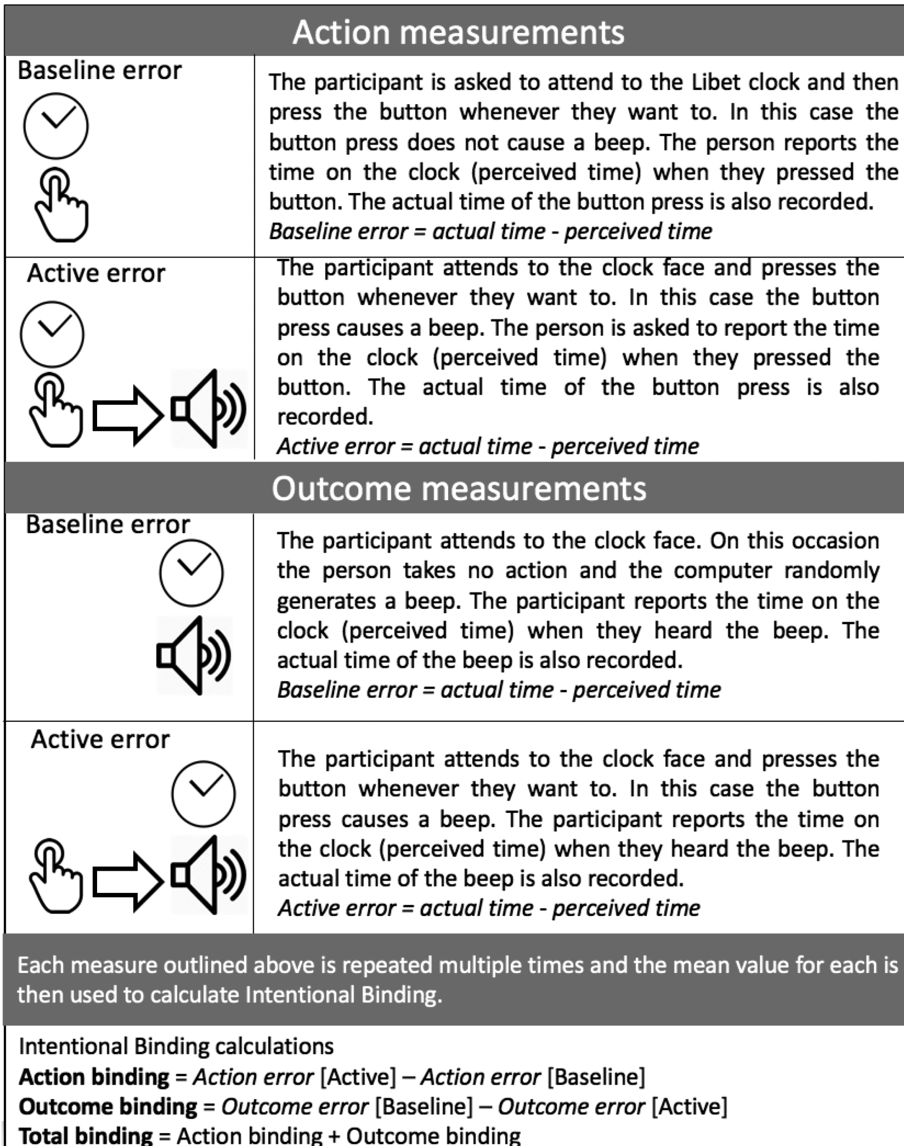

Fig. 3. The measurements and calculations used to assess Intentional Binding for an action/outcome condition, using the Libet Clock method. A fixed action/outcome interval of 250 ms is used for all active trials.

signals and recognise taps on the skin, and register actions from button presses. The software was extended to support the Touchpad condition.

3.5 Procedure

Before starting the Libet Clock trials, the experimenter ensured that all the input interfaces worked reliably. The signal processing features for detecting taps on the skin were adjusted (increasing sensitivity from the default settings) until 10 subsequent taps could be detected.

The participants were first introduced to the tasks of recording the perceived action and outcome times from the Libet Clock. The participants then had a practice round for each of the four

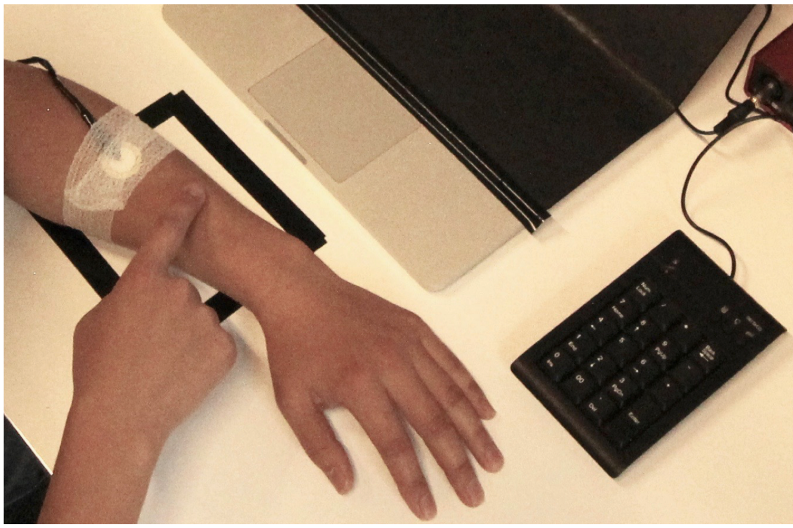

Fig. 4. Setup for the three input conditions: Skin, Touchpad, and Button. The input interface was always brought in front of the subject on a designated area (the black rectangle marked on the table). A separate screen displayed the interface and the Libet Clock. On the image, the Skin interface is shown.

measurement types with each interface. The touchpad was used by tapping (i.e., lifting the finger instead of holding it on the surface and clicking), and the skin was instructed to be used in the same way.

Before beginning the blocks of 30 trials for each interface and each measure, the participants viewed on-screen instructions telling them to record either (a) the perceived action or (b) the perceived outcome time. After reading the instructions, the experimenter revealed the clock face on the screen.

The procedure for the trials is depicted in Figure 2. To start a trial, the participant pressed a foot switch. This caused the clock arm to appear at a random position and start to rotate. In the three trials requiring an action, participants were instructed to wait for the clock arm to rotate through at least one full round before performing the action. The random starting point and waiting for the arm to rotate before performing an action helped to discourage planning of action times.

At a fixed 250 ms interval after the action, a buzzer sound was played as an outcome. In the outcome baseline condition where no action was performed, the sound was played at a random time between 5,000 and 7,500 ms after the foot switch was pressed to start the trial. After the action or the outcome, the clock arm continued to rotate for a random time (between 1,500–2,500 ms) and then disappeared. When the clock arm disappeared, the participant told the experimenter the perceived time of action or outcome, depending on the current condition. The experimenter typed the time on an input dialogue on the screen. The purpose of telling the perceived times to the experimenter instead of simply typing those themselves was to encourage the participants to pay attention to the task. This way, the participant was also able to double-check that there are no errors in the reported times.

3.6 Participants

Data were collected from 24 right-handed participants (10 females, with an average age of 30.95 years). All of them used a computer keyboard and a laptop touchpad daily.

In two cases our system for capturing touches on the skin was not working effectively. However, this behavior was detected already at the beginning of the experiment, and the experiment was

	Action binding	Outcome binding	Total binding
Button	17.68ms (44.75ms)	36.60ms (56.01ms)	54.28ms (67.56ms)
Skin	25.02ms (42.08ms)	70.52ms (57.97ms)	95.53ms (81.95ms)
Touchpad	21.68ms (49.03ms)	19.64ms (60.16ms)	41.32ms (80.61ms)

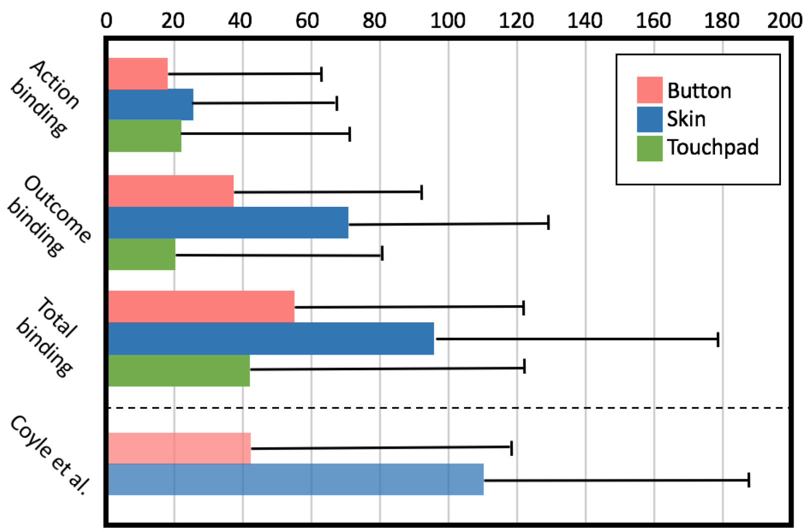

Fig. 5. The upper values and the bars denote the mean action, outcome, and total binding times in ms for each input condition in the present experiment, as well as the total binding times from Coyle et al. Higher binding values, indicate a greater sense of agency. Standard deviations in brackets and in error bars.

discontinued. The next participants were allocated to replace the discontinued ones, resulting in a complete data set of 24 participants.

Three participants, however, misinterpreted experimental instructions in one of the conditions. By examining the action and outcome binding values, we found that two participants had over 200 ms values in action binding, and one participant in outcome binding with one of the input interfaces. Because the outcome was played always at a 250 ms interval after the actions, these high values imply that the participants had incorrectly reported either outcome values (beep time) when the instructions were to report action values (input), or the opposite. Therefore, all data from these three participants were excluded, leaving 21 participants for the analysis.

4 RESULTS OF EXPERIMENT 1

Figure 5 shows the results from the Libet Clock experiment. A one-way repeated measures ANOVA shows a significant difference between the intentional binding values with the three interfaces, $F(2,20) = 8.19$, $p = 0.001$, showing a medium effect size of eta squared $\eta^2 = 0.09$. A posthoc analysis with a Bonferroni corrected paired sample t-test shows this difference is significant between two pairs: Skin and Button ($p = 0.005$), and Skin and Touchpad ($p < 0.001$), with respective effect sizes of Cohen's $d = 0.55$ and $d = 0.67$. No significant difference was found between the mean total binding values with Button and Touchpad.

A Bonferroni corrected paired sample t-test further shows significantly higher outcome binding with Skin input compared to Button ($p = 0.007$) and to Touchpad ($p < 0.001$).

These results replicated the significant difference Coyle and colleagues found between the intentional bindings with the Skin and the Button. Moreover, the average values and the difference are consistent: Input on the skin (mean total binding of 95.53 ms) in this experiment increased intentional binding compared to Button (54.28 ms) and Touchpad (41.32 ms), while Coyle et al. [7] reported the total binding of 109.47 ms for Skin and 42.92 ms for Button. Their effect size was $d = 0.91$. The total binding is further consistent with earlier binding experiments using button input for actions [18, 34].

Our findings on outcome binding values (70.52 ms for Skin and 36.60 ms for Button) are also consistent with those of Coyle and colleagues (79.82 ms for Skin and 36.11 ms for Button). In contrast, the action binding values for Button in our study were on average higher than in Coyle's, thus also leaving the difference to Skin smaller in action binding.

Our effect is medium-sized, and falls within the confidence intervals of the effect size of the study by Coyle et al. (ours: $d = 0.55$, 0.95 CI = [0.19, 0.90]; Coyle et al.: $d = 0.91$, 0.95 CI = [0.37, 1.44]). We obtained the per participant averages of total binding values found in the original study. We used these to calculate the overall estimation of the effect size. For these two studies, the combined effect size is $d = 0.710$.

We also conceptually replicate the finding of Coyle et al. by showing that intentional binding is higher with the skin than when interacting with a touchpad ($d = 0.67$, again a medium effect). This is significant because it rules out the explanation that the difference was incidental to the button used. In particular, it invalidates the concern that taps and presses make a difference to this effect. Thus, our findings support the idea that the difference in the sense of agency is due to using the skin as an input device.

5 EXPERIMENT 2

The purpose of this experiment was to understand if and how interfaces that provide different levels of sense of control with objective measures also differ in subjective measures of the sense of control and other factors of user experience they provide.

We study interfaces that Experiment 1 [2] has shown to differ in the provided implicit sense of agency, namely on-skin, touchpad, and button input. The same interfaces and input methods (binary tapping or pressing) that were used in previous work were also used here. In sense of agency studies using the Libet Clock, the measurement procedure is the task; in contrast, we introduce a real-world task in the form of a game. The game allows us to measure user experience using a variety of questionnaires that focus on technology perception, workload, and overall satisfaction. This has not been possible within Libet Clock experiments.

Our main hypothesis in Experiment 2 is that **H2: Differences in experienced control correlate positively with the sense of agency**. We also expect that interfaces associated with a higher sense of control (e.g., on-skin input) will be experienced more positively (as per other components of user experience) than interfaces associated with a lower sense control (e.g., button input). The rationale behind this hypothesis is that for the measures of experienced control to be relevant in HCI, they should reflect the lower-level sense of control, and have some effect on higher-level overall experiences with user interfaces.

5.1 Participants

We collected data from 42 participants (21 females, with a mean age of 27.10 ± SD of 4.89 years). All of them used a keyboard daily, and 27 of them a touchpad.

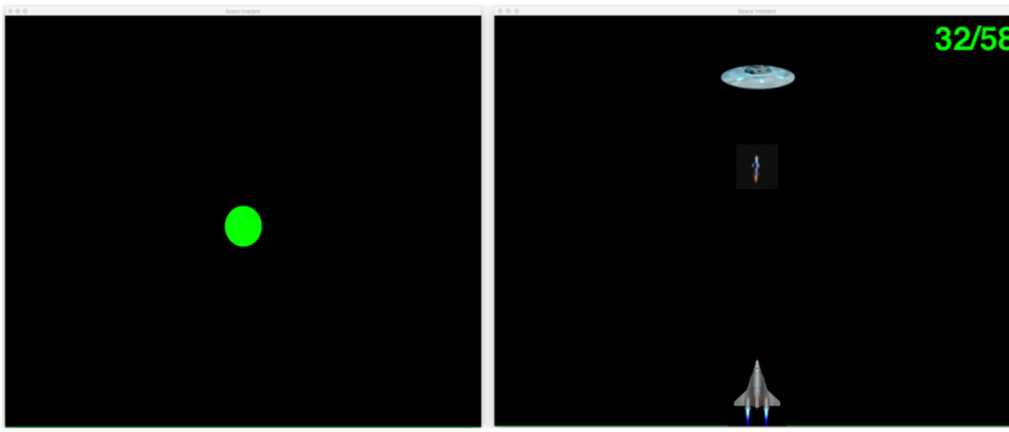

Fig. 6. The screenshot on the left shows the reaction time test to ensure the equivalent technical performance of the input interfaces. The screenshot on the right shows the game used as the experiment task.

We aimed for the statistical power required to detect effects of a size similar to those found in previous studies of the agency. Earlier studies of agency in HCI have typically found sense of agency to be a medium to large effect: Coyle and colleagues [7] found an effect of $d = 0.91$ and we in Experiment 1 [2] found an effect of $d = 0.55$, with a combined effect size of 0.71 [2]. Of course, these effect sizes might not hold for explicit user experience measures as they are likely more variable, and we collect them only once instead of repeated measures as in the earlier Libet Clock studies. But the previous effects provide an initial estimate. According to an a priori power analysis using G*Power [13], we needed a sample size of 37 to find pair-wise differences using t-tests at 0.90 power, given the effect size of $d = 0.55$. To have a sample size above that 37, and to counterbalance for the three input interfaces, we recruited 42 participants.

5.2 Design

The experiment followed a within-subjects design with three input interfaces (Button, Skin, and Touchpad). The order of the interfaces was counterbalanced. After the participants completed the task with one of the interfaces, they filled in a set of questionnaires measuring the explicit user experiences with that interface.

5.3 Task

The experiment task was to play a simple game, using either the button, the skin, or the touchpad for input. In the game, the participants shot at passing targets (Figure 6 on the right).

In the game participants saw a spaceship in a static position at the bottom of a window. A tap or a press launched a bullet from the tip of the spaceship. The task was to hit as many targets as possible. The targets were UFOs that traveled with constant speed on a horizontal path across the window. The UFOs flew into the view from a random position above the tip of the spacecraft. Every other UFO flew in from the left, and every other from the right. The participant had only one bullet available for each target UFO. If a target was hit, it froze for 1,000 ms before the next target appeared, and if it was missed, it continued its travel outside the window before the next target flew in. The interface displayed the hit ratio of the targets.

The game served three purposes in the experiment. First, the experiment needed to incorporate real-world interaction in order to measure user experiences with the input interfaces. The game could provide such interaction instead of the experimental tasks used in measuring implicit sense of agency. Second, the interaction needed to be similar to that in previous experiments of sense of agency: binary selection. The game was therefore controlled by pressing or tapping one of the three input interfaces. This was needed to allow relating user experience to findings on implicit sense of agency. Third, a measure of performance was needed as a control measure for the measures of user experience. We found temporal pointing to provide such a measure for interacting with the interfaces with binary input. However, merely using a temporal pointing task would not have met the first requirement for real-world interaction. We, therefore, designed a game that only takes binary input, provides a performance measure based on temporal pointing, and provides the needed level of interaction to measure user experience.

5.4 Procedure

Participants performed four reaction time tests in addition to the experiment task. The first reaction time test was performed before the experiment started. It served as a practice round for on-skin input, which the participants were not familiar with. Lower performance with the skin condition was a concern, as participants were likely experts with keyboards and touchpads. In that initial reaction time test, the participants practiced arm taps in blocks of 10 trials until one block was performed without missing any taps. The three other reaction time tests were performed before each input condition and consisted of one block of 20 reaction time trials. The purpose of these was to ensure high input accuracy for all input interfaces.

In the reaction time tests, participants were asked to fixate on a blank window displayed on a screen until a green circle appeared in the middle of it (Figure 6 on the left). We asked the participants to react to circle appearances by pressing the button, tapping their arm, or tapping the touchpad as quickly as possible. We registered success for such a trial where a tap or press happened within 700 ms. This value was determined through pilot studies. If the participants needed to attempt a second tap for a missed one, reaction times were always longer than 700 ms. A correctly registered tap with a normal reaction and attention resulted in reaction times faster than 700 ms, even when the participants fixated elsewhere on the screen. The green circle appeared with a random delay of 1,500 ms to 2,000 ms after the previous response (or after starting a block of trials).

The main experiment task—playing the game—was performed with each input interface after these blocks of 20 reaction time trials. The participants were instructed to try to hit as many targets as possible and informed about having only one bullet available for each target. The game always lasted for three minutes for each input interface. This allowed us to ask participants about their perceived game duration. A fixed duration also allowed to minimize interference of performance and duration with the user experience measures.

5.5 Questionnaire on User Experience

User experience was measured with questions 1–10 presented in Table 1. We used AttrakDiff [22] because it focuses on technology perception (rather than on system features or a task). It further allows us to distinguish pragmatic qualities, which should also reflect the task performance with the interfaces, and hedonic qualities, which may reflect experiences specific to interfaces, such as novelty. We used the English version of AttrakDiff [21].

We also asked about the usability of the interface with one two-question UMUX-LITE [27]. A recent study found UMUX-LITE to work well for measuring perceived usability [27].

To measure intentional physical effort, we used NASA-TLX to assess experienced workload in terms of physical demand. NASA-TLX also provides further measures on perceived temporal demand, performance, and effort. We used the standard NASA-TLX, including all the scales, but

Table 1. Our Questionnaire for Measuring Explicit User Experience

Nr	Questionnaire	Type	Question	Scale	Levels
1	**AttrakDiff**	**Pragmatic**	I found the device	Confusing—Structured	7
2	**AttrakDiff**	**Pragmatic**	I found the device	Impractical—Practical	7
3	**AttrakDiff**	**Pragmatic**	I found the device	Complicated—Simple	7
4	**AttrakDiff**	**Pragmatic**	I found the device	Unpredictable—Predictable	7
5	**AttrakDiff**	**Hedonic**	I found the device	Dull—Captivating	7
6	**AttrakDiff**	**Hedonic**	I found the device	Tacky—Stylish	7
7	**AttrakDiff**	**Hedonic**	I found the device	Cheap—Premium	7
8	**AttrakDiff**	**Hedonic**	I found the device	Unimaginative—Creative	7
9	**UMUX-LITE 1**	**Usability**	This system's capabilities meet my requirements	Strongly disagree—Strongly agree	7
10	**UMUX-LITE 2**	**Usability**	This system is easy to use	Strongly disagree—Strongly agree	7
11	**NASA-TLX**	**Mental Demand**	How mentally demanding was the task	Low—High	21
12	**NASA-TLX**	**Physical Demand**	How physically demanding was the task?	Low—High	21
13	**NASA-TLX**	**Temporal Demand**	How hurried or rushed was the pace of the task?	Low—High	21
14	**NASA-TLX**	**Performance**	How successful were you in accomplishing what you were asked to do?	Low—High	21
15	**NASA-TLX**	**Effort**	How hard did you have to work to accomplish your level of performance?	Low—High	21
16	**NASA-TLX**	**Frustration**	How insecure, discouraged, irritated, stressed, and annoyed were you?	Low—High	21
17	**Body Ownership**	**Ownership**	It felt like the device I was using was part of my body	Strongly disagree—Strongly agree	7
18	**Agency 1**	**Agency**	It felt like I was in control of the movements during the task	Strongly disagree—Strongly agree	7
19	**Agency 2**	**Agency**	What is the degree of control you felt	Lowest (1)—Highest (6)	7
20	**Agency 3**	**Agency**	Indicate how much it felt like pressing/tapping the button/touchpad/arm caused the space craft to shoot	Not at all—Very much	7
21	**Time perception**	**Time**	Perception of task duration	Seconds	numerical

The second column shows the name of the questionnaire, the third the quality it addresses, the fourth the actual question, the fifth its scale as expressed on the questionnaire, and the last the levels of the measure.

without weighting. Weighting was not included as the experiment task was a game, and therefore the importance of sub-questions of NASA-TLX were considered irrelevant for participants to evaluate.

The explicit sense of agency was measured with items related to the *experience of being in control*, *consistency between actions and outcomes,* and *exerting intentional physical effort.* These are facets of the explicit sense of agency that have been shown in previous studies as indicators of authorship of one's actions [11, 30, 33]. The feeling of control judgments has reliably been shown to be related to an explicit sense of agency. For instance, Barlas and Kopp manipulated action and outcome congruencies and measured the explicit sense of agency with one item asking participants to indicate the degree of control they have felt over the changes on the screen [1]. We used this item also as an agency measure (Question 19 in Table 1).

Longo and Haggard [30] also developed items to measure senses of ownership and agency over a virtual hand. One example item for sense of agency was *"It felt like I was in control of the hand I was looking at"* rated on a 7-point Likert scale, +3 indicating strong agreement and −3 indicating strong disagreement. We modified their questions to match the present study's input interfaces (Questions 17 and 18 in Table 1).

In Ebert and Wegner's studies, consistency between actions and consequences was manipulated during a push/pull task using a joystick as the input device. The authors used one self-report question *"Indicate how much it felt like moving the joystick caused the object on the computer screen to move"* as a measure of explicit sense of agency, also rated on a 7-point scale anchored at 1, 4, and 7, with "not at all", "somewhat", and "very much", respectively [11]. We also modified the wording of this item to match the input categories we have instead of the joystick (Question 21 in Table 1).

Time perception is also related to both agency and user experience; the implicit agency is measured by perceived time in both Libet Clock and interval estimation paradigms, and experienced task duration reflects task performance in interaction. We use the game with a constant duration

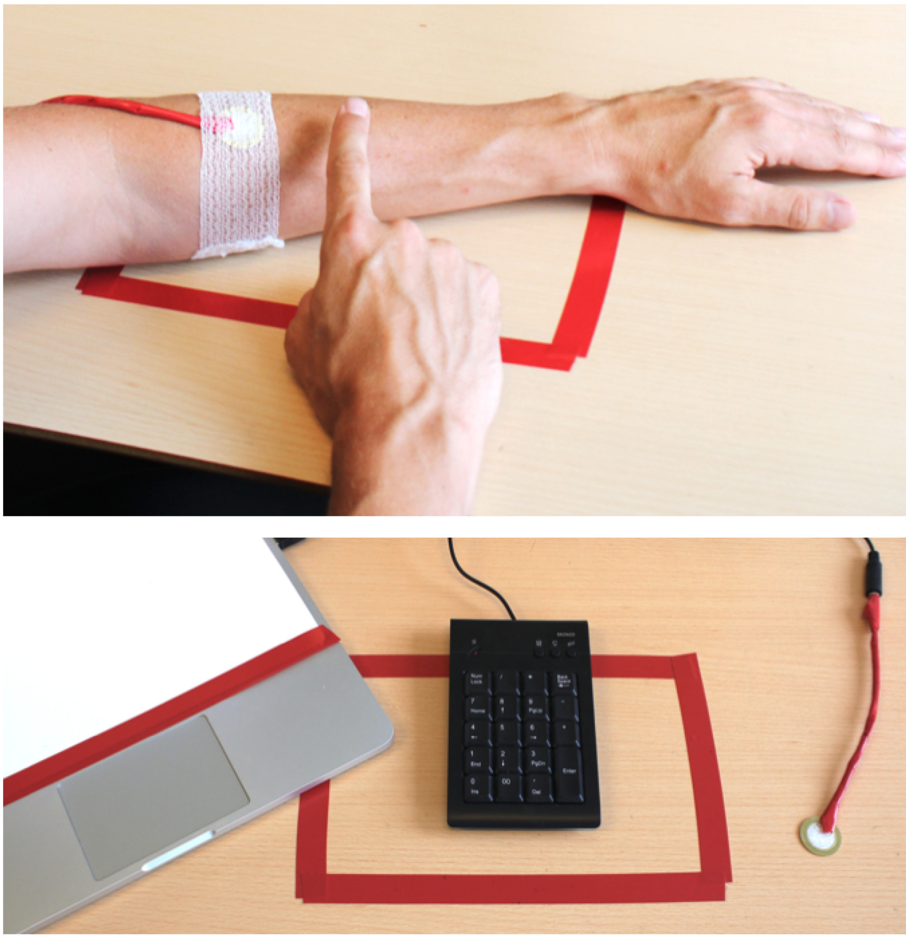

Fig. 7. The study setup. The image above presents a participant tapping their arm for input. The image below shows the three input interfaces: a touchpad, a numpad for button input, and the piezo-electric sensor for on-skin input.

of 3 minutes, and an explicit question on game duration to measure how time is subjectively perceived. This is similar to a measure called Subjective Duration Assessment [8].

5.6 Apparatus

The input interfaces are shown in Figure 7. We used the Enter key of a USB-connected numpad for button input, and a MacBook for touchpad input. The whole area of the touchpad was available for tapping, and the users were instructed to tap, not click it. The screen and keyboard of the MacBook were covered to prevent distraction.

For sensing taps on the arm, we used a piezo-electric contact microphone with a diameter of 27 mm attached to the left arm of the participant with a medical bandage, similar to previous work [2, 7]. It was placed on the non-dominant arm to tap on the skin with the dominant index finger. The piezo sensor was connected to the MacBook Pro (2013) running the experimental software via a high-fidelity audio interface (Focusrite Scarlett 6i6 2nd Gen).

To track taps on the skin, we continuously sampled the microphone at 44.1 kHz frequency and performed peak detection on blocks of 1,000 samples. We defined a peak as a sample that exceeds an absolute signal magnitude threshold as well as an absolute gradient threshold. These thresholds were defined by piloting. Thus, a tap needs to be of sufficient magnitude, but also needs to exhibit a sharp rise in amplitude. The low signal-to-noise ratio of the microphone allowed us to configure the two thresholds for reliable detection with no false positives, and yet to be sensitive enough for light taps. Every tap results in a series of peaks in the signal. We only considered the first peak for a tap and blocked subsequent peaks in a 200 ms window. This further helped sensitivity and was allowed by the experiment task design, as only one bullet, that is, one tap was in use for each target in the game.

We marked an area on the table with tape to ensure that all input was given at the same location, at a constant distance from the display, and at a comfortable distance from the user to rest the arm and the wrist of the inputting hand on the table. The input interfaces were stowed away when not used. The experiment tasks were displayed on a Samsung SyncMaster 244T LCD display with a 1,920 × 1,200 resolution. The table height was adjusted for each participant as needed.

6 RESULTS OF EXPERIMENT 2

We first analyse data from the reaction time tests to ensure a high level of precision in sensing input with each interface. We then report the results on task performance with the game. Finally, we report the user experiences based on questionnaire data.

6.1 Input Accuracy

Reaction time tests were used for ensuring high accuracy rates of input with all interfaces. In reaction time tests, the average error rates (reaction times > 700 ms) were 7.98% for touchpad, 7.86% for on-skin, and 2.26% for button input. A one-way repeated measures ANOVA shows a significant difference between the error rates with the three interfaces, $F(2, 41) = 11.43, p < 0.001$. A posthoc analysis with a Bonferroni corrected paired sample t-test shows the difference in error rates is significant between two pairs: touchpad and button ($p < 0.001$), and skin and button ($p < 0.001$). These results suggest that while taps are registered equally well between skin and touchpad, button input outperforms skin and touchpad in binary input.

Average reaction times (excluding the trials with over 700 ms reaction times) were 424.12 ms for touchpad, 364.17 ms for on-skin, and 340.47 ms for button input. A one-way repeated measures ANOVA shows a significant difference between the reaction times with the three interfaces, $F(2, 41) = 40.33$, $p < 0.001$. A posthoc analysis with a Bonferroni corrected paired sample t-test shows the difference in reaction times is significant between all pairs: touchpad and skin, touchpad and button, and skin and button ($p < 0.001$).

As we found no difference in the accuracy between touchpad and skin, the differing reaction times are likely caused by inherent differences of the input modalities rather than user performance. For example, a finger can be rested on the button to allow fast reaction time, whereas with skin and touchpad input the finger waits in the air at a distance to react, possibly increasing the reaction time.

6.2 Task Performance

We collected hit rates in the game to analyse task performance with each interface because performance can influence user experiences. Even if input accuracies do not differ, the performance in the game may vary due to other factors, including previous experience with an input interface.

In the game, the mean performances as a hit rates were 92.06% (73.86 hits and 6.36 misses) for touchpad, 92.19% (74.64 hits and 6.29 misses) for on-skin, and 95.92% (78.33 hits and 3.33 misses)

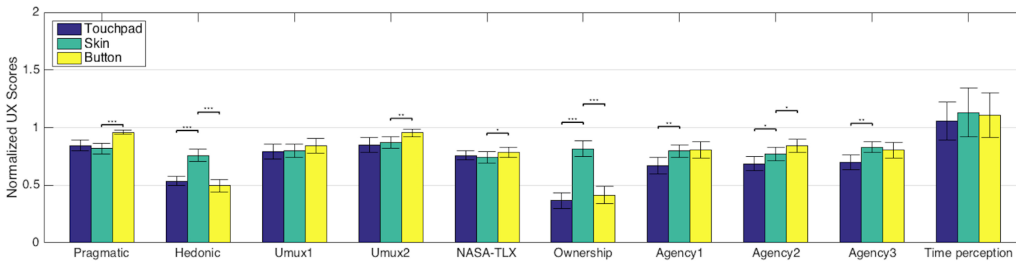

Fig. 8. The results from the User Experience questionnaires for the three input interfaces. The scores are normalized to 0..1 (except for the time perception with an open answer, which is normalized to the proportion of the actual duration of the task). The variables correspond to questions on Table 1 as follows: Pragmatic is an average of Q1–Q4, Hedonic is an average of Q5–Q8. Umux 1 is Q9 and Umux 2 is Q10. NASA-TLX is normalised and averaged for Q11–Q16. Ownership is Q17. Agency1, 2, and 3 are Q18–Q20, respectively. Time perception is Q21. The error bars are 0.95 confidence intervals. The asterisks denote significant differences between pairs in t-test: * corresponds to $p < 0.05$, ** to $p < 0.01$, and *** to $p < 0.001$.

for button input. A one-way ANOVA shows a significant difference between the hit rates with the three input types, $F(2,41) = 6.27, p = 0.003$. A posthoc analysis with a Bonferroni corrected paired sample t-test shows this difference is significant between two input interface pairs: touchpad and button ($p = 0.001$), and skin and button ($p < 0.001$). This result allows us to eliminate the influence of performance on our measures of user experience for analysing differences between the touchpad and the skin. However, the better performance with button input may positively influence user experience with it, in particular when compared with either of the two other interfaces.

6.3 User Experiences

For analysing the user experiences with the questionnaire data, we first normalized all scores. The scores were normalized by dividing each individual measure by the levels of their scale (the scales are presented in Table 1). Except for five scales of the NASA-TLX questions, all scales position positive end at the higher boundary. Therefore, we reversed the scores of NASA-TLX for all questions but the question on Performance, so as to have the positive end of the scale at 1 (after normalizing the scores). Note, that the scores on Time perception are based on open numerical answers, and can thus have values over 1 (indicating that game duration was perceived longer than it actually was).

Figure 8 shows the results from the user experience questionnaire. All data from the questionnaires were analysed with paired t-tests between on-skin and touchpad input, and between on-skin and button input. We are only interested in these comparisons because a sense of agency in earlier studies was found higher for on-skin input compared to the two other input interfaces. Below, we also report the effect size for all significant differences as Cohen's d values.

Pragmatic quality of button input (mean score = 0.960) was found to be significantly higher than that of on-skin input (mean = $0.819, p < 0.001, d = 1.24$). In contrast, no difference was found between on-skin and touchpad input. On-skin input was rated higher in hedonic quality (mean = 0.759) compared to both button (mean = $0.495, p < 0.001, d = 1.51$) and touchpad input (mean = 0.537, $p < 0.001, d = 1.42$).

UMUX-LITE question 1 did not show differences between input interfaces, but button was rated as easier to use (mean = 0.956) in UMUX-LITE question 2 compared to on-skin input (mean = 0.871, $p = 0.002$, $d = 0.63$).

The overall normalized mean scores for NASA-TLX were 0.758 for touchpad, 0.741 for on-skin, and 0.786 for button input. The difference in these scores is significant between on-skin and button input ($p = 0.035, d = 0.30$). Among the individual NASA-TLX questions, differences were found in scores for perceived mental demand and effort, and between both pairs (on-skin and touchpad, and on-skin and button input). Mean scores for mental demand were 0.730 for on-skin, 0.797 for touchpad, and 0.8141 for button input. The difference in mental demand was significant between touchpad and on-skin input ($p = 0.045$), and between on-skin and button input ($p = 0.012$). Mean scores for effort were 0.712 for on-skin, 0.644 for touchpad, and 0.721 for button input. The difference in perceived effort was significant between touchpad and on-skin input ($p = 0.048$) and between on-skin and button input ($p = 0.031$).

The skin was rated significantly higher in a question for ownership of the input interface (mean = 0.816) compared to both touchpad (mean $= 0.367, p < 0.001, d = 2.06$) and button input (mean $= 0.415, p < 0.001, d = 1.72$). In all of the explicit questions on control over the actions and outcomes, on-skin input (means for Agency 1 = 0.816, Agency 2 = 0.796, and Agency 3 = 0.769) was rated higher compared to touchpad (Agency 1: mean $= 0.367, p = 0.003, d = 0.62$; Agency 2: mean $= 0.670, p = 0.028, d = 0.43$; Agency 3: mean $= 0.687, p = 0.002, d = 0.75$). The button was found to provide a higher degree of control (mean = 0.844) than on-skin input (mean = 0.769) in question 2 on agency ($p = 0.029, d = 0.40$), and no significant difference was found between the on-skin and button input in the two other questions on agency.

No differences were found in perception of the duration of the game between the three input interfaces.

In summary, experienced control seems to reflect the difference in the sense of control between the on-skin input and touchpad. However, such a relation was not found between on-skin input and button input. The only measures where on-skin input resulted in more positive user experiences compared to both button and touchpad input were hedonic quality and ownership. Therefore, other factors appear to overcome the differences in the sense of agency when measured subjectively as experiences of agency, also hiding how the experienced control may spill over to other components of user experience.

6.4 Performance and User Experience

Performance in both hit rates and reaction times was higher with button input. Therefore, performance might be one of the factors which overcome how the sense of control is reflected in the higher-level experience of control. Therefore, we analysed the possible effects of performance on user experience measures with partial correlations adjusted for internal variables: hit rate and reaction times. Here we report the correlation coefficients (Pearson's r) and their significance as p-values using a Student's t distribution.

The correlation coefficients and corresponding p -values for partial correlation between the hit rate and each of the user experience measures (adjusted for the reaction times) showed significance for four measures: Pragmatic quality ($r = 0.208, p = 0.022$), UMUX-LITE question 2 ($r = 0.205, p = 0.024$), NASA-TLX ($r = 0.331, p < 0.001$), and Agency 2 ($r = 0.219, p = 0.015$). These correlations are unsurprising as the differences in experiences between the better performing button and on-skin input were found in scores for these same measures.

However, no correlations could be found between reaction times and user experiences. This suggests that such low-level performance in input as the reaction time may not affect subjective experiences. In contrast, the higher-level performance measure of hit rate appears to influence subjective experiences.

7 DISCUSSION

We set out to explore the links between objective and subjective measures of the sense of control, and their relationship to user experience. Specifically, we were interested in whether different levels of sense of agency are also reflected in experienced control and in other components of user experience. We have done so by relying on known differences in sense of agency between input interfaces. Those interfaces were compared experimentally by having participants play a game and produce reports on user experience. Next, we discuss the key findings about this link, address some limitations of our work, and outline future work.

7.1 Links from Agency to User Experience

The studies show two main results across the interfaces, one related to the comparison of skin and touchpad **(i)**, and one related to the comparison of skin and button **(ii)**. In Experiment 1, the skin showed a higher sense of agency in comparison with both the touchpad and the button.

For **(i)**, the results are clear. In Experiment 2, we find an increase in the subjective sense of control with on-skin input over touchpad input, measured by three questionnaire items about control. These findings echo those captured by the objective sense of agency measures in Experiment 1, which show increased binding of 96 ms for skin input, compared to 41 ms for touchpad input. We also find a significant difference in the hedonic quality of on-skin input, as well as in ownership. All of these measures favor the skin.

These findings suggest that the benefits of skin input for our sense of control spill over into the user experience. The three questions on explicit agency in Experiment 2 (questions 18–20 in Table 1) show medium to large effect sizes ($d = 0.62, d = 0.43$, and $d = 0.75$, respectively). These are of similar size to the effects found in implicit agency measures for this comparison in Experiment 1 ($d = 0.55$, 0.95 CI= $[0.19, 0.90]$). The finding on hedonic quality may be related to differences in agency, but could also be a form of novelty effect, in that participants find the interface captivating and fun.

We find it important that the objective data on performance for skin and touchpad were similar; hit rate is less than 1% different and reaction times varied less than 60 ms. This similarity makes the comparisons above very clean.

For **(ii)**, the comparison of skin and button, results are less clear. In Experiment 2, we find a significant difference in one of the three questions about control, where button scores 9.7% percent *higher* than the skin; the other two questions are not different. In Experiment 1, the button scored significantly lower than the skin in the objective measure of control. The result of Experiment 2 is therefore opposite to the hypothesis that the differences found in the objectively measured sense of control are also shown in subjective measures of control.

In the other measures of user experience in Experiment 2, the skin is associated with higher hedonic quality (being more "creative", "stylish", and "captivating") and a higher degree of ownership than the button. In contrast, the button performs better on pragmatic quality, NASA TLX (particularly, mental demand and effort), and one of the UMUX components (about whether the system is "easy to use"). All of these last differences are about classic aspects of usability. Recall, for instance, that pragmatic quality is about being "structured", "practical", and "predictable".

The differences in performance between skin and button are small. Despite our pilots and attempts to make performance similar, the button was 4% more accurate than the skin but had only a 24 ms difference in reaction time. The 4% difference in accuracy means that participants in the skin condition missed 3 UFOs more during the 3-minute game than with the button.

We found correlations between performance in the game and measures of user experience. However, we found no correlation between reaction times and user experience. One explanation of this difference is that those low-level performance differences in reaction times are not perceived by users, while the higher-level performance of hit rates are. In the game, the participants perceived feedback of missing a target, and the hit ratio was also displayed to them. In reaction time tests, however, participants could not perceive success in similar ways as in the hit-or-miss setting of the game. Therefore, performance in reaction time may have been left unnoticed by the participants, unlike their performance in the game, and that perceived performance may have increased their positive user experiences with the button, including that of control.

7.2 Is There a Link, Then?

Across similar performance rates, the implicit sense of control emerges as explicit experiences of control in the responses to our agency questions across skin and touchpad input. However, with differences in performance, these same correlations are not observed. As such, our current interpretation of the data suggest that *implicit sense of control influences explicit sense of control; however, differences in the perceived performance appear to moderate this influence.*

The finding of the moderating effect of performance is consistent with the literature on objective/performance versus subjective/experience components of usability [e.g., 23, 43]. Although this literature generally finds separate components, in all studies there is a substantial correlation between these components. We suspect such a correlation might be influencing our study as well, so that with the higher performance with the button, the differences on the experience of control are diminished below significance. In general, however, the inference process involved in user experience questionnaire answers is complex [47] so it is difficult to propose a more specific explanation.

In summary, where the literature has not found a clear link between intentional binding and explicit judgements of control [10], we find a correlation between objective and subjective measures of control, as well as between the objective measure of control and other components of user experience. However, this correlation may be moderated by performance. These findings, in concert with those of the literature, then, do not fully support a clear argument for either measure of control as a standalone metric of quality in HCI.

7.3 Limitations

Experiment 2 and our interpretation of it have a number of limitations. Most importantly, we assume that input interfaces can be equated with differences in sense of agency. This assumption is based on Experiment 1 and earlier work which has investigated these differences [2, 7], and their findings seems large and robust. However, we do not have actual, per-participant agency measures due to this setup, which would have allowed us to do a more detailed analysis of correlations.

Another issue, which we worked extensively with in piloting the study for Experiment 2, is that controlling performance between conditions is difficult. The differences that do remain in the study, for instance in hit rate, likely shape our findings. They appear in part, however, to be related to inherent characteristics of button and touchpad input (for example, the finger rests on the button, whereas in tapping a touchpad the aiming phase happens without contact and may thus increase reaction times and accuracy of the temporal tap) so in the current setup they are difficult to mitigate.

We used a gaming task in Experiment 2. Games likely come with particular expectations on part of the participants about control. Shneiderman [44], for instance, has suggested that game players compete with the system, while other (application) users prefer a stronger sense of control and of being in charge. Therefore, game playing may change the correlations between experiences of control and other user experiences, compared to interacting with an application. This might have influenced the results.

Finally, in this study, we considered three input styles that all exhibited intentional bindings in their original studies [2, 7]. Our pursuit of correlations with other user experience measures is based on significant differences in those bindings, but bindings existed for all three input styles nevertheless. It could be that the moderating effects of performance that we see are a result of only small (yet significant) differences in binding effect (approximately 40 ms between skin and button, for example). A deeper exploration of scales of binding effect with user experience measures remains an avenue for future work.

8 CONCLUSION

Both objective and subjective measures of control are increasingly applied in HCI to compare the qualities of user interfaces. To date, however, the wider literature has been unable to establish a clear link between the implicit sense of agency and higher-level user experience, calling into question the suitability of the current measures of the sense of control in HCI. We have added to the knowledge of the relationship between implicit and explicit senses of control, but further questions remain. We examined the user experience of three input interfaces with known differences in objective measures of sense of agency: button, skin, and touchpad input. We also examined the complementary effect of task performance on user experience. We found that, across input styles that performed similarly (touchpad and skin input), sense of agency did indeed correlate with known UX measures of control. However, when performance differed (between skin and button input), these correlations could not be found. Our results suggest that neither subjective nor objective measures alone make that clear contribution to our understanding of control and user experience as we previously thought.

REFERENCES

[1] Zeynep Barlas and Stefan Kopp. 2018. Action choice and outcome congruency independently affect intentional binding and feeling of control judgments. *Frontiers in Human Neuroscience* 12 (2018), 137.

[2] Joanna Bergstrom-Lehtovirta, David Coyle, Jarrod Knibbe, and Kasper Hornbæk. 2018. I really did that: Sense of agency with touchpad, keyboard, and on-skin interaction. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 378.

[3] Matthew Botvinick and Jonathan Cohen. 1998. Rubber hands 'feel' touch that eyes see. *Nature* 391, 6669 (1998), 756–756.

[4] Niclas Braun, Stefan Debener, Nadine Spychala, Edith Bongartz, Peter Sorös, Helge Heinrich Oskar Müller, and Alexandra Philipsen. 2018. The senses of agency and ownership: A review. *Frontiers in Psychology* 9 (2018), 535.

[5] Patricia Ivette Cornelio Martinez, Silvana De Pirro, Chi Thanh Vi, and Sriram Subramanian. 2017. Agency in mid-air interfaces. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 2426–2439.

[6] Patricia I. Cornelio Martinez, Emanuela Maggioni, Kasper Hornbæk, Marianna Obrist, and Sriram Subramanian. 2018. Beyond the libet clock: Modality variants for agency measurements. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 541.

[7] David Coyle, James Moore, Per Ola Kristensson, Paul Fletcher, and Alan Blackwell. 2012. I did that! measuring users' experience of agency in their own actions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2025–2034.

[8] Mary Czerwinski, Eric Horvitz, and Edward Cutrell. 2001. Subjective duration assessment: An implicit probe for software usability. In *Proceedings of the IHM-HCI 2001 Conference*. 167–170.

[9] Nicole David, Albert Newen, and Kai Vogeley. 2008. The "sense of agency" and its underlying cognitive and neural mechanisms. *Consciousness and Cognition* 17, 2 (2008), 523–534.

[10] John A. Dewey and Günther Knoblich. 2014. Do implicit and explicit measures of the sense of agency measure the same thing? *PLoS One* 9, 10 (2014), e110118.

[11] Jeffrey P. Ebert and Daniel M. Wegner. 2010. Time warp: Authorship shapes the perceived timing of actions and events. *Consciousness and Cognition* 19, 1 (2010), 481–489.

[12] Kai Engbert, Andreas Wohlschläger, and Patrick Haggard. 2008. Who is causing what? The sense of agency is relational and efferent-triggered. *Cognition* 107, 2 (2008), 693–704.

[13] Franz Faul, Edgar Erdfelder, Axel Buchner, and Albert-Georg Lang. 2009. Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods* 41, 4 (2009), 1149–1160. DOI:https://doi.org/10.3758/BRM.41.4.1149

[14] Tiare Feuchtner and Jörg Müeller. 2017. Extending the body for interaction with reality. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 5145–5157.

[15] Tiare Feuchtner and Jörg Müller. 2018. Ownershift: Facilitating overhead interaction in virtual reality with an ownership-preserving hand space shift. In *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology*. 31–43.

[16] Mar Gonzalez-Franco and Tabitha C. Peck. 2018. Avatar embodiment. Towards a standardized questionnaire. *Frontiers in Robotics and AI* 5 (2018), 74.
[17] Patrick Haggard, Sam Clark, and Jeri Kalogeras. 2002. Voluntary action and conscious awareness. *Nature Neuroscience* 5, 4 (2002), 382.
[18] Patrick Haggard and Manos Tsakiris. 2009. The experience of agency: Feelings, judgments, and responsibility. *Current Directions in Psychological Science* 18, 4 (2009), 242–246.
[19] Takayoshi Hagiwara, Maki Sugimoto, Masahiko Inami, and Michiteru Kitazaki. 2019. Shared body by action integration of two persons: Body ownership, sense of agency, and task performance. In *Proceedings of the 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR).* IEEE, 954–955.
[20] Sandra G. Hart and Lowell E. Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Proceedings of the Advances in Psychology.* Elsevier, 139–183.
[21] Marc Hassenzahl. 2004. The interplay of beauty, goodness, and usability in interactive products. *Human-computer Interaction* 19, 4 (2004), 319–349.
[22] Marc Hassenzahl, Michael Burmester, and Franz Koller. 2003. AttrakDiff: Ein fragebogen zur messung wahrgenommener hedonischer und pragmatischer Qualität. In *Proceedings of the Mensch & Computer 2003.* Springer, 187–196.
[23] Kasper Hornbæk and Effie Lai-Chong Law. 2007. Meta-analysis of correlations among usability measures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* ACM, 617–626.
[24] Shunichi Kasahara, Jun Nishida, and Pedro Lopes. 2019. Preemptive action: Accelerating human reaction using electrical muscle stimulation without compromising agency. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems.* ACM, 643.
[25] Shunichi Kasahara, Kazuma Takada, Jun Nishida, Kazuhisa Shibata, Shinsuke Shimojo, and Pedro Lopes. 2021. Preserving agency during electrical muscle stimulation training speeds up reaction time directly after removing EMS. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems.* 1–9.
[26] Simone Kühn, Ivan Nenchev, Patrick Haggard, Marcel Brass, Jürgen Gallinat, and Martin Voss. 2011. Whodunnit? electrophysiological correlates of agency judgements. *PLoS One* 6, 12 (2011), 1–6. DOI:https://doi.org/10.1371/journal.pone.0028657
[27] James R. Lewis. 2018. Measuring perceived usability: The CSUQ, SUS, and UMUX. *International Journal of Human–Computer Interaction* 34, 12 (2018), 1–9.
[28] Benjamin Libet, Curtis A. Gleason, Elwood W. Wright, and Dennis K. Pearl. 1983. Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential) the unconscious initiation of a freely voluntary act. *Brain* 106, 3 (1983), 623–642.
[29] Hannah Limerick, James W. Moore, and David Coyle. 2015. Empirical evidence for a diminished sense of agency in speech interfaces. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems.* ACM, 3967–3970.
[30] Matthew R. Longo and Patrick Haggard. 2009. Sense of agency primes manual motor responses. *Perception* 38, 1 (2009), 69–78.
[31] Jean-Luc Lugrin, Maximilian Ertl, Philipp Krop, Richard Klüpfel, Sebastian Stierstorfer, Bianka Weisz, Maximilian Rück, Johann Schmitt, Nina Schmidt, and Marc Erich Latoschik. 2018. Any "body" there? avatar visibility effects in a virtual reality game. In *Proceedings of the 2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR).* IEEE, 17–24.
[32] Antonella Maselli and Mel Slater. 2013. The building blocks of the full body ownership illusion. *Frontiers in Human Neuroscience* 7 (2013), 83.
[33] Rin Minohara, Wen Wen, Shunsuke Hamasaki, Takaki Maeda, Motoichiro Kato, Hiroshi Yamakawa, Atsushi Yamashita, and Hajime Asama. 2016. Strength of intentional effort enhances the sense of agency. *Frontiers in Psychology* 7 (2016), 1165.
[34] James Moore and Patrick Haggard. 2008. Awareness of action: Inference and prediction. *Consciousness and Cognition* 17, 1 (2008), 136–144.
[35] James Moore, Christoph Teufel, Naresh Subramaniam, Gregory Davis, and Paul Fletcher. 2013. Attribution of intentional causation influences the perception of observed movements: Behavioral evidence and neural correlates. *Frontiers in Psychology* 4 (2013), 23. DOI:https://doi.org/10.3389/fpsyg.2013.00023
[36] James W. Moore. 2016. What is the sense of agency and why does it matter? *Frontiers in Psychology* 7 (2016), 1272.
[37] James W. Moore, D Middleton, Patrick Haggard, and Paul C. Fletcher. 2012. Exploring implicit and explicit aspects of sense of agency. *Consciousness and Cognition* 21, 4 (2012), 1748–1753.
[38] Jakob Nielsen. 1994. *Usability Engineering.* Morgan Kaufmann.
[39] Dario Pittera, Elia Gatti, and Marianna Obrist. 2019. I'm sensing in the rain: Spatial incongruity in visual-tactile mid-air stimulation can elicit ownership in VR users. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems.* 1–15.

[40] S. K. Poonian and Ross Cunnington. 2013. Intentional binding in self-made and observed actions. *Experimental Brain Research* 229, 3 (2013), 419–427. DOI:https://doi.org/10.1007/s00221-013-3505-5

[41] Nina Rosa, Remco C. Veltkamp, Wolfgang Hürst, Tanja Nijboer, Carolien Gilbers, and Peter Werkhoven. 2019. The supernumerary hand illusion in augmented reality. *ACM Transactions on Applied Perception* 16, 2 (2019), 1–20.

[42] Naho Saito, Keisuke Takahata, Toshiya Murai, and Hidehiko Takahashi. 2015. Discrepancy between explicit judgement of agency and implicit feeling of agency: Implications for sense of agency and its disorders. *Consciousness and Cognition* 37 (2015), 1–7. DOI: https://doi.org/10.1016/j.concog.2015.07.011

[43] Jeff Sauro and James R. Lewis. 2009. Correlations among prototypical usability metrics: Evidence for the construct of usability. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* ACM, New York, NY, 1609–1618. DOI:https://doi.org/10.1145/1518701.1518947

[44] Ben Shneiderman. 1983. Direct manipulation: A step beyond programming languages. *Computer* 16, 8 (1983), 57–69.

[45] Ben Shneiderman. 2010. *Designing the User Interface: Strategies for Effective Human-Computer Interaction.* Pearson Education India.

[46] Matthis Synofzik, Gottfried Vosgerau, and Albert Newen. 2008. Beyond the comparator model: A multifactorial two-step account of agency. *Consciousness and Cognition* 17, 1 (2008), 219–239.

[47] Paul van Schaik, Marc Hassenzahl, and Jonathan Ling. 2012. User-Experience from an inference perspective. *ACM Transactions on Computer-Human Interaction* 19, 2 (2012), 25 pages. DOI:https://doi.org/10.1145/2240156.2240159

[48] Johann Wentzel, Greg d'Eon, and Daniel Vogel. 2020. Improving virtual reality ergonomics through reach-bounded non-linear input amplification. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems.* 1–12.

Received July 2020; revised August 2021; accepted October 2021