Using the Implicit Association Test for Interface-Based Evaluations

Tiago Devezas
Research Center for Assistive Information and Communication Solutions (AICOS)
Fraunhofer Portugal
Porto, Portugal
tiago.devezas@fraunhofer.pt

Bruno Giesteira
Faculty of Fine Arts
University of Porto
Porto, Portugal
bgiesteira@fba.up.pt

***Abstract*— Non-instrumental dimensions, the aspects of a product that go beyond its ability to help achieve goals efficiently, are increasingly important in User Experience (UX) research. These dimensions, which include qualities like aesthetics and symbolism, are mainly assessed by self-reports, research has shown. However, respondents can provide wrong answers, willingly, due to concerns like social desirability and self-presentation, or unwillingly, due to the inability to access their inner states. We explored if one implicit measuring method, the Implicit Association Test (IAT), can be used to complement or replace self-report measures. Participants completed six IATs and explicit measures to determine their attitudes toward products represented by pictures of their interfaces. Two non-instrumental dimensions were assessed: valence and self-identification. Overall, implicit and explicit measures displayed a medium correlation. When comparing the correlations between the IATs for the two assessed dimensions and the corresponding explicit measures, similar strong effects were found. This suggests that the IAT bears further exploration as a complement or alternative to self-report methods.**

Keywords-Implicit Association Test; interface evaluation; aesthetics; User Experience

I. INTRODUCTION

In this paper, we explore if one implicit measuring method, the IAT, is a valid complement or alternative to explicit measures for interface-based product evaluations. Empirical UX research relies heavily on self-report measures, such as questionnaires and interviews [1]. However, during such data collection methods, responses can be distorted, deliberately or unconsciously [3].

The IAT is a measure of association strength between concepts that relies on response latency [8]. It is believed to better reflect automatic attitudes than explicit measures and reduce the influence of self-report biases. It has also been shown to be resistant to deliberate faking [25].

The IAT is scientifically accepted as an implicit measure due to its high internal consistency, a rare occurrence for latency-based measures, and satisfactory test-retest reliability [23]. Additionally, the IAT is simple and fast to administer, requiring only a common PC.

We assessed the participants' aesthetic judgments of valence and self-identification towards interface images using the IAT and explicit measures. Interface aesthetics is believed to play a major role in users' perception of the system's usability [26] [29]. There's evidence in the literature indicating that aesthetic evaluations can occur in a spontaneous and automatic manner [21] [24]. We believe that the latency based nature of the IAT makes it an adequate method for measuring such dimensions.

In Section II, we present some of the limitations of self-report measures and how they led to the development of implicit measuring techniques. The IAT, the implicit measure used in this study, is described in Section III. In Section IV, we discuss how aesthetics can affect the perceived usability of a system. Section V talks about how judgments of aesthetic nature can be formed very quickly and why the IAT is an apt method to evaluate them. In Section VI, we describe the design and methodological implementation of the study, followed by a presentation of results, in Section VII. Finally, in Section VIII, we discuss how the results we obtained suggest that the IAT is a method that bears further study for UX research purposes.

II. SELF-REPORT LIMITATIONS AND IMPLICIT MEASURES

In a 2011 study [1], the authors found that two self-report methods - questionnaires and interviews - comprised more than half of the data collection methods used for empirical UX research. The authors, who reviewed 66 empirical studies from 51 publications, also found that the most assessed dimensions are emotions, enjoyment and aesthetics.

There's ample evidence that self-report techniques satisfy important psychometric criteria such as usefulness and efficiency. On the other hand, it is also known that they come accompanied by some limitations [3].

These issues include social-desirability and self-presentation motivations, which can make respondents purposefully distort their answers, particularly when these answers are believed to violate social norms, jeopardize one's self-image or go against the stereotypical answer [3].

Wrong answers can also be provided not deliberately but due to the respondents' inability to introspectively access previously formed attitudes [7].

In order to be able to infer mental contents while overcoming these limitations, researchers, mainly from the field of psychology, developed several techniques known as implicit measures [6].

These methods are called implicit because they aim to measure implicit attitudes, which, according to [7], are "*introspectively unidentified (or inaccurately identified) traces of past experience that mediate favorable or* *unfavorable feeling, thought or action toward social objects*".

They're also implicit in the sense that they don't depend on the awareness of the participants relatively to what's being assessed with the procedure. Implicit measures are thus more sensitive to the spontaneous, automatic evaluations that can be assumed to guide real-life behavior and less likely to be influenced by factors like social desirability and self-presentation [4].

III. IMPLICIT ASSOCIATION TEST

The IAT, developed by Greenwald, McGhee and Schwartz [8] is probably the most well-known of all the implicit measuring methods.

The IAT indirectly measures the strengths of associations among concepts by requiring participants to sort stimulus exemplars from two pairs of concepts (e.g., Flower and Insect and Good and Bad) using just two response options [23].

The rationale behind the IAT is that sorting is facilitated (i.e., response latency should be lower) when two strongly associated concepts (e.g., Flower + Good) require pressing a response key and another pair (e.g., Insect + Bad) requires pressing the other response key. In contrast, when strongly associated concepts require a different response key (e.g., Insect + Good and Flower + Bad), sorting should be slower. The difference in response latency between these two tasks, called the IAT effect, is taken as an indicator of the degree of the strength of association between concepts.

As a measure of association strength between concepts, the IAT can potentially reveal different associations than the ones available introspectively and explicitly reported [14].

Among the factors that contribute to the IAT's acceptance as an implicit measure, are its high internal consistency, a rare occurrence for latency-based measures, and its satisfactory test-retest reliability [23].

It has also been shown that the IAT effect is resistant to several procedural artifacts. These include the hand assigned to each category, the variability in the number of items used to represent the concepts, the subject's familiarity with the items used to represent the concepts, the variability in the response-stimulus interval and the order of the mixed categorization task (as long there's counterbalancing of the order of the study) [3].

There is also evidence of the IAT's high resistance to faking when compared to self-report measures. For example, Steffens [25] found that the IAT, while not immune to faking, is much harder to fake than explicit measures.

Lucas and Baird [15] mention that potential self-report errors are unlikely to be shared across different measuring methods. They advocate a multi-method approach where self-report measures are complemented or validated by other techniques like implicit measures.

With this in mind, coupled with the fact that there aren't examples in the literature of the IAT being used in a UX research context, we aim to explore the potential of this method for interface-based product evaluations.

We assessed two non-instrumental dimensions regarding the evaluated products: valence and self-identification. In the valence IATs, interface pictures of the target products were paired with words pertaining to the attribute categories "Good" and "Bad".

For assessing the self-identification construct, i.e., the degree through which one associates itself with a given product, the image stimuli representing the target products were paired with words representing the attribute categories "Me" and "Others".

These are two constructs that the literature indicates the IAT is capable of assessing. According to Brunel et al. [3], "*the IAT can provide implicit measures of automatic attitudes, self-concepts, self-esteems, and stereotypes.*"

Since we wanted the evaluation to be based as much as possible on aesthetic judgments, no interaction with the products depicted by their interface pictures happened during the study.

In UX practice and research, it is common to design interfaces in non-interactive mediums (e.g., paper, digital images). The ability to evaluate non-interactive interfaces adds to the IAT's usefulness for UX purposes.

IV. AESTHETICS AND PERCEIVED USABILITY

There's evidence that interface aesthetics play a major role in perceived usability, influencing the users' perceptions regarding a system's ease of use not only before, but also after the user has interacted with the system.

This relationship between interface aesthetics and perceived usability has been demonstrated by Tractinsky [26]. The author conducted a study to validate and replicate a study conducted in Japan by Kurosu and Kashima that found that interface aesthetics play a major role in people's perceptions of apparent usability. The author was able to replicate the results in a different cultural setting (Israel), thus concluding that the relationship between perceived interface aesthetics and apparent usability is culturally independent. Moreover, the close relationship between these two dimensions also increases the likelihood that aesthetics may considerably impact system acceptability.

In a subsequent study, Tractinsky, Katz and Ikar [29] explored the relationship between interface aesthetics and usability both before and after users interacted with the system. In addition to corroborate that the users' perception of interface aesthetics is highly correlated with the system's perceived usability, the authors also found, to their surprise, that post-experiment perceptions of the system usability are influenced by the interface's aesthetics, not by the actual usability of the system.

As such, products more strongly associated with positive aesthetic valence might be perceived as easier to use and more easily accepted.

V. AUTOMATIC AESTHETIC JUDGMENTS

The IAT is a timed task which requires the participants to act as quickly as possible while avoiding making mistakes.

Thus, one can question if a limited exposure time to stimuli of variable complexity suffices for evaluative purposes.

We don't believe this to be an issue, due to the evidence indicating that consistent aesthetic evaluations can be formed rather quickly.

Lindgaard et al. [15] found that evaluations of the visual appeal of web homepages after only 50ms of exposure were highly correlated with judgments made after 500ms. The authors state that as little as 50ms might be enough for users to form a highly consistent aesthetic impression of a web homepage.

In a study aiming to replicate and expand Lindgaard et al. findings [15], Tractinsky and colleagues [28] found that aesthetic judgments could be formed after an exposition of only 500ms and that these perceptions are fairly stable, particularly in the case of extreme evaluations. The authors thus suggest that visual aesthetics plays an important role in user's evaluation of web pages and interactive systems in general.

Similarly, Locher et al. [16] found that ratings of paintings made after 100ms were highly correlated with ratings made after unlimited exposure.

Regarding the processes involved, Hekkert [11] believes that aesthetic pleasure or displeasure results uniquely from sensory perception and occurs at initial, mostly automatic and perceptual levels.

According to Norman's [22] framework, these automatic, sensory-based judgments are processed at the "visceral" level. Jordan [13] also speaks of pleasure originated by sensory perception, the physio-pleasure.

Tractinsky [27] speaks of the need to consider this "visceral" beauty as having a major influence on evaluations of beauty.

We believe that the IAT is an adequate tool to assess judgments of aesthetic valence, since it is believed to tap automatic processes.

This assumption is the behind some studies from the field of experimental aesthetics, which employed the IAT for assessing automatic aesthetic evaluations of paintings and architectural styles [21] [24], as well as visual patterns [20] [2].

VI. METHOD

A. Procedure

Eight participants (four female and four male) volunteered to participate in the study after being contacted by e-mail. All had normal or corrected to normal vision and were collaborators of the Fraunhofer Portugal AICOS institute, in Portugal. The average age was of 27.25 (SD = 4.06). The study took place at the institute's facilities. Participants were seated in front of a laptop running the Windows 7 operating system with a screen resolution of 1920x1080 pixels. A trial version of Millisecond's Inquisit software was installed in the laptop and used to administer the IATs and self-report measures.

At the beginning of each session, participants were asked to follow the on-screen instructions and given an opportunity to clarify any doubts. Each session took about 35 minutes.

The recommendations found in [23], regarding the IAT's structure and the need to counterbalance the order of the combined tasks and the presentation of measures were followed. Likewise, more than a single stimulus was used per category in order not to compromise IAT effect's magnitude and reliability.

B. Measures

1) IAT: All the IATs used in this study were based on the Picture IAT script available on Millisecond's Inquisit website.

Some adaptations were made, namely the translation of all the text displayed from English to European Portuguese, the participants' native tongue. The stimuli were also altered to fit the study's purposes.

Four pairs of target concepts were assessed: two mobile operative systems – Android and iOS; two Portuguese newspaper web sites – P3 and i Newspaper; two versions of an interface from a Web application in development – Prototype and Mockup; and two modifications of the Android mobile OS which provide a set of apps targeted at older adults – Smart Companion and Fujitsu. Figure 1 and Figure 2 show some exemplars of the image stimuli used in the IATs.

The order in which the IATs and corresponding explicit measures were presented to half of the participants is displayed in Table I. The other half completed them in reverse order.

Four IATs assessed the valence construct by pairing interface pictures of each target product with words representing the attribute categories "Good" and "Bad".

The self-identification construct was assessed in two IATs, where interface pictures from each product were paired with words from the categories "Me" and "Others".

TABLE I. ADMINISTERED IATS AND EXPLICIT MEASURES

Target concepts	Image stimuli	Attribute categories	Word stimuli
Android/iOS	6	Good/Bad	16
Android/iOS	6	Me/Others	10
Android/iOS valence and self-identification explicit measures			
P3/i newspaper	6	Good/Bad	16
P3/i newspaper valence explicit measure			
Prototype/Mockup	4	Good/Bad	10
Prototype/Mockup valence explicit measure			
Smart Companion/Fujitsu	8	Good/Bad	16
Smart Companion/Fujitsu	8	Me/Others	10
Smart Companion/Fujitsu valence and identification explicit measures			

The script used in this study computes the IAT scores using the improved algorithm developed by Greenwald, Nosek and Banaji [9]. No modifications were made to the scoring procedures.

Table I shows the number of stimuli used to represent the IATs target concepts (images) and attribute categories (words).

Even though the degree of familiarity of the participants with the concepts varied, this is a procedural factor the IAT is resistant to [3].

 ISBN: 978-1-61208-325-4

Figure 1. Exemplars of the image stimuli used in the Smart Companion/Fujitsu and Android/iOS IATs

2) Explicit measures: Two bipolar scales with five and seven points were used for assessing, respectively, the valence and self-identification constructs. Like with the IAT, the self-identification dimension was only assessed for two of the four product pairs.

The scale answers were coded from negative to positive values to represent a relative measure conceptually similar to the IAT D score (e.g., for the targets Smart Companion and Fujitsu, a score of -1 would indicate a moderate preference for Fujitsu relative to Smart Companion).

Since the order of presentation of measures was counterbalanced, half of the participants answered the self-report measures before the IAT. In order to provide reference to these users, one picture of each product's interface was displayed alongside the scales.

C. Stimuli

A total of 24 image stimuli representing the target concepts were used.

The images' width and height were normalized between IATs.

Since the degree of familiarity of the participants with the products varied, a cue was provided to reduce task confusion and facilitate categorization, as recommended by Brunel, Tietje and Greenwald [3].

The cue consisted of a label identifying the category membership of each picture stimuli placed below each image (e.g., images of the interface of Smart Companion had a label beneath saying "Smart Companion").

Labels were not used for the Android/iOS IATs, since these two concepts and the images used to represent them were assumed to be familiar to the users.

When translated from the original Portuguese used in the study to English, the items for the "Good" attribute category were "Joy", "Love", "Peace", "Wonderful", "Pleasure", "Glorious", "Laugh", and "Happy".

For the "Bad" attribute category, the following word stimuli were used: "Agony", "Terrible", "Horrible", "Bad", "Awful", "Failure", "Injured", and "Evil".

Regarding the self-identification IATs, the words for the attribute category "Me" were "Self", "Me", "Am", "Mine" and "I". For the attribute category "Others" they were "His", "Their", "Other", "Them" and "Others".

The words used for the attribute categories were mainly from Inquisit's default set, with some being adapted from literature examples.

D. IAT Design

Table II demonstrates the IAT structure used in this study.

TABLE II. IAT STRUCTURE

Block	Trials	Items assigned to left-key response [e]	Items assigned to right-key response [i]
B1	20	Android pictures	iOS pictures
B2	20	Good words	Bad words
B3	20	Android pictures + Good words	iOS pictures + Bad words
B4	40	Android pictures + Good words	iOS pictures + Bad words
B5	20	iOS pictures	Android pictures
B6	20	iOS pictures + Good words	Android pictures + Bad words
B7	40	iOS pictures + Good words	Android pictures + Bad words

A trial corresponds to the time that mediates from the display of a single stimulus until the correct categorization of that stimulus.

Every time an error was made, a red "X" was displayed below the stimuli and the participant was required to correct the error before proceeding to the next task.

The order of the combined tasks (pictures + words) was counterbalanced. Half of the participants completed the IATs in a similar order to the one above, while the other half completed it with the blocks B1, B3 and B4 switched with B5, B6 and B7.

The script used for the IATs in this study applies the improved scoring algorithm from Greenwald, Nosek and Banaji [9] when computing the IAT effect (D score).

Figure 2. Exemplars of the image stimuli used in the Prototype/Mockup and P3/i newspaper IATs

This improved scoring algorithm provides several benefits when compared to the previous version, namely: a higher resistance to the response speed artifact, almost eliminating the production of extreme scores for slow responders; a higher resistance to prior IAT experience (although not totally eliminating this effect); a better reflection of the underlying association strengths; a more powerful assessment of the relations between association strengths and other variables of interest; an increased power to observe the effect of experimental manipulations on association strengths; and a better insight regarding the individual differences that are due to association strengths rather than other variables.

There's one procedure that is not applied automatically by Millisecond's Inquisit software, which is the elimination of participants with response latency lower than 300ms in more than 10% of the trials.

Raw data was analyzed for every participant and no such cases were found.

VII. Results

A. IAT

Table III shows the average D scores for the IATs. The interpretation of the IAT effect below each scored follows the conventional method: a D score lower or equal to 0.15 in absolute value is considered to indicate little no preference.

Absolute values between 0.16 and 0.35 and 0.36 to 0.65 correspond, respectively, to slight and moderate preference.

Values greater than absolute 0.65 are interpreted as indicating a strong preference.

The IAT effect was only noticeable in two cases: the Smart Companion/Fujitsu and Android/iOS valence IATs.

In both cases a slight preference was found for Smart Companion relative to Fujitsu (D = 0.33) and for Android relative to iOS (D = 0.30).

The remaining average D scores revealed no preference (D <= 0.15) for any of the target concepts.

TABLE III. IAT Average D Scores

Target concepts	Valence	Self-identification
Smart Companion/Fujitsu	D = 0.33 (SD=0.38) Slight preference for Smart Companion	D = -0.09 (SD=0.43) Little to no preference
Android/iOS	D = 0.30 (SD=0.55) Slight preference for Android	D = 0.11 (SD=0.45) Little to no preference
P3/i newspaper	D = -0.07 (SD=0.51) Little to no preference	
Prototype/Mockup	D = 0.003 (SD=0.13) Little to no preference	

B. Explicit measures

Table IV displays the average scores of the self-report scales administered for each product pair evaluated. The scores were coded to directly map to the IAT's D score, providing a conceptually equivalent measure of relative preference.

Smart Companion was moderately preferred to Fujitsu on the valence scale (score = 1) and slightly preferred on the identification scale (score = 0.875). In all the other product pairs evaluated, no relative preference was reported.

TABLE IV. Explicit measures Average Scores

Target concepts	Valence	Self-identification
Smart Companion/Fujitsu	Average score = 1 (SD=1.07) Moderate preference for Smart Companion	Average score = 0.875 (SD=1.73) Slight preference for Smart Companion
Android/iOS	Average score = -0.25 (SD=1.70) Little to no preference	Average score = -0.25 (SD=2.43) Little to no preference
P3/i newspaper	Average score = 0 (SD=1.20) Little to no preference	
Prototype/Mockup	Average score = 0 (SD=0.13) Little to no preference	

C. Data Correlation

1) IAT/Explicit measures: The average correlation coefficient between the IATs and explicit measures was 0.42 ($p > 0.05$), a relatively high correlation coefficient for implicit-explicit measures when compared to the average value of 0.24 found by Hofmann et al. [8].

Implicit and explicit measures tend to correlate when the evaluated objects are not socially controversial [8].

Since interface-based evaluation isn't a particularly sensitive topic, this outcome could be expected.

The correlations between the IAT and the corresponding explicit measure for each product pair evaluated are shown in Table V.

TABLE V. IAT/EXPLICIT MEASURES CORRELATION

Target concepts	Valence	Self-Identification
Smart Companion/Fujitsu	r = 0.12 (p > 0.05)	r = 0.40 (p > 0.05)
Android/iOS	r = 0.89 (p < 0.01)	r = 0.66 (p > 0.05)
P3/i newspaper	r = 0.56 (p > 0.05)	
Prototype/Mockup	r = -0.01 (p > 0.05)	

The lack of correlation in the Smart Companion/Fujitsu valence IAT leads us to hypothesize that it might have been a case self-report bias.

Smart Companion is an Android modification developed by Fraunhofer Portugal AICOS. All the participants in this study were collaborators of the institute and most of them were aware of this fact.

This might have led the participants to explicitly report a preference for the product developed by their employer instead of the competing solution.

After looking at each participant's scores we found that the majority (75%) explicitly reported a moderate to strong preference for Smart Companion. However, some of the individual implicit scores directly contradicted the provided explicit answers.

Summing all this up, it's a reasonable hypothesis that this might have been a case where the IAT scores reflect the method's resistance to self-report biases, thus resulting in a low correlation.

The existence of varying correlations between the IAT and explicit measures has been subject of ample discussion in the literature.

According to Brunel, Tietje and Greenwald [3], the correlation between IAT and explicit measures can be limited by several factors. These include factors related with self-presentation and social desirability concerns, poor introspective access to attitudes that may cause inaccuracy in self-reports and the existence of homogeneous attitudes across specific populations.

Hofmann et al. [12] refer that the IAT and explicit measures tap distinct constructs which can be more or less linked. According to Greenwald and Banaji [7], implicit measures can tap unconscious processes not accessible to explicit, and conscious, self-report. Thus, the correlation between the IAT and explicit self-report measures can depend of the correspondence between conscious and unconscious cognition.

They also mention procedural factors that can interfere with the correlations between measures, like the order of the IAT's combined tasks or the order of presentation of measures.

Even though methodological precautions were taken – including counterbalancing the order of the IAT's combined blocks and the presentation of measures – we can't rule out the possible interference of procedural factors.

2) Valence/self-identification IATs: On average, a strong correlation was found between the valence and self-identification IATs ($r = 0.70$; $p = 0.05$).

Hassenzhal's model of aesthetic experience [10] provides a potential explanation for these results. The author suggests that beauty is strongly related with the hedonic attribute of identification, which he describes as the communication of personally relevant values through objects. The correlation values for the Smart Companion/Fujitsu and Android/iOS pairs are displayed on Table VI.

TABLE VI. VALENCE/SELF-IDENTIFICATION IAT CORRELATION

Target concepts	Valence/Self-identification
Smart Companion/Fujitsu	r = 0.65 (p > 0.05)
Android/iOS	r = 0.75 (p < 0.01)

3) Valence and self-identification explicit measures: In order to see if the valence and self-identification explicit measures displayed a different relationship than the IATs, we also computed correlation coefficients for them.

On average, the valence and self-identification explicit measures were strongly correlated ($r = 0.82$; $p < 0.05$).

The individual correlation values can be found on Table VII.

TABLE VII. VALENCE/IDENTIFICATION EXPLICIT MEASURES CORRELATION

Target concepts	Valence/Self-identification
Smart Companion/Fujitsu	r = 0.70 (p = 0.05)
Android/iOS	r = 0.93 (p < 0.01)

The strong correlation between the valence and self-identification explicit measures parallels what happened with the IATs scores.

 ISBN: 978-1-61208-325-4

As mentioned previously, some authors believe that beauty is strongly related with the concept of identification, so a high correlation between these two constructs both for implicit and explicit measures doesn't comes as a surprise.

Moreover, this convergence between implicit and explicit measures can also be seen as indicative of the potential of the IAT as a complementary or substitutive method for self-report measures.

VIII. CONCLUSION

This study's main goal was to explore if the IAT could be a robust supplement or alternative to self-report measures for UX purposes, particularly for assessing non-instrumental aspects like valence and self-identification during interface-based evaluations. We believe our results suggest that possibility.

The main argument for our assumption is the fact that the IATs and self-report measures scores displayed a medium correlation ($r = 0.42$; $p > 0.05$). This value is substantially higher than the average correlation coefficient ($r = 0.24$) between the IAT and explicit measures found by a meta-analysis conducted by Hofmann et al. [8].

The assumption that the IAT can be used in parallel, or alternatively, to self-report measures is reinforced by the fact that a strong correlation was found both between the valence and self-identification IATs ($r = 0.70$; $p = 0.05$) and the equivalent explicit measures ($r = 0.82$; $p < 0.05$).

As such, these results indicate that the IAT was able to measure similar constructs to the ones evaluated through explicit reports in the context of this study.

Evidence from the literature indicates that the IAT taps spontaneous and automatic processes, which some authors believe that are responsible for guiding behavior [4] and judgments of aesthetic valence [11].

For interface-based evaluations with no interaction, where we assume that the aesthetic dimension plays a major role in the evaluative process, the IAT might represent more adequately user attitudes than explicit measures.

Another point in favor of the IAT is its resistance to unwilling or deliberate errors in self-report measures.

We hypothesize that there might have been a case of self-report bias in our study. Most of the participants knew that one of the products was developed by their employer, which might have led them to explicitly report preference for it relative to the other. However, the IAT didn't mirror those explicitly reported attitudes, resulting in a weak correlation between implicit and explicit measures.

Since there aren't, to our knowledge, previous studies using the IAT for UX research, we believe that this paper might contribute to pave the way for exploring this implicit measuring method as a complement or alternative to self-report measures, particularly for assessing non-instrumental qualities.

As future work, we aim to integrate the IAT in the testing and development cycles of Ambient Assisted Living (AAL) applications alongside self-report measures to assess non-instrumental qualities. By testing and following up with bigger user populations, our goal is to find if the IAT is indeed a more reliable way to collect data about user attitudes than explicit methods.

REFERENCES

[1] J. A. Bargas-avila and K. Hornbæk, "Old Wine in New Bottles or Novel Challenges ? A Critical Analysis of Empirical Studies of User Experience," *Proc. 2011 Annu. Conf. Hum. factors Comput. Syst. - CHI '11*, pp. 2689–2698, 2011.

[2] M. Bertamini, A. Makin, and A. Pecchinenda, "Testing whether and when abstract symmetric patterns produce affective responses.," *PLoS One*, vol. 8, p. e68403, 2013.

[3] F. F. Brunel, B. C. Tietje, and A. G. Greenwald, "Is the Implicit Association Test a Valid and Valuable Measure of Implicit Consumer Social Cognition?," *Journal of Consumer Psychology*, vol. 14. pp. 385–404, 2004.

[4] J. De Houwer, "What are implicit measures and why are we using them?," in *Handbook of implicit cognition and addiction*, 2006, pp. 11–28.

[5] R. H. Fazio and M. A. Olson, "Implicit measures in social cognition research: their meaning and use.," *Annu. Rev. Psychol.*, vol. 54, pp. 297–327, 2003.

[6] B. Gawronski, "Ten frequently asked questions about implicit measures and their frequently supposed, but not entirely correct answers.," *Canadian Psychology/Psychologie canadienne*, vol. 50. pp. 141–150, 2009.

[7] A. G. Greenwald and M. R. Banaji, "Implicit social cognition: attitudes, self-esteem, and stereotypes.," *Psychol. Rev.*, vol. 102, pp. 4–27, 1995.

[8] A. G. Greenwald, D. E. McGhee, and J. L. Schwartz, "Measuring individual differences in implicit cognition: the implicit association test.," 1998.

[9] A. G. Greenwald, B. A. Nosek, and M. R. Banaji, "Understanding and using the implicit association test: I. An improved scoring algorithm," *J. Pers. Soc. Psychol.*, vol. 85, pp. 197–216, 2003.

[10] M. Hassenzahl, "The Interplay of Beauty, Goodness, and Usability in Interactive Products," *Human-Computer Interaction*, vol. 19. pp. 319–349, 2004.

[11] P. Hekkert, "Design aesthetics : principles of pleasure in design Design aesthetics : principles of pleasure in design," *Psychol. Sci.*, vol. 48, pp. 157 – 172, 2006.

[12] W. Hofmann, B. Gawronski, T. Gschwendner, H. Le, and M. Schmitt, "A meta-analysis on the correlation between the implicit association test and explicit self-report measures.," *Personal. Soc. Psychol. Bull.*, vol. 31, pp. 1369–1385, 2005.

[13] P. W. Jordan, "Pleasure with products: Human factors for body, mind and soul," in *Human factors in product design: Current practice and future trends*, 1999, pp. 206-217.

[14] K. A. Lane, M. R. Banaji, B. A. Nosek, and A. G. Greenwald, "Understanding and Using the Implicit Association Test: IV: What We Know (So Far) about the Method.," in *Implicit measures of attitude*, 2007, pp. 59–102.

[15] G. Lindgaard, G. Fernandes, C. Dudek, and J. Brown, "Attention web designers: You have 50 milliseconds to make a good first impression!," *Behaviour & Information Technology*, vol. 25. pp. 115–126, 2006.

[16] P. Locher, E. A. Krupinski, C. Mello-Thoms, and C. F. Nodine, "Visual interest in pictorial art during an aesthetic experience.," *Spat. Vis.*, vol. 21, pp. 55–77, 2007.

[17] R. E. Lucas and B. M. Baird, "Global self-assessment," in *Handbook of multimethod measurement in psychology*, 2005, pp. 29–42.

[18] S. Mahlke, "Aesthetic and Symbolic Qualities as Antecedents of Overall Judgements of Interactive Products," in People and

Computers XX - Engage. Proceedings of HCI 2006, 2007, pp. 57–64.

[19] D. Maison, A. G. Greenwald, and R. H. Bruin, “Predictive Validity of the Implicit Association Test in Studies of Brands, Consumer Attitudes, and Behavior,” *Journal of Consumer Psychology*, vol. 14. pp. 405–415, 2004.

[20] A. D. J. Makin, A. Pecchinenda, and M. Bertamini, “Implicit affective evaluation of visual symmetry.,” *Emotion*, vol. 12. pp. 1021–1030, 2012.

[21] S. Mastandrea, G. Bartoli, and G. Carrus, “The automatic aesthetic evaluation of different art and architectural styles.,” *Psychology of Aesthetics, Creativity, and the Arts*, vol. 5. pp. 126–134, 2011.

[22] D. A. Norman, *Emotional Design: Why We Love (Or Hate) Everyday Things*. New York: Basic Books, 2004.

[23] M. R. Banaji, “The Implicit Association Test at Age 7: A Methodological and Conceptual Review,” *Test*, vol. 663. pp. 265–292, 2007.

[24] M. Pavlović, and S. Marković, “Automatic processes in aesthetic judgment: Insights from the implicit association test”, *Psihologija*, vol. 45. pp. 377-393, 2012.

[25] M. C. Steffens, “Is the implicit association test immune to faking?,” *Exp. Psychol.*, vol. 51, pp. 165–179, 2004.

[26] N. Tractinsky, “Aesthetics and apparent usability: empirically assessing cultural and methodological issues,” in *CHI’97 Proceedings of the ACM SIGCHI Conference on Human factors in computing systems*, 1997, pp. 115–122.

[27] N. Tractinsky, “A Few Notes on the Study of Beauty in HCI,” *Human-Computer Interaction*, vol. 19. pp. 351–357, 2004.

[28] N. Tractinsky, A. Cokhavi, M. Kirschenbaum, and T. Sharfi, “Evaluating the consistency of immediate aesthetic perceptions of web pages,” *International Journal of Human-Computer Studies*, vol. 64. pp. 1071–1083, 2006.

[29] N. Tractinsky, A. . Katz, and D. Ikar, “What is beautiful is usable,” *Interacting with Computers*, vol. 13. pp. 127–145, 2000.