Designing Proactive Objects with Artificial Eyes Based on Perceptual Crossing Paradigm

Siti Aisyah binti Anas[1], Rong-Hao Liang[2], Jun Hu[2] and Matthias Rauterberg[2]

***Abstract*—This paper presents a crowd-sourcing video-based investigation on the users' perception of proactiveness towards an object with artificial eyes through winking. Based on both the session initiation protocol (SIP) and the perceptual crossing paradigm, we synthesized a set of minimally-designed video clips and tested them in Amazon Mechanical Turk with 240 participants. The results show that, in both single- or multi-user scenarios, winking can be a useful expression that makes the user view the object as being proactive and encourages reciprocal input. We also discuss how the object-initiated interaction extends the perceptual crossing paradigm in human-object communication.**

I. Introduction

Nowadays, objects have become smarter and have gradually reduced the need for human intervention [24]. As a result, the objects *disappear* [30] into the background and perceived as passive-reactive objects [1] which impedes the object's ability to learn and develop its understanding of the user. The quality and manner of interaction between a smart object and the user is a parameters not many take advantage of and therefore there are not many discussions that focus on the matter [24].

Perceptual crossing paradigm [2] is a paradigm to show if a person can recognize between an intentional subject or a reactive object. The paradigm emphasized between the different experiences that people have when interacting with an object that shows intention to communicate and an object that is solely reacting/functioning to their presence.

In previously conducted research, Deckers et al. [8] designed an artefact embedded with different perceptive behaviors in the form of dynamic light movements when the presence of a person is detected while Marti [17] developed a companion robot capable of experiencing perceptual crossing with a child to stimulate the child's reflection during playtime and to allow the child to learn social competence with a companion robot. (i.e. reactive object). Deckers et al. [8] and Marti claim their experiment uses perceptual crossing paradigm which engages the object with a human subject. Nevertheless, the framework of Deckers et al. [8] and Marti [17] were focused mainly on the human presence and engagement. This, nonetheless only illustrated the unidirectional activity of the reactive human-object interaction and not the intention of the intentional object.

In reference to perceptual crossing paradigm [2], to distinguish the interaction between an intentional subject and a reactive object, the communication must be proactive. Therefore, to experience perceptual crossing with an object based on the perceptual crossing paradigm, the object has to proactively initiate communication with a person by conveying signs or signals to the person of its intention to communicate. Therefore, to exploit the perceptual crossing paradigm into the human-object interaction design practice, this paper will describe the design of a proactive object-initiated interactions using a visible expressive perceptual quality similar to that of a human eye contact [9] (i.e. artificial eyes). A minimal expression through eye winking was chosen for the object to express its intention to initiate engagement with the user. Winking is a minimalist expression that can capture and retain the object-to-human visual attention in a subtle manner, and works without eyebrow movement [19][23]. Winking also generates less distractions compared to audios.

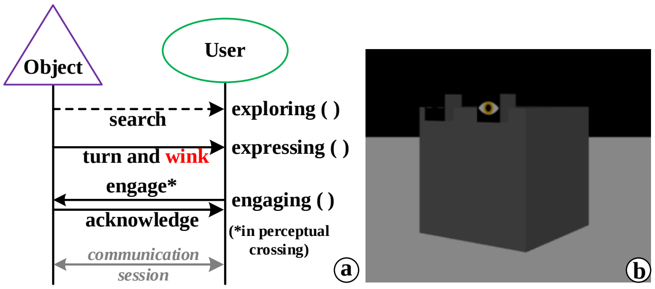


Figure 1. SIPO (Session Initiation for Proactive Object) model based on perceptual crossing. (a) A proactive object finds a user of its interest and expresses its intention by turning its orientation and winking at the user. The user received the intention and engage in the further communication session. (b) A proactive object that turns and winks.

To investigate the proactive object prototype with artificial eyes, we first propose a conceptual model, Session Initiation for Proactive Object (SIPO) based on perceptual crossing paradigm (Fig. 1a). The model, which is adapted from the INVITE method of the Session Initiation Protocol [26], depicts the protocol of the interaction between an object and the user. A face features detection and gaze estimation algorithm were embedded into the object allowing the object to search and find a user using its sensors (i.e., camera, microphone). Once it detects a person is looking at it, the object will then wink to gain further interest from the person as it also conveys the object's intention to engage. If the person maintains his/her gaze with the object after the wink, the object will understand that the person is interested to continue with the interaction, establishing perceptual crossing

*Research supported by Ministry of Higher Education, Malaysia (MOE).

[1] Siti Aisyah binti Anas is with Industrial Design Department, Eindhoven University of Technology, 5612AZ Eindhoven, The Netherlands and Universiti Teknikal Malaysia Melaka, FKEKK, 76100, Durian Tunggal, Melaka, Malaysia (phone: +6062702382; e-mail: aisyah@utem.edu.my).

[2] Rong-Hao Liang, Jun Hu and Matthias Rauterberg are with Industrial Design Department, Eindhoven University of Technology, 5612AZ Eindhoven, The Netherlands
(e-mail: {J.Liang,J.Hu,G.W.M.Rauterberg}@tue.nl).

and the communication session starts. The user is able to terminate the session at any point of the communication.

Based on the SIPO model, an artefact with gender-neutral primitive shape and a pair of expressive gender-neutral cartoonish eyes placed on the top of a primitive shape to avoid the impact of gender stereotyping [5] was developed as shown in Fig. 1(b). To look at its user, the black box orients its body and look to the front instead of rolling its eyes to avoid the accidental negative expressions. It also blinks periodically to keep its appearances natural. To understand the affect of winking to users' perception, a crowd-sourcing video-based user study using Amazon Mechanical Turk (MTurk) was conducted [21]. It enables us to reach a large group of audience which will produce results with cognitive diversity. In the video clips, minimal expression was deliberately used to keep the user's focus on the expression rather than other cultural symbols, as shown in Fig. 1(b).

We synthesized 10 video clips that included single- and multi-user scenarios with or without blinking, and tested them with 240 participants in MTurk. The quantitative results show that, in general, the participants perceived winking to be significantly more proactive than the non-winking video clips, regardless of how many users were in the scenario and how the object turned to the users. Based on the findings, we discuss the limitations and further generalizations to inform future research directions.

II. Design and implementation

A. The SIPO Interaction Model

The state and state change of the object is described in state diagram as shown in Fig. 2. The state diagram is a complement of the protocol diagram shown in Fig. 1(a). A proactive object in its idle state before moving to the 1) exploring state by searching for a user of interest. Once found, it moves to the 2) expressing state and show its intention. It then moves to the 3) engaging state when the user gives reciprocal input (e.g., looks back, vocally responds to it, turns toward the object) within a time threshold, T. Only then perceptual crossing is then established and the communication session starts. The object 4) terminates the communication when the user ignores the signal of the object, or when the user decides to terminate the communication yielding the response time $t_r > T$. After a session is terminated, the object will continue to explore the environment.

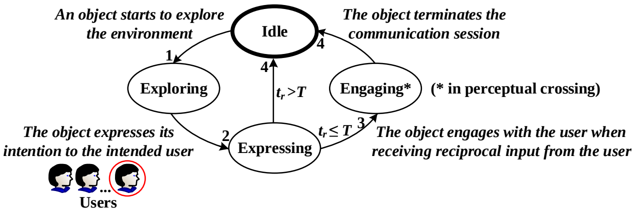

An object starts to explore the environment
Idle
The object terminates the communication session
1
4
4
Exploring
$t_r > T$
Engaging*
(* in perceptual crossing)
The object expresses its intention to the intended user
2
Expressing
$t_r \leq T$
3
The object engages with the user when receiving reciprocal input from the user
...
Users

Figure 2. State diagram of the SIPO model (t_r: response time of the user).

B. Proof-of-Concept Implementation and Example Scenario

To illustrate the application, we choose an everyday object for implementation of a proof-of-concept prototype (e.g., [4][7][15]). Productivity studies suggest that periodically taking a proper break (such as 52 minutes of work followed by a break of 17 minutes) [25] is good for a worker's health and productivity at work. We therefore implement a proactive coffee machine that can remind workers when it is a suitable time for a coffee break.

1) Expressive Perceptual Quality

To enable perceptual crossing, we need to embed at least one perceptual quality into an artefact to make it perceivable by users. Of the possible modalities of perceptual qualities, we choose a visual approach as it is less obtrusive to the environment than sound. For rich expressivity, we chose eyes as the perceptual quality. The reason for is two-fold: firstly, it is natural for people to engage in social interaction by making eye contact, and secondly, eyes are expressive in terms of signaling interest to people. Some conventional smart things such as Furby[1] and Cabbage Patch Kids[2], interactive electronic toys and the Ulo smart surveillance camera[3] have also embedded animated eyes into their expressions. Embedding a pair of eyes into an object allows it to engage people's social interaction skills in human-object interactions.

2) Expressing Intentions

The intention can be enhanced by eye gestures. A single wink is a common signal that could mean a silent agreement between two people, and is usually a friendly gesture implying a degree of intimacy [3]. The simplest way to realise a wink is through a longer blink in one eye, and the duration of winking must be noticeably longer than normal blinking so that people can recognize it. After the user is engaged, the object can also dilate the pupil as a subtle acknowledgement and continue gazing at the user. The intention of initiative taking should be expressed by paying attention to a person. Based on the body formation theory [27], this attention can be expressed by turning to someone to look straight at him/her (body movement).

3) Implementation

Fig. 3 shows the prototype design with a pair of electronic animated eyes attached at the top. A mechanical rotating plate is placed at the bottom to control its body orientation. An Omron HVC-P2 multi-function image sensor module, a compact sensor module that natively supports face detection and gaze estimation, is used to search for users and identify their engagement. A Teensy3.2 microcontroller is used for signal processing. The machine is covered with fabric to hide the buttons, which might tempt the users to push them. The mechanical rotating plate and the image sensor are both well-hidden to avoid distracting the users' attention during interaction with the object.

Fig. 4 shows how the coffee machine initiate the session. The coffee machine explores the environment for people using face tracking, which indicates the number of faces detected and the respective pixel position, $p_i(x,y)$ of each face i. Based on p_i, a linear function is used to estimate the angle θ_i between the sensor module and each face i and to identify the nearest neighbour as a user of interest $U=i$. The coffee machine then shows its intention by either turning its body to the angle θ_U or simply by rolling its animated eyes towards the direction of θ_U if possible. It may then wink at the user as

[1] https://furby.hasbro.com/en-us

[2] https://cabbagepatchkids.com/

[3] https://mu-design.lu/

a friendly invitation, as shown in Fig. 5a. Since the direction of gaze of the user *U* can also be obtained from the imaging sensor, the gaze engagement of the user *U* can also be detected. If the coffee machine finds that *U* is looking back, it dilates its pupils as a subtle acknowledgement and continues to gaze at *U* to maintain the engagement, as shown in Fig. 5b. As investigated by Sejima et al. [28], in human-human communication, a person who dominantly in control of the interaction, their pupillary area will enlarge about 1.5 times. This finding showed that the pupil response has relationships in human-human communication. Therefore, pupil dilation can be used as a signal for maintaining engagement.

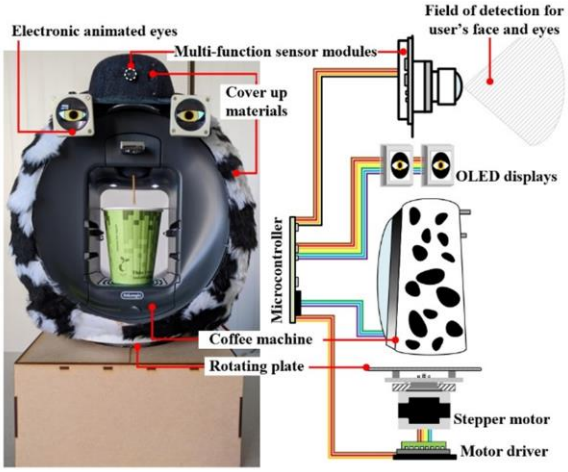


Figure 3. Prototype design of the proactive object.

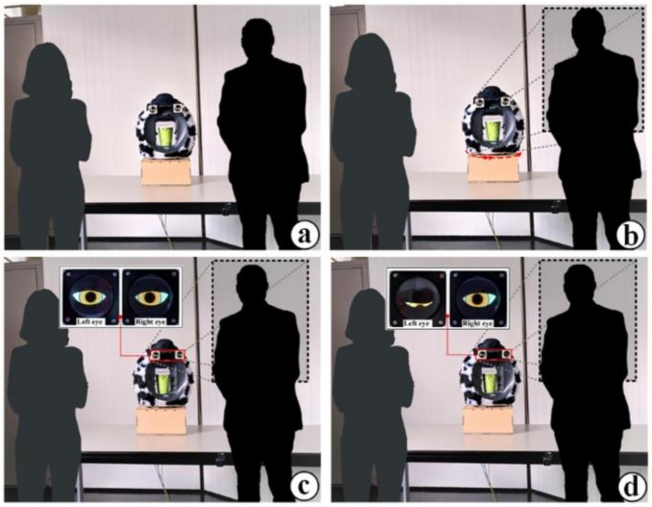


Figure 4. Initiating a session: (a) an idle coffee machine searches for potential users; (b) turns to a user of interest; (c) looks straight; and (d) winks at him.

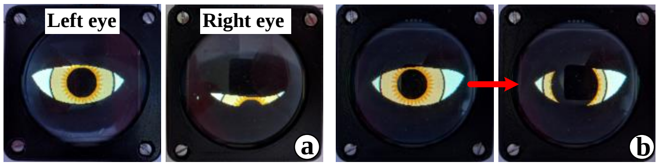


Figure 5. Eye expressions: (a) winking; (b) pupil dilation.

After the connection established through perceptual crossing, the user can decide whether to communicate with the coffee machine using a supported modality, such as using voice commands as input, or to terminate the session of his or her own will. Notably, when the coffee machine finds that *U* is ignoring the invitation by not looking back, it turns to another nearby user and tries to establish a conversation. If it finds that no one has an interest in having a cup of coffee, it hibernates by turning its eyes off until the next coffee break, as shown in Fig. 6.

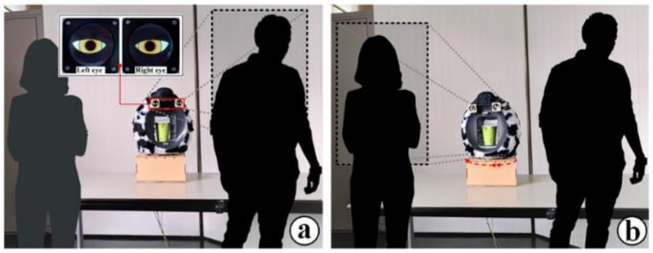


Figure 6. Terminating a session: (a) the user of interest looks away, so the coffee machine (b) turns to the next user of interest.

III. Evaluation

We choose to evaluate this design through a video-based study based on crowd-sourcing for several reasons. First, crowd-sourcing presents an effective paradigm that enables us to gain results from a large group in such a way as to maximise cognitive diversity [11] [13] and enhance group performance [21]. Second, crowd-sourcing has comparable validity to that of a real-world study in terms of testing the graphical design [11] and the results obtained here were also comparable to laboratory studies [11]. Last but not least, representation of future design in the form of video has been proven to be one of the best ways to gather the possible varieties of participants' responses to which they have no direct experience about it [20].

The main assumption of this approach is that participants' reactions to videos provide an efficient way to capture how they would perceive an actual object, similar to previous work (e.g., [14][22]). Moreover, video-based study allows us to exclude unwanted environmental factors and to control the experimental parameters in order to precisely retain the validity of the designed object.

A. Video Synthesis

We tested the basic *Motion+gaze* expressions, *turn (T)* and *wink (W)*, in both single-user (*SU*) and two-user (*TU*) scenarios, where *TU* can be considered a basic multi-user scenario. Since the multi-user scenario involved three entities in a triadic interaction, we acknowledge Box as the main entity wanting to initiate an interaction with either the user (who sees the Box's actions from a first-person viewpoint) or the white figurine (who sees the Box's actions from a third-person viewpoint). In *SU*, Box expresses its intention to engage with the participant. Two possible expressions were therefore tested: 1) $SU(T_1,S_1)$: turning toward the participant without winking, and 2) $SU(T_1,W_1,S_1)$: turning towards the participant and winking. Each of these expressions were followed with a *stare (S)* to maintain the engagement.

In *TU*, Box expressed its intention to engage either the participant or the third person. In addition to the cases in

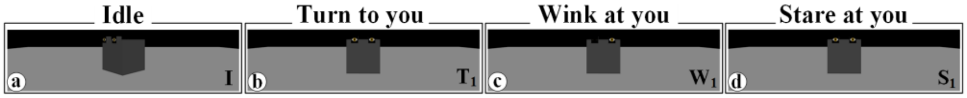


Figure 7. Video animations for single-user scenarios.

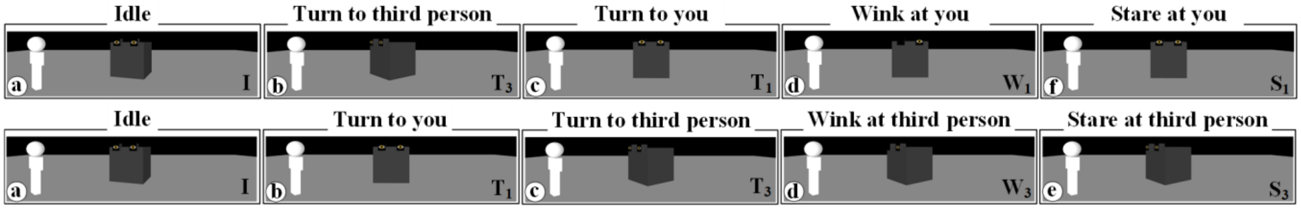


Figure 8. Video animations for two-user scenario

which Box turned to the intended participant directly with a wink *($TU_1(T_1,W_1,S_1)$, $TU_3(T_3,W_3,S_3)$)* or without a wink *($TU_1(T_1,S_1)$, $TU_3(T_3,S_3)$)*, it was also interesting to see how people felt if the object also turned to another participant before turning to the intended participant (the other four cases). Hence, a total of eight expressions were tested. Each expression was followed by a *stare (S)* to maintain the engagement.

TABLE I. TIMING DIAGRAM AND SEQUENCE OF BEHAVIORS IN SINGLE-USER AND TWO-USER SCENARIOS.

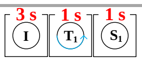
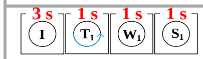
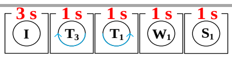
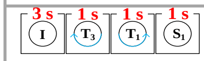
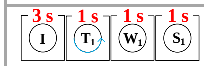
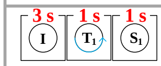
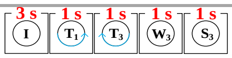
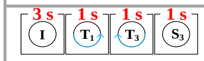
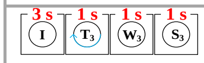
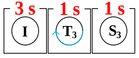
Scenarios	Sequence of behaviors	Interpretation
Single user		$SU(T_1,S_1)$
		$SU(T_1,W_1,S_1)$
Two users (Object initiating interaction with the first person)		$TU_1(T_3,T_1,W_1,S_1)$
		$TU_1(T_3,T_1,S_1)$
		$TU_1(T_1,W_1,S_1)$
		$TU_1(T_1,S_1)$
Two users (Object initiating interaction with the third person)		$TU_3(T_1,T_3,W_3,S_3)$
		$TU_3(T_1,T_3,S_3)$
		$TU_3(T_3,W_3,S_3)$
		$TU_3(T_3,S_3)$

Table I shows the timing diagrams to illustrate the sequence of behaviours for each scenario in Fig. 7 and Fig. 8. For each video, the idle behaviour was set to 3 *s*, whereas the other behaviours (turn, wink, and stare) were set to 1 *s*. The durations within and between natural eye blinking were set to 200 *ms* and 2-3 *s*, respectively, to avoid any confusion between winking and blinking.

B. Participants

A total of 240 participants were recruited from Amazon Mechanical Turk. The participants were evenly separated into two 120-participant groups, A and B.

C. Procedures

Each participant was required to watch a total of six videos, where two videos in *SU* and four videos in *TU*. Each participant first watched the two videos in *SU* and then watched a set of four videos in *TU*. Before participants started to observe the videos, they read a set of instructions that introduced the purpose of the study and explained the meaning of proactive and reactive behaviour using examples of preference functions [18]. Also mentioned in the instructions were the user as the first person, the abstract white figurine as the third person and the Box as the interactor, who wants to interact with either the first or the third person. All video clips have a video cover image of the Box gazing at the participant (Fig. 9). Before pressing the play button, participants were instructed to position themselves at a distance at which they could engage in eye-to-eye interaction with the Box. After watching the video, they were asked to give a response to it before moving to the next step. After approval, compensation of USD$2.50 was given to each participant who submitted a completed survey.

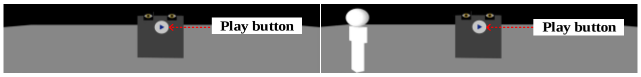


Figure 9. Video cover image in single-user (left) and two-user (right) scenarios.

D. Task and Stimuli

Of the 10 video clips, the eight videos in *TU* were split into two sets. One set was related to initiating interaction with the participant *(TU_1)* while the other was related to initiating interaction with the third person *(TU_3)*. *TU_1* and *TU_3* were assigned to groups A and B, respectively. This is to avoid the participants rated the interaction that is intended toward them more favourable than interaction that is intended toward the third person if the same participants are being asked to evaluate both *TU_1* and *TU_3*. Apart from that, the 2 *(SU)* × 4 *(TU)* × 2 group = 16 conditions were tested and counterbalanced using Latin Square, so that the between-group ordering effects in both *SU* and *TU* were eliminated. For instance, 120 (out of 240) participants watched *$SU(T_1,S_1)$* as the first video and the other 120 watched *$SU(T_1,W_1,S_1)$* as the second.

E. Measurement

Each participant was given a seven-point Likert scale of proactive-reactive measures, where a score of 7 stands for ‘very proactive’ and 1 stands for ‘very reactive’. Along with the rating, each participant was also required to provide an explanation of at least one sentence to each question about the reason for the rating given.

IV. Results

TABLE II. Statistics for the participants' responses to the seven-point scaled Likert proactive-reactive measure in the crowdsourcing study. Wink-related results are highlighted.

Condition	Scenarios	*M*	*Mdn*	*SD*
First impression	$SU(T_1,S_1)$	4.03	4	2.15
	$SU(T_1,W_1,S_1)$	**4.63**	**5**	**1.86**
Single-user scenarios	$SU(T_1,S_1)$	3.90	3.5	2.08
	$SU(T_1,W_1,S_1)$	**4.71**	**5**	**1.91**
Two-user scenarios TTW	$TU_1(T_3,T_1,W_1,S_3)$	**4.57**	**5**	**2.07**
	$TU_3(T_1,T_3,W_3,S_3)$	**4.49**	**5**	**2.20**
TTW̄	$TU_1(T_3,T_1,S_1)$	4.19	5	2.13
	$TU_3(T_1,T_3,S_3)$	4.04	4	2.05
TW	$TU_1(T_1,W_1,S_1)$	**4.32**	**5**	**2.13**
	$TU_3(T_3,W_3,S_3)$	**4.58**	**5**	**1.96**
TW̄	$TU_1(T_1,S_1)$	3.62	3	1.95
	$TU_3(T_3,S_3)$	3.98	4	1.98

TTW: Turn twice and wink

TTW̄: Turn twice, no wink

TT: Turn twice (TTW & TTW̄)

T: Turn once (TW & TW̄)

TW: Turn once and wink

TW̄: Turn once, no wink

W: Wink (TTW & TW)

W̄: No wink (TTW̄ & TW̄)

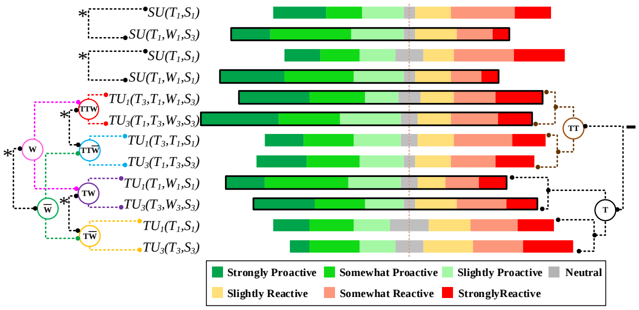


Figure 10. Diverging stack bar chart of participants' responses to the seven-point Likert proactive-reactive measure. Wink-related results are highlighted by an asterisk (*): significant difference (*p*<.05), insignificant difference (*p*>=.05).

A summary of the crowdsourcing study results is given in Table II and Fig. 10. In this section, we describe the quantitative results obtained from the proactive-reactive measure and then discuss the qualitative findings from the user's explanations.

A. Quantitative Results

1) Winking is Generally Considered Proactive

In *SU*, the results of the Shapiro-Wilk test indicated that the distribution of ratings was not statistically normal in both videos (p>0.05). The results of a Mann-Whitney U test showed that the rating of $SU(T_1,W_1,S_1)$ $(Mdn=5,SD=1.91)$ was significantly higher $(Z=–4.13,p<0.05)$ than $SU(T_1,S_1)$ $(Mdn=3.5,SD=2.15)$. Fig. 10 shows that 64% of the participants in $SU(T_1,W_1,S_1)$ rated winking as a proactive behavior, compared to only 39% in $SU(T_1,S_1)$ who rated the object that stared at them as being proactive. The answers for the first video regarding the first impressions of the participants also concurred with this result. The results of a Mann-Whitney U test showed that the participants who watched $SU(T_1,W_1,S_3)$ $(Mdn=5,SD=1.86)$ first also rated it significantly higher $(Z=–2.07,\ p<0.05)$ than those who started from $SU(T_1,S_1)$ $(Mdn=4,SD=2.15)$. Fig. 10 shows that for the first impression, 63% of the participants in $SU(T_1,W_1,S_3)$ found the object to be proactive when it winked at them, whereas in $SU(T_1,S_1)$, the participants were equally inclined to rate the Box as being proactive or reactive (48% for each).

In *TU*, the results of the Shapiro-Wilk test indicated that the distribution of ratings was not statistically normal in all eight videos $(p<0.05)$, and the results of a Mann-Whitney U test showed that the rating of all videos with a wink W $(Mdn=5,SD=2.08)$ was significantly higher $(Z=4.02,p<0.05)$ than all videos without a wink W̄ $(Mdn=4,SD=2.03)$. Based on the Likert scale shown in Fig. 10, 60% of the participants rated the object with a wink W as slightly to strongly proactive, compared to 35% who perceived winking as a reactive behaviour (rating scale between 1–3). Without a wink W̄, 48% of the participants experienced staring as a reactive behaviour, while 44% of them considered the object to be proactive. Collectively, the results suggest that a proactive object that winks makes the user see it as more proactive than one that does not wink.

2) Making More Turns Does Not Affect the Perception

In *TU*, the results of a Mann-Whitney U test show no significant difference $(Z=1.57,p=0.12>0.05)$ between turning to each participant TT $(Mdn=5,SD=2.12)$ and turning only to the user of interest T $(Mdn=4,SD=2.03)$. Fig. 10 reveals that 54% of the participants in TT perceived the object that acknowledged both participants as being proactive, whereas 41% of them considered making more turns to be reactive behaviours. A total of 49% of the participants who experienced fewer turns T reported the object as being proactive, while 43% reported the object as being reactive. In summary, the results suggest that making fewer or more turns does not influence the participants' rating score to infer the object as more proactive or reactive. Therefore, regardless of winking, making one more turn does not make the object appear more proactive. Considering making more turns in multi-user scenarios could increase the time in turn taking, directly turn to the user of interest is more time-saving.

3) First and Third Person Perceived Similar Proactiveness

In *TU*, the results of a Mann-Whitney U test show no significant difference $(p=0.49)$ between participant group A $(Mdn=5,SD=2.06)$ and participant group B $(Mdn=5,SD=2.10)$. Therefore, in the *TU* scenario, regardless of who is the user of interest, there is no effect on the objects' perceived proactiveness.

The diverging stack bar in Fig. 10 shows that participants in $SU(T_1,W_1,S_3)$ and $SU(T_1,W_1,S_1)$ rated the object as proactive with scores of 75% and 77%, respectively. For the TTW scenarios, 57% of the participants in $TU_1(T_3,T_1,W_1,S_3)$ and 60% of the participants in $TU_3(T_1,T_3,W_3,S_3)$ reported the object as proactive. In TW scenarios, 63% of the participants in $TU_1(T_1,W_1,S_1)$ and 53% in $TU_3(T_3,W_3,S_3)$ emphasized the object as being more proactive than reactive. These results therefore suggested that regardless of whether the object winked at the participants or the third person, the participants perceived winking as proactive behavior.

For $TT\overline{W}$ scenarios, 48% of the participants in $TU_1(T_3,T_1,S_1)$ rated the object as being proactive and 46% as reactive, whereas in $TU_3(T_1,T_3,S_3)$, 53% of the participants considered the object to be proactive and 43% as reactive. The results show that without winking and making more turns, participants have a tendency to categorise the object as either proactive or reactive, and therefore no significant differences were found between participant groups A and B. For $T\overline{W}$ scenarios, 64% of the participants in $TU_1(T_1,S_1)$ and 53% in $TU_3(T_3,S_3)$ recognised gazing and turning towards the person of interest as reactive behaviours. The results show that behaviour without winking and less turning influenced the participants to rate the object as more reactive.

B. *Qualitative Results*

1) *First Impressions on Turning with a Wink*

For participants experiencing $SU(T_1,W_1,S_1)$ as their starting point, 75 out of 120 ranked Box as being a proactive object (rating scale between 5–7). Thirty-three participants mentioned that Box was capable of initiating interaction when it winked at them, e.g., *"Box turned in my direction and tried to communicate by winking at me"* (P16) and *"Box turned and winked at me; I felt like Box initiated contact"* (P99). Seven participants reported that Box tried to influence them to collaborate: *"I was surprised Box looked at me and then winked, so I winked back"* (P35), *"When Box winked, I believe it was trying to invite me to do something"* (P79), and *"Box noticed me and we interacted by winking. It tried to get me involved in something interesting"* (P109). Five participants realised that they needed to cooperate: *"Box winked, trying to get my attention and ready for a command from me"* (P3) and *"When Box winked, it affected me and I felt that I needed to do something to respond"* (P141). However, 40 participants perceived the Box as a reactive object (rating scale between 1–3). Twenty-seven participants mentioned winking as reactive behaviour: *"Box's movements felt orchestrated and I thought that it only winked at me because it was told to"* (P6), and "*Box seems to only react when I am making eye contact with him"* (P81).

2) *First Impressions on Turning without a Wink*

Of the participants experiencing $SU(T_1,S_1)$ as their starting point, 57 (out of 120) participants ranked Box as being proactive (rating scale between 5–7). Thirty-four participants mentioned that Box took the initiative and made eye contact, meaning that it was proactive: *"Box turned toward me and we made eye contact"* (P51) and *"Box engaged with me first and looked at me"* (P125). Nonetheless, 58 (out of 120) participants perceived Box as a reactive object (rating scale between 1–3). Thirty-three (out of 120) participants reported that Box was aware of their presence and reacted by turning towards them, e.g., *"Box seemed more reactive as though it became aware of my presence and then looked directly at me"* (P18) and *"Box felt very reactive. It sensed my presence and turned to look at me"* (P48). Nine (out of 120) participants stated that Box was acting in response to their presence rather than controlling the situation, e.g., *"Box was being reactive as if it was waiting for me to act first before doing its next action"* (P42) and *"Box does not do much, but it did seem at least slightly interested in what I needed to do first"* (P50).

3) *User Experiences of Different Types of Turns*

Thirty-seven (out of 60) participants in the TTW and $TT\overline{W}$ experiments pointed out that acknowledgement was proactive, e.g., "*Box displayed proactive action because it acknowledged both our presences"* (P23) and *"Box considered both of us by acknowledging us"* (P7). In addition, 21 (out of 60) participants in TTW and $TT\overline{W}$ reported that the Box showed proactive behavior because it was able to give equal attention to both of them. For instance, *"I appreciate that Box gave its attention to both of us" (P9),* and *"Box seemed to be balancing out engaging both of us"* (P165). Nonetheless, 82 participants in TTW and $TT\overline{W}$ mentioned that since Box was responsive to their presence and incapable of making its own decisions, it was reactive: *"Box responded to our presence, but looked confused as to which one it should interact with"* (P11) and *"Box made a clear effort to observe both of us, but it I think it was unable to decide"* (P159).

Eleven (out of 30) participants in $T\overline{W}$ speculated that the Box was proactive because it showed intention to get a reply from them, "*Box acknowledged me; I assume it was waiting for me to do something"* (P17), and "*Box turned its body to square up with my gaze. I suppose Box was waiting for me to make a move"* (P73). Forty-one (out of 60) participants in TW and $T\overline{W}$ reported that the Box was proactive because it was able to make its own decisions independently, e.g., *"Box seemed more active and interested in interacting with me compared to the other person"* (P50) and *"Box was proactive by aiming its attention at the other person and ignored me"* (P198). Nonetheless, twenty-four (out of 60) participants in TW and $T\overline{W}$ stated that since Box was unable to notice the presence of all the observers, it was reactive. For instance, *"Box only ever turned towards the other person and never myself"* (P144) and *"Box is reactive because it did not acknowledge the other person and was staring blankly at myself"* (P81). Twenty-eight (out of 60) participants in $T\overline{W}$ mentioned that Box reacted to them with no clear action, e.g., "*Box turned to me without giving any clue as to why it was looking at me*" (P19), and *"Box only looked at the other person but did not act towards him or engage"* (P168).

V. Discussion

A. *Quantitative and Qualitative Findings*

Both findings show that the participants significantly able to differentiate the object proactive behavior with a reactive one. Through winking, the object is considered proactive whereas staring is considered reactive. However, directing its body towards the user does not make the object as more proactive. Nonetheless, for a person to experience perceptual crossing with an object, the object needs to express its proactiveness of initiative-taking. In this way, users can experience expressive interaction with the object and the quality of human-object interaction can be improved.

B. *The Appropriateness of Winking*

Winking may be perceived as a controversial behaviour, and can convey a variety of messages that might be either harmless or offensive [19]. For example, embedding winking eyes into a gender-specific object might create sexual implications (e.g., flirting), and winking eyes on a data-sensitive object might create an inappropriate expression of secret-sharing or agreement-making, as winking is often interpreted as meaningful behaviour by a receiver [6]. The meaning behind winking can also vary depending on the situation, culture and gender differences. For example, winking, especially coming from the opposite sex, is perceived as impolite gesture in many Asian countries [3] but the results may vary if an non-gendered object winks as a way of establishing contact. We therefore deliberately designed the artificial eyes with a gender-neutral appearance. In this way, the eyes can be easily mounted on any object that can be considered gender-neutral in most cases. Our crowd-sourced user study evaluated the perceived proactiveness in a gender-neutral setting (i.e., Box and an abstract figurine) to prevent affecting the participants' emotions, which may in turn influence their ratings of the perceived scenarios. Future research may perpetuate on our results in a more context-dependent setting.

C. *Communication Session during Eye-Contact*

The main focus of this study was to enable an object to initiate eye-contact engagement in perceptual crossing with an intention to be perceived. Once eye-contact engagement is detected, feedback (e.g., pupil dilation) should be provided in order to confirm the engagement. Thereafter, during the human-object communication session, natural interactions such as bodily gestures, facial expressions or voice responses could be used to enhance the embodiment while maintaining eye-contact. For example, a simple nod from the user or a beeping sound from the object may be used to show agreement while conversing with each other. Our session initiation technique provides a solid foundation for these interaction schemes.

D. *Scalability of The Interaction Model*

The viability of the SIPO interaction model was confirmed using two-user scenarios. This model extends the traditional dyadic perceptual crossing paradigm to a triadic one, and can be extended to a multi-triadic scenario where a single object interacts with multi-users. However, it is insufficient to further extend the perceptual crossing models for fully connected of many objects and many user scenarios due to the one-to-one perceptual crossing conflicts among these objects. Inter-object connectivity and more accurate sensory and object-level negotiation are requirements for such solution.

E. *Alternative Design and Implementations*

Winking is a subtle expression that can be realised using a simple infrastructure, i.e. an add-on LED matrix. A proactive object also can be augmented by other modalities of perceptual qualities, such as a pair of ears [31]. For example, an object with ears can turn its body orientation to show attention to the user, and then move its ear to invite the users to provide reciprocal input, such as talking to it. The object can also initiate the interactions by making sounds or verbally asking users' attention, but it can be more obtrusive if the context does not fit [10]. There are also other plausible gestures for expressing confirmation or agreement, such as head nodding or hand gestures, but these designs require more mechanical joints or an extra display. Further forms of confirmation, such as providing GUIs such as icons or progress bars on the visual display, can also express confirmation, but this introduces another layer of information that may break natural eye contact and gaze engagement.

The one degree-of-freedom rotation can be realized by simple mechanisms driven by a single motor and is a simple yet effective way to express intention based on the body formation theory [27]. It should be noted that the rotation platform was based on a noisy stepper motor that could be unpleasant for people working in a quiet environment. Silent actuation methods such as joule-heating actuators [29] is possible solutions, although the low torque could mean that the applications are limited to lightweight objects. The expressiveness of an object can be enriched by increasing the number of degrees of freedom, such as adding more joints, the object might be perceived more like a robot. This research attempts to present design guidelines that are generalizable to object design so that even an abstract primitive may be designed to take initiative in engagement.

The camera module used for gaze and face tracking generally lead to privacy issues [12], especially when a device is internet-connected. LIDAR [16] can be used as an alternative solution for tracking user's body formation, although the granularity is lower than camera. Otherwise, data protection and obfuscation services are required, and the system should drain the data without making it for another uses. However, these solutions need to be made explicit to the users, so that they are aware of them.

F. *Future Work*

We used initiation-taking as our starting point for designing human-object perceptual crossing interaction design. We proposed the SIPO interaction model and evaluated the key aspect of expressing the intention. Although we carried out a 240-participant crowdsourcing study for validation, further evaluation in a real-world setting would be useful in understanding the issues in real-world deployment. Understanding how to apply the enabled human-object perceptual crossing interaction design in longer-term problem solving is a possible direction for future work. Future research should also consider enabling the object to be aware of the context in which it is situated. Even though the initiation occurs silently in the background, it might still disturb an ongoing discussion or introduce an unnecessary distraction. Being able to identify and adapt to the social norms of the context could further improve the social appropriateness of session initiation.

VI. CONCLUSION

Based on SIPO model, we have investigated how an object with artificial eyes can simply affect the user perception of the proactiveness (i.e. winking) through a 240-participant crowd-sourcing video-based study. The results show that, in both single- or multi-user scenarios, winking can be a useful expression that makes the user view the object as being proactive and encourages reciprocal input. The results of a crowd-sourcing video evaluation show clear significance, which could warrant larger-scale real-world deployment for behavioural measures in the same setting. We also revealed the ambiguity of explanation through the qualitative study, and discussed how the object-initiated interaction extends the perceptual crossing paradigm in human-object communication. We hope researchers can make use of our results in designing a proactive object that can improve communication between humans and smart objects.

ACKNOWLEDGMENT

This research received funding from the Ministry of Higher Education, Malaysia and Universiti Teknikal Malaysia Melaka (UTeM).

REFERENCES

[1] Angelini, L., Couture, N., Khaled, O.A. and Mugellini, E. 2017. Internet of Tangible Things (IoTT): Challenges and Opportunities for Tangible Interaction with IoT. *CoRR*. abs/1708.02664, (2017).

[2] Auvray, M., Lenay, C. and Stewart, J. 2009. Perceptual interactions in minimalist virtual environment. *New Ideas Psychol.* 27, (2009), 79–97.

[3] Bamio Ares, E.C.J.S.L.A.V.S.S. 2012. *World without words*. Hogeschool van Amsterdam.

[4] Barreiros, C.A.S., Veas, E.E. and Pammer, V. 2017. BioIoT: Communicating Sensory Information of a Coffee Machine Using a Nature Metaphor. *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (Denver, Colorado, USA, 2017), 2388–2394.

[5] Brahnam, S. and Angeli, A.D. 2012. Gender affordances of conversational agents. *Interacting with Computers*. 24, 3 (2012), 139–153.

[6] Burgoon, J.K. and Hale, J.L. 1988. Nonverbal expectancy violations: Model elaboration and application to immediacy behaviors. *Communication Monographs*. 55, 1 (1988), 58–79.

[7] Burneleit, E., Hemmert, F. and Wettach, R. 2009. Living Interfaces: The Impatient Toaster. *Proceedings of the 3rd International Conference on Tangible and Embedded Interaction* (Cambridge, United Kingdom, 2009), 21–22.

[8] Deckers, E., Wensveen, S., Ahn, R. and Overbeeke, K. 2011. Designing for Perceptual Crossing to Improve User Involvement. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vancouver, BC, Canada, 2011), 1929–1938.

[9] Hamilton, A.F. de C. 2016. Gazing at me: the importance of social meaning in understanding direct-gaze cues. *Philosophical Transactions of the Royal Society B: Biological Sciences*. 371, 1686 (2016).

[10] Hansson, R., Ljungstrand, P. and Redström, J. 2001. Subtle and Public Notification Cues for Mobile Devices. *Ubicomp 2001: Ubiquitous Computing* (Berlin, Heidelberg, 2001), 240–246.

[11] Heer, J. and Bostock, M. 2010. Crowdsourcing Graphical Perception: Using Mechanical Turk to Assess Visualization Design. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Atlanta, Georgia, USA, 2010), 203–212.

[12] Hong, J. 2013. Considering Privacy Issues in the Context of Google Glass. *Commun. ACM*. 56, 11 (2013), 10–11.

[13] Ikediego, H.O., Ilkan, M., Abubakar, A.M. and Bekun, F.V. 2018. Crowd-sourcing (who, why and what). *International Journal of Crowd Science*. 2, 1 (2018), 27–41.

[14] Ju, W., Takayama, L. and Nass, C. 2008. Approachability: How People Interpret Automatic Door Movement as Gesture. *Proc. of Design and Emotion* (Hong Kong, China, 2008).

[15] Kao, H.-L.C. and Schmandt, C. 2014. MugShots: Everyday Objects As Social Catalysts. *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication* (Seattle, Washington, 2014), 75–78.

[16] Laput, G. and Harrison, C. 2019. SurfaceSight: A New Spin on Touch, User, and Object Sensing for IoT Experiences. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk, 2019).

[17] Lehmann, H., Iacono, I., Robins, B., Marti, P. and Dautenhahn, K. 2011. Make It Move: Playing Cause and Effect Games with a Robot Companion for Children with Cognitive Disabilities. *Proceedings of the 29th Annual European Conference on Cognitive Ergonomics* (Rostock, Germany, 2011), 105–112.

[18] Lin, Z. and Carley, K. 1993. Proactive or Reactive: An Analysis of the Effect of Agent Style on Organizational Decision-making Performance. *Intelligent Systems in Accounting, Finance and Management*. 2, 4 (1993), 271–287.

[19] Lindsey, A.E. and Vigil, V. 1999. The interpretation and evaluation of winking in stranger dyads. *Communication Research Reports*. 16, 3 (1999), 256–265.

[20] Mancini, C., Rogers, Y., Bandara, A.K., Coe, T., Jedrzejczyk, L., Joinson, A.N., Price, B.A., Thomas, K. and Nuseibeh, B. 2010. Contravision: exploring users' reactions to futuristic technology. *CHI* (2010).

[21] Page, S.E. 2007. *The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools, and Societies (New Edition)*. Princeton University Press.

[22] Pedersen, E.W., Subramanian, S. and Hornbæk, K. 2014. Is My Phone Alive?: A Large-scale Study of Shape Change in Handheld Devices Using Videos. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Toronto, Ontario, Canada, 2014), 2579–2588.

[23] Peterson, J. and Allison, L.W. 1931. Controls of the eye-wink mechanism. *Journal of experimental psychology*. 14, 2 (1931), 144–154.

[24] Petrov, V., Mikhaylov, K., Moltchanov, D., Andreev, S., Fodor, G., Torsner, J., Yanikomeroglu, H., Juntti, M.J. and Koucheryavy, Y. 2017. When IoT Keeps People in the Loop: A Path Towards a New Global Utility. *CoRR*. abs/1703.00541, (2017).

[25] Randolph, S.A. 2016. The Importance of Employee Breaks. *Workplace Health & Safety*. 64, 7 (2016), 344–344.

[26] Rosenberg, J., Schulzrinne, H., Camarillo, G., Johnston, A., Peterson, J., Sparks, R., Handley, M. and Schooler, E. 2002. *SIP: session initiation protocol*.

[27] Roten, Y. de, Darwish, J., Stern, D.J., Depeursinge, E.F. and Warnery, A.C. Nonverbal communication and alliance in therapy: The body formation coding system. *Journal of Clinical Psychology*. 55, 4, 425–438.

[28] Sejima, Y., Sato, Y. and Watanabe, T. 2015. Development of an expressible pupil response interface using hemispherical displays. *2015 24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)* (2015), 285–290.

[29] Wang, G., Cheng, T., Do, Y., Yang, H., Tao, Y., Gu, J., An, B. and Yao, L. 2018. Printed Paper Actuator: A Low-cost Reversible Actuation and Sensing Method for Shape Changing Interfaces. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada, 2018), 569:1–569:12.

[30] Weiser, M. 1991. The Computer for the 21 st Century. *Scientific american*. 265, 3 (1991), 94–105.

[31] Yeo, K.P., Nanayakkara, S. and Ransiri, S. 2013. StickEar: Making Everyday Objects Respond to Sound. *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology* (St. Andrews, Scotland, United Kingdom, 2013), 221–226.