INTERNATIONAL DESIGN CONFERENCE – DESIGN 2022
https://doi.org/10.1017/pds.2022.97

Interaction and Perception of Interaction with 3D Objects during Design Activities

T. Vuletic ✉, A. Duffy, C. McTeague, L. Hay, G. Campbell and M. Grealy
University of Strathclyde, United Kingdom
✉ tijana.vuletic@strath.ac.uk

Abstract
Post study questionnaires are used in design studies to uncover data about design reasoning and intent. A study was conducted where activities the study participants performed were compared to the participants' statements about those activities, collected immediately after the study via a questionnaire. The goal was to explore the reliability of post study evaluations. Disagreements between performed and reported activities were identified, and recommendations made to, where possible, include more objective measures of design activity.

Keywords: design activities, research methodologies and methods, design research, gesture research

1. Introduction

Post study questionnaires are often used in design research (Shavelson et al., 2003, Pedgley, 2007, Jiang and Yen, 2009). At times, they are a necessity, as the primary concern of the study is the focus on the design task, and not the concurrent reporting of it. Post study questionnaires are also used to collect richer information about the design process that is not easily observable (Jansson and Smith, 1991, Badke-Schaub and Frankenberger, 1999, Shavelson et al., 2003). Sometimes studies are performed using surveys or questionnaires as a sole method of data collection, in order to collect feedback from design professionals (Römer et al., 2001), and they can be a limitation of the study (Price and Murnan, 2004). Inclusion of think-aloud protocols is acceptable in specific contexts where instead of recording developments contemporaneously or an observational approach, a post-hoc recording is required (Perry and Krippendorff, 2013). It has, however, been established that protocol analysis is "weak in capturing non-verbal thought processes" (Cross, 2001). Additionally, protocol studies often include small samples of designers, and can contain "subjective inference from verbal and behavioural data" (Hay et al., 2017). Design activities consist of visual and spatial elements that are hard to record verbally (Jiang and Yen, 2009). Designers do not always show evidence of conscious reflection on procedural skills or activities they performed during design process, as they are performing them (Yilmaz and Seifert, 2011). Participant perceptions and actions are intrinsically linked (Krippendorff, 1989), and ensuring both are explored and measured is paramount. In cases where vocalisation of actions would disrupt the design process observed (Chandrasegaran et al., 2018), but were activities explored would still need to be externalised, post study interviews allow us to recover information about designers' intentions, thoughts and reasoning.

Discrepancies between actual and reported data have not been extensively studied in the design context, but in other fields concerns have been raised about the validity of findings relying solely on self-reported measures. Observing media use (Parry et al., 2021) or social network use (Junco 2013), significant discrepancies and only moderate correlations between stated and observed activities have been found.

Actual and hypothetically stated values in behavioural studies have shown similar discrepancies, and even found that certain elicitation methods can induce them (List and Gallet, 2001). These discrepancies, if present, in the context of design studies, could significantly influence the findings.
The study reported in this paper explores the discrepancies between study participants' activities and their perceptions of those activities, reported post study, in order to explore how reliable post study evaluations are. Recommendation for studies where core activity focuses on uninterrupted product design is identified from the study findings.

2. Study Methodology

A user centred study was performed, aiming to explore participants' perception of performed activities. Three manipulative activities (Zoom in, Translate up, and Rotate clockwise) using hand gestures are performed using three different 3D shapes (a chair, a phone, and a sphere) to position and orient them, during a conceptual design activity. These activities and objects were chosen as they were simple and easy to perceive and would not add more complexity to the study than needed. Then the participants answered questions about these activities post study.
The study was performed as an extension of a larger study, aiming to identify designers' intuitive and natural response to the design problem at task, disregarding what is currently achievable by available technology (Vuletic et.al., 2021). It envisaged that in the future conceptual product design would be performed in 3D space employing hand gestures as a sole or one of the modes of interaction. The objective of the larger study was to observe if the shape of the object being manipulated or its recognisability have an effect on the gestures used for the interaction i.e. object positioning and orientation, during conceptual design. If the participant performed the same gesture for the same activity for different objects, it was assumed that the shape of the object did not affect the gesture use. If gestures were different, it was assumed there may have been an effect introduced by the object shape. Recognisability referred to a specific function the manipulated object would have in the physical world. Sphere was the only object that did not have a recognisable usable everyday life counterpart and did not have an associated function it performed. Therefore, it was assumed that if recognisability was not playing a part in the gesture interaction with the object, a larger proportion of gestures used to interact with the sphere would be performed in the same manner as the gestures used to interact with the phone and the chair. The choice of shape and recognisability as parameters to observe was made with regards to the goal of the larger study. However as both observed and self-reported measures of activities were taken, the data was also used to explore the discrepancies between the actual activities and participants perceptions of them. The questions posed to the participants did not mention recognisability to avoid misunderstandings. They were asked:

- “Did the shape of the object influence the gestures you made?”
- “Had the phone/chair been a rectangular box/chair shaped box, would you interact with it the same way?”

Then the gestures performed for the same activity for different objects were compared with the verbal statements provided about the activities, post study, in order to compare stated and observed activities.
Participants in the study were 44 3rd to 5th year Product Design Engineering (PDE) students and graduates. Fifteen were female, and 29 were male. Seven were left handed, 33 right handed and one participant was ambidextrous. They had 4.9 years of CAD (Computer Aided Design) experience on average. They also had an average of 1.4 years of design experience in the professional environment, including internships. Their average age was 22.4.

2.1. Instructions to Participants

Participants were asked to observe an animation of a 3D rendered part being manipulated (this is referred to as an activity), and then use their hands to perform the gesture they believe would result in this activity i.e. imagine they were causing the activity. A flowchart illustrating this process is shown in Figure 1.

https://doi.org/10.1017/pds.2022.97 Published online by Cambridge University Press

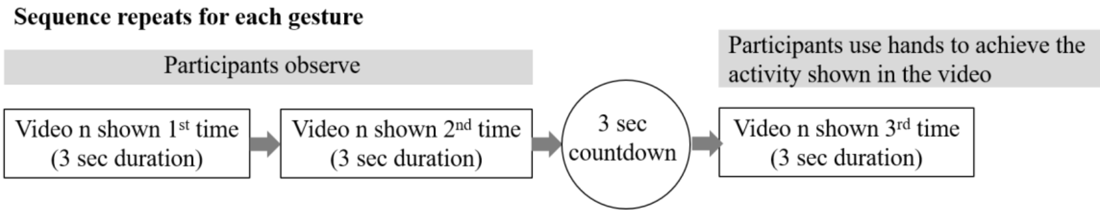

Figure 1. Sequence of activities

The activities and shapes shown to the participants are illustrated in Figure 2. Each activity was shown to the participants three times. They were asked to simply observe it the first two times they see it. Then when they viewed it for the third time, they were asked to pretend they were performing it using their hands i.e. imagine that their hand motions were causing and resulting in the activity shown in the screen. Before they were shown the three activities they will have been asked to perform, they were shown two or three different activities using the same objects e.g. translation in a different direction, in order to test whether they have understood the instructions. When they confirmed they were comfortable with the activity they moved on to the set they were assigned. Sequences of the three activities were randomised, so every tenth participant would perform the same sequence, to reduce the effects of previous activities propagating in the same way through the entire study. For example, participants 1, 11, 21, 31, 41 would have performed the activities in a specific sequence, only participants 2, 12, 22, 32, 42 would have performed the activities in a different specific sequence, etc. The video recordings of the participants pretending they were interacting with the object were analysed to identify their preferred gestures for each activity. Each video was three seconds long, and the countdown before the video was three seconds. This time limitation was chosen in order to record the participants' initial reaction, and reduce likelihood of creation of analogies with CAD interaction.

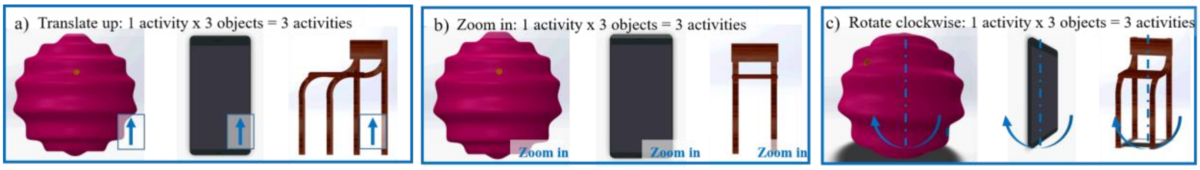

Figure 2. Activities performed for three shapes (images are screenshots from the videos shown)

The participants were not told what the goal of the study was, to avoid influencing their actions. Some participants asked if they were to imagine the object was on the table in front of them, and they were told to interact with it wherever they perceive it.

3. Comparing Type of the Object and Effect on Gesture

In total, 132 gestures were collected. Gestures were first identified and described. If one hand was used the description followed the sequence of "*Hand activity, axis (palm facing/fingertips facing, plane, open/closed/fingers)*". If both hands were used and they were performing the same activity (symmetric), the description followed the same sequence to describe the behaviour of both hands. If both hands were used and they performed different activities (asymmetric), the description followed the same sequence to describe the behaviour of left hand first, and then the right hand. Then each gesture was sketched on a post-it. Gestures were then parsed by grouping the post-its, and at this point only the identical gestures were grouped together. The only interpretation involved was when the same gesture was performed in different planes. These were considered to be the same, since if the point of view was changed the gesture performed would fundamentally be the same. The participants were asked if they perceived the object as 2D or 3D, and where they stated they did not perceive it as 3D or were not sure (five participants in total) three coders were involved in the coding the gestures as 2D or 3D, to ensure that the coding was as objective as possible. Krippendorff's Alpha reliability estimate of 0.846 was calculated. Agreement of $\alpha \geq 0.8$ is customarily required (Krippendorff, 2004). The categorisation process that was to follow was discussed among the three coders for 10% of the sample

https://doi.org/10.1017/pds.2022.97 Published online by Cambridge University Press

of a larger study this study was an extension of. Following the discussion, a baseline was agreed. Small discrepancies between the gestures were not considered e.g. if rotation of a hand around an axis followed a slightly more irregular path or an ellipsoid path it was considered to be a roughly circular motion. Once this was done, the gestures were categorised and analysed for patterns and relations by one coder. Sketches were then assigned unique identifiers. Gestures that performed the same activity following the same path were grouped in the same category. However, each variant (and variants were based on hand pose) was given its own code expressed by the decimal value.

For Translate Up activity, some exemplary gestures are shown in Figure 3. In the first category translation was performed by a hand tracing a vertical line upwards, and the same gesture was performed in fourteen variants – TU01.1-14. The difference between them was the shape the hand forms. None of these hand shapes indicated that the hand "held" the object being moved, which was the case for TU02.1-5, where, while the same vertical upwards trajectory was followed, the hand pinched or grasped or otherwise encircled the object being moved. Similar approach was identified in work by Wobbrock et al. (2009) and Piumsomboon et al. (2013), however they have not presented the sub-variants, and did not differentiate the gestures in terms of the object being held or not. Wobbrock et al. (2009) also draw upon the work of Beringer (2001) who found that pointing is often performed using arbitrary number of fingers, hence concluding that as long as a full hand is not used number of fingers can be disregarded and categorised as a same gesture. This approach was adopted in this study as well. Additional justification for disregarding the number of fingers used was that they do not fundamentally change the gesture performed, while the use of full hand occasionally can indicate a different activity.

Each unique activity was given a unique code and number. For "Rotate clockwise" code was RCW XX.n, for "Translate up" the code was in the form of TU XX.n, and for "Zoom out" the code was in the form of ZO XX.n, where XX is the number of a category, and n is the number of a sub variant.

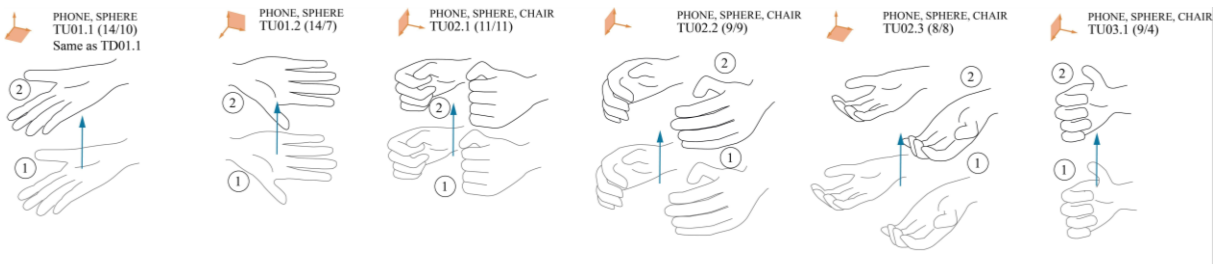

Figure 3. Examples of a number of Translate Up gestures

The gesture classification and categorisation process are illustrated in the flowchart shown in Figure 4.

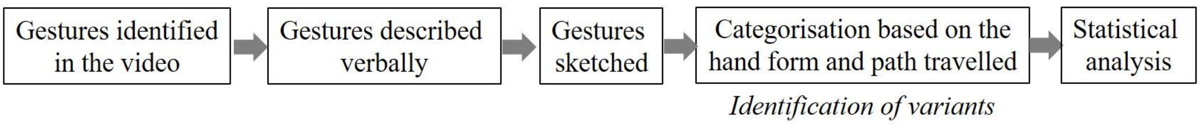

Figure 4. Gesture classification process

Once the gestures were categorised and grouped, those performed by each participant, for three different objects, were observed in order to establish if they interacted with all of the objects using the same gestures for the same activities.

3.1. Statistical Analysis

Regardless of the repetition rates for different shapes and recognisability, all performed gestures were statistically analysed in order to determine if there was agreement between the participants, and if the distribution of gestures by categories could have happened by chance.

Agreement rate was calculated for each of the activities and the categories within it using Agreement Rate (AR) calculation derived by Findlater et al. (2012) and adopted by Vatavu and Wobbrock (2015). AR rate essentially measures the homogeneity for nominal data.

https://doi.org/10.1017/pds.2022.97 Published online by Cambridge University Press

The formula for the AR is given in Equation 1:

$$\boldsymbol{AR_i} = \sum_{k=1}^{q} \frac{n_{ik}(n_{ik}-1)}{n_i(n_i-1)} \quad (1)$$

Where q is the total number of gestures produced by the gesture classification process, $\boldsymbol{n_{ik}}$ is the number of occurrences of a gesture Gk for referent Ri and $\boldsymbol{n_i}$ is the total number of gesture proposals for referent Ri. Guidance for the interpretation of calculated values is: Agreement rates below 0.1 require further data collection. Agreement rates between 0.1 and 0.3 indicate medium agreement. Rates between 0.3 and 0.5 indicate high agreement, and AR above 0.5 indicates a very high agreement.

Agreement Rate is widely used, but not universally accepted as a measurement for selection of appropriate gestures for the inclusion in the consensus set. Tsandilas (2018) suggests that an additional measure should be used to chance-correct the coefficients and specific agreement. Use of Fleiss'κ or Krippendorf's α is suggested. In this study Fleiss' κ was calculated to correct for chance of agreement (see Equation 2):

$$\kappa = \frac{p_a - p_e}{1 - p_e};\ p_e = \sum_{k=1}^{q} \pi_k^2,\ \pi_k = \frac{1}{m}\sum_{i=1}^{m} \frac{n_{ik}}{n_i}; \quad (2)$$

Where m is the total number of items, $\boldsymbol{n_{ik}}$ is the number of gestures per item i having category k, and $\boldsymbol{n_i}$ is the total number of gestures for item i. Guidance for the interpretation of the results is: If κ < 0 agreement is poor. If κ > 0.01 and κ < 0.20 agreement is slight. If κ > 0.21 and κ < 0.40 agreement is fair. If κ > 0.41 and κ < 0.60 agreement is moderate. If κ > 0.61 and κ < 0.80 agreement is substantial. If κ > 0.81 agreement is almost perfect. In this study fair agreement was required at the minimum e.g. κ > 0.21.

AR and κ values for all three activities are given in Table 1.

Table 1. Values for AR and Fleiss κ

Gesture	AR (>0.1)	Fleiss κ
Rotate clockwise	0.307	0.923
Translate up	0.245	0.963
Zoom out	0.143	0.789

AR and Fleiss'κ establish if there is sufficient agreement between the participants proposing gestures from a theoretically infinite set of gestures, and if the agreement is statistically significant, respectively. They however, they do not define what number of repetitions for a specific gesture would be significant enough to consider it for inclusion in a consensus set, if one was to be created for these activities. To determine this, Chi square Goodness of Fit analysis was performed to determine if the number of repetitions for different categories within each activity was likely to happen by chance, and what number of repetitions was expected for each category. Calculations for the Chi square Goodness of Fit follows Equation 3:

$$X^2 = \sum \frac{(O-E)^2}{E}; \quad (3)$$

Where $X^2$is Chi-Square goodness of fit test, O – observed frequency, E – expected frequency. Chi-Square goodness of fit was calculated for one shape at the time where gestures were performed for a number of different shapes. Values for different activities and objects can be found in Table 2. All p values are significantly lower than the Bonferroni corrected p value, indicating that the repetition of gestures was unlikely to happen by chance. However, for Zoom Out activity expected number of repetitions was lower than five, meaning that Chi-Square cannot provided definitive conclusions. Chi-Square calculation for Zoom Out was thus recalculated using the exact one calculation, and those values were reported instead.

https://doi.org/10.1017/pds.2022.97 Published online by Cambridge University Press

Table 2. Values for Chi-Square

Gesture	Bonferroni corrected p value	Chi Sq Phone	Chi Sq Sphere	Chi Sq Chair	Expected number of repetitions
Rotate Clockwise	0.017	0.47×10^{-5}	0.01×10^{-5}	0.02×10^{-5}	8.8; 6.1; 8.6
Translate Up	0.017	0.182×10^{-5}	0.001×10^{-5}	0.001×10^{-5}	6.1; 5.8; 5.5
Zoom Out	0.017	0.541×10^{-4}	0.473×10^{-4}	0.073×10^{-4}	3.3; 3.4;4.4

Observing the gestures that were repeated, it was noticeable that nine unique gestures were repeated for Translate Up activity, six for the Rotate Clockwise, and seven for Zoom Out, but that only five unique gestures were repeated for the same activity for all shapes and for the recognisable shapes only, as shown in Figure 5.

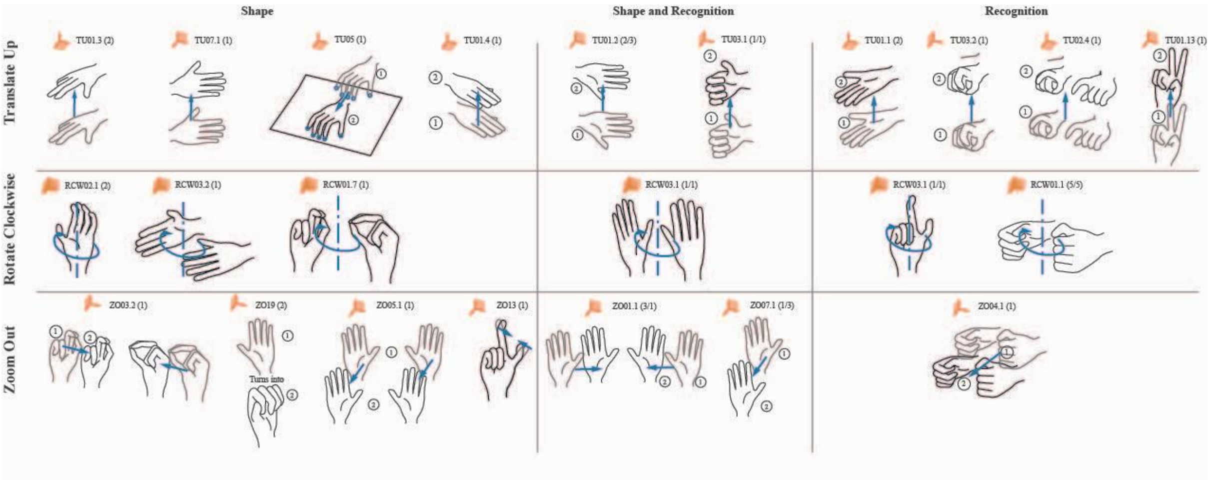

Figure 5. Gestures that were repeated regardless of shape or recognition

3.2. Effect of the Shape of the Object

To observe the effect of the shape of the object Zoom Out, Rotate Clockwise, and Translate Up activities were performed for the irregular sphere, the chair, and the phone. Out of 132 cumulative gestures across all three activities (44 participants performed three gestures each), 22 used the same gestures for the same activities regardless of the shape of the object (16.67%).

Table 3. Agreement between statement about interaction and actual interaction based on shape of the object (for all three objects)

Gesture	Participants who claimed shape influenced interaction (and interaction was different)	Participants who claimed shape didn't influence interaction (and interaction was the same)	Participants who didn't know if there was an influence (interaction was the same)
Rotate clockwise	34	0	1
Translate up	33	0	2
Zoom out	31	1	1

https://doi.org/10.1017/pds.2022.97 Published online by Cambridge University Press

When asked if "the shape of the object influenced the gestures they made", 38 participants stated it did, two stated it did not and four did not know if it did. Comparing stated and observed activities, out of 38 participants that said they would perform different gestures for different shapes, 74% did (averaged out across three different activities), as shown in the first column in Table 3. Two participants stated they performed the same gesture regardless of the object shape, and only one did so and only to zoom out. Among the four that could not tell what they did, 33% performed the same gesture for all shapes.

3.3. Effect of the Recognisability of the Object

The sphere was the only object that was not used in everyday life or had an assigned function. Therefore, it was assumed that if recognisability was not playing a part in the gesture interaction with the object, larger proportion of gestures used to interact with the sphere would be performed in the same manner as the gestures used to interact with the phone and the chair. When asked if "they would interact with a phone/chair the same way they would have if they had been a rectangular box/chair shaped box", 21 participants said they would, 15 said they would not, and eight did not know. Out of 21 participants that claimed they would interact with both objects in the same way only 17% did (averaged out across the three activities). Out of 15 participants who claimed they would perform a different gesture for objects if they were not recognisable, 89% of them have performed different gestures for the sphere and the remaining two objects. Eight participants did not know what they did, and 20% have actually performed the same gesture regardless of recognisability of the object.

Table 4. Agreement between statement about interaction and actual interaction based on the recognisability of the object (for all three objects)

Gesture	Participants who claimed would interact with recognisable and non-recognisable object in the same manner (and interaction was the same)	Participants who claimed would interact with recognisable and non-recognisable object in a different manner (and interaction was different)	Participants who didn't know if there was an influence (interaction was the same)
Rotate clockwise	3	15	2
Translate up	5	13	1
Zoom out	3	12	2

Overall, it seems that different gestures have been performed if the object was not recognisable. However, the difference in shapes of the object might influence the interaction regardless of the recognisability and this may be influencing the participants' activities. Recognisable objects used in this study, the phone and the chair, have very different shapes and perhaps the shape does play a more important role for the choice of gesture than the recognisability of the object does. For example, it has been observed that in some cases the participants interacted with a chair using different gestures than with the other two objects, for example by "holding" the legs of the chair while rotating it or zooming it. These gestures are illustrated in Figure 5, RCW01.1 and ZO04.1. In certain circumstances, a physical chair could be moved around a room using a similar motion. On the other hand, the sphere was in some instances rotated like a desktop globe, with participants indicating an immovable axis in its centre with one hand and using the other hand to "flick" the surface of the globe to initiate the rotation around that axis. This gesture was not repeated frequently, and was thus not further analysed, but it does provide an example of a function being assigned to an object without an obvious function due to its shape.

It should be noted here that recognisability was considered to show influence of the gesture performed if the same gesture was performed for the phone and chair, and different gesture was performed for a sphere. If the same gesture was performed for a sphere and phone and different for a chair, or a same gesture was performed for a sphere and chair and different for a phone, then that is taken as an indicator that recognisability may not have an effect. It was not explored what the influence for the differences were in those cases.

https://doi.org/10.1017/pds.2022.97 Published online by Cambridge University Press

4. Implication of Discrepancy Between Stated and Observed Activities

Participants that perceived their activities as different for different objects majorly did perform different activities, for both shape and recognisability around 74-89%. Of participants who believed they interacted in the same way with a recognisable and non-recognisable object, only 16-19% actually did perform the same gesture. These ratios are shown in Figure 6. While these findings do not provide definitive conclusions on participants' perception of their own activities, they do indicate that perhaps more focus should be put on objective measures rather than participants' perceptions of their own activity.

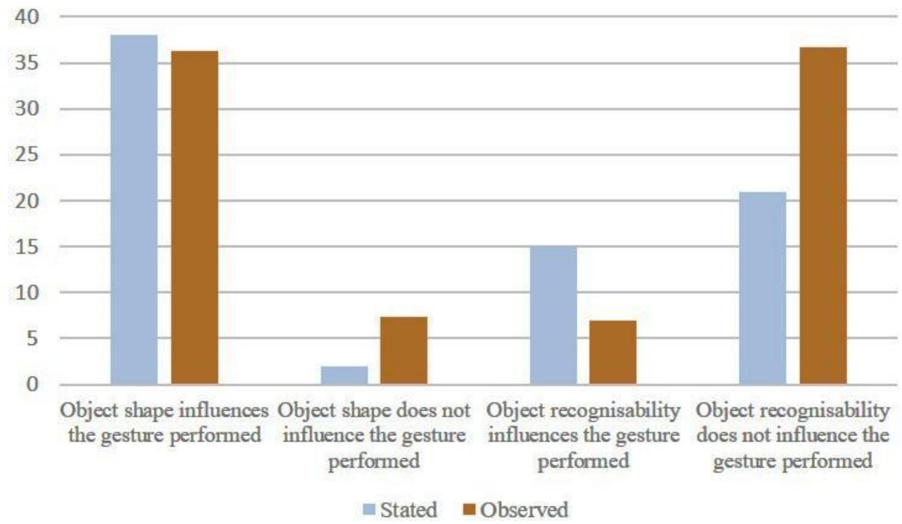

Figure 6. Comparison of stated perceived activities and performed activities

While, generally speaking, the statements and the activities match for the participants that believed shape or recognisability of the object was influential, if the outcomes were calculated using the measures used in Section 3.1, the differences in the context of the full sample might lead to different results.
Chi Square goodness of fit calculations were performed to examine the likelihood of participants performing the gestures by chance for each of the four categories illustrated in Figure 6. Results of the analysis are given in Table 5 and Table 6, for shape-based statements and observations and recognisability-based statements and observations, respectively.

Table 5. Chi Square calculations for shape-based statements and observations

Shape	Stated	Observed
Object shape influences the gesture performed	38	36.33
Object shape does not influence the gesture performed	2	7.33
Analysis results	Chi-sq = 32.4; p=0.000	Chi-sq = 19.26; p=0.000

Table 6. Chi Square calculations for recognisability-based statements and observations

Recognisability	Stated	Observed
Object recognisability influences the gesture performed	15	7
Object recognisability does not influence the gesture performed	21	36.67
Analysis results	Chi-sq = 1; p=0.317	Chi-sq = 20.15 p=0.000

https://doi.org/10.1017/pds.2022.97 Published online by Cambridge University Press

For the shape-based calculation, the stated and observed analysis both show that the results are not likely to have happened by chance, as $p<0.05$. For the recognisability-based calculation, analysis of the observed values would indicate the results were not likely to have happened by chance, as $p<0.05$. However, for the stated values $p=0.317$ is larger than 0.05, and would lead the researchers to the conclusion that the effects may have happened by chance. Hence, if the study was performed without the observation, and the results were based on verbal statements from the questionnaire only, the researchers would have concluded that the repetitions may have happened by chance and that findings cannot be extrapolated from the data. If the study was based on the observation only, the researchers would conclude that the repetitions did not happen by chance and that recognisability may be influencing the gestures. As both types of data were collected, it is now known that the results are conflicting, perhaps indicating that the question in the questionnaire was not clear enough or alternatively showing that perceptions participants have of their own activities are not always reliable.
Combining both methods, activity performance analysis and the post study questionnaire, to infer findings in design studies is likely to lead to information containing more depth. Basing the findings on post study questionnaires only may not lead to reliable results.
This study is an extension of a larger study, focusing on identification of intuitive and natural response to the design problem at task, avoiding influence of previous experience. As natural and intuitive response was the key focus of the study, its design and scope could not include any questions before the study or interactions with participants during the study. In the future it would be interesting to repeat a similar study with three cohorts, one that responds post study, one that is also asked questions prior to the study, and one where there is interaction during the study. That could provide more information on potential reasons for the discrepancies found. Currently, it appears they are due to different participants' abilities to recall or foresee actions, however this is a speculative conclusion only.
Choice of design students was made focusing on the larger study, where it was an appropriate one as by the 3rd the students are considered to have a sufficient grasp of design, but have still not fully adopted the traditional way of working. It would however be beneficial to explore if the outcomes of the study would be different if participants were experienced designers, in the future studies. This would also explore the effect of experience and skill on participants' ability to differentiate between their actions and perceptions.

5. Conclusion

Activities performed in order to interact with different 3D objects and participant statements in the post study questionnaire, reporting on the manner participants believed they interacted with different objects, were compared in the study reported in this paper. Some level of disagreement was identified between the number of activities actually performed, and those reported as performed by participants in the post study questionnaire, as up to 85% of participants correctly reported their activities. The disagreement was more pronounced where participants thought the shape or recognisability of an object did not influence the manner of interaction. While this level of agreement initially may seem acceptable, if the data is to be statistically analysed the outcomes of the analysis of performed and reported data sets could differ significantly. While it is not possible to generalise findings of one study to the entire field of design, a tentative recommendation at this stage would be to consider the use of multiple data collection tools in design studies focusing on the activities that cannot be easily verbalised. Use of post study questionnaires as a sole data collection tool may lead to significant distortion of findings. It would be beneficial to use a second data collection tool e.g. visual or audio, or include an independent second measurement that does not rely on participant perceptions, which could be used to validate the findings.

Acknowledgment

This research was made possible by funding from University of Strathclyde.

References

Badke-Schaub, P. and Frankenberger, E., 1999. Analysis of design projects. Design Studies, 20(5), pp.465-480. https://doi.org/10.1016/s0142-694x(99)00017-4

https://doi.org/10.1017/pds.2022.97 Published online by Cambridge University Press

Beringer, N., 2001, April. Evoking gestures in SmartKom-Design of the graphical user interface. In International Gesture Workshop (pp. 228-240). Springer, Berlin, Heidelberg. https://doi.org/10.1007/3-540-47873-6_25

Chandrasegaran, S., Ramanujan, D. and Elmqvist, N., 2018, June. How Do Sketching and Non-Sketching Actions Convey Design Intent?. In Proceedings of the 2018 Designing Interactive Systems Conference (pp. 373-385). https://doi.org/10.1145/3196709.3196723

Cross, N., 2001. Design cognition: Results from protocol and other empirical studies of design activity. In Design knowing and learning: Cognition in design education (pp. 79-103). Elsevier Science. https://doi.org/10.1016/B978-008043868-9/50005-X

Findlater, L., Lee, B. and Wobbrock, J., 2012, May. Beyond QWERTY: augmenting touch screen keyboards with multi-touch gestures for non-alphanumeric input. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 2679-2682). ACM. https://doi.org/10.1145/2207676.2208660

Hay, L., Duffy, A.H., McTeague, C., Pidgeon, L.M., Vuletic, T. and Grealy, M., 2017. A systematic review of protocol studies on conceptual design cognition: Design as search and exploration. Design Science, 3. https://doi.org/10.1017/dsj.2017.11

Jansson, D.G. and Smith, S.M., 1991. Design fixation. Design studies, 12(1), pp.3-11. https://doi.org/10.1016/0142-694x(91)90003-f

Jiang, Hao, and C. Yen. "Protocol analysis in design research: a review." Journal Paper 78, no. 24 (2009): 16.

Junco, R., 2013. Comparing actual and self-reported measures of Facebook use. Computers in Human Behavior, 29(3), pp.626-631. https://doi.org/10.1016/j.chb.2012.11.007

Krippendorff, K., 1989. Product semantics: A triangulation and four design theories.

Krippendorff, K., 2004. Reliability in content analysis: Some common misconceptions and recommendations. Human communication research, 30(3), pp.411-433. https://doi.org/10.1093/hcr/30.3.411

List, J.A. and Gallet, C.A., 2001. What experimental protocol influence disparities between actual and hypothetical stated values?. Environmental and resource economics, 20(3), pp.241-254. https://doi.org/10.1023/A:1012791822804

Parry, D.A., Davidson, B.I., Sewall, C.J., Fisher, J.T., Mieczkowski, H. and Quintana, D.S., 2021. A systematic review and meta-analysis of discrepancies between logged and self-reported digital media use. Nature Human Behaviour, 5(11), pp.1535-1547. https://doi.org/10.1038/s41562-021-01117-5

Price, J.H. and Murnan, J., 2004. Research limitations and the necessity of reporting them. https://doi.org/10.1080/19325037.2004.10603611

Pedgley, O., 2007. Capturing and analysing own design activity. Design studies, 28(5), pp.463-483. https://doi.org/10.1016/j.destud.2007.02.004

Perry, G.T. and Krippendorff, K., 2013. On the reliability of identifying design moves in protocol analysis. Design Studies, 34(5), pp.612-635. https://doi.org/10.1016/j.destud.2013.02.001

Piumsomboon, T., Clark, A., Billinghurst, M. and Cockburn, A., 2013, September. User-defined gestures for augmented reality. In IFIP Conference on Human-Computer Interaction (pp. 282-299). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-40480-1_18

Römer, A., Pache, M., Weißhahn, G., Lindemann, U. and Hacker, W., 2001. Effort-saving product representations in design—results of a questionnaire survey. Design studies, 22(6), pp.473-491. https://doi.org/10.1016/s0142-694x(01)00003-5

Shavelson, Richard J., Dennis C. Phillips, Lisa Towne, and Michael J. Feuer. "On the science of education design studies." Educational researcher 32, no. 1 (2003): 25-28. https://doi.org/10.3102/0013189x032001025

Tsandilas, T., 2018. Fallacies of agreement: A critical review of consensus assessment methods for gesture elicitation. ACM Transactions on Computer-Human Interaction (TOCHI), 25(3), p.18. https://doi.org/10.1145/3182168

Vatavu, R.D. and Wobbrock, J.O., 2015, April. Formalizing agreement analysis for elicitation studies: New measures, significance test, and toolkit. In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (pp. 1325-1334). ACM. https://doi.org/10.1145/2702123.2702223

Vuletic, T., Duffy, A., McTeague, C., Hay, L., Brisco, R., Campbell, G. and Grealy, M., 2021. A novel user-based gesture vocabulary for conceptual design. International Journal of Human-Computer Studies, 150, p.102609. https://doi.org/10.1016/j.ijhcs.2021.102609

Wobbrock, J.O., Morris, M.R. and Wilson, A.D., 2009, April. User-defined gestures for surface computing. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 1083-1092). ACM. https://doi.org/10.1145/1518701.1518866

Yilmaz, S., Seifert C.M., 2011. Creativity through design heuristics: A case study of expert product design. Design Studies, Volume 32, Issue 4, Pages 384-415, https://doi.org/10.1016/j.destud.2011.01.003.