The role of tangibility and iconicity in collaborative modelling tasks

Dan Ionita[1], Deniece S. Nazareth[2], Alexandr Vasenev[1], Frank van der Velde[2], and Roel Wieringa[1]

[1] University of Twente, Department of Services, Cybersecurity and Safety
Enschede, The Netherlands
d.ionita,a.vasenev,r.j.wieringa@utwente.nl
[2] University of Twente, Department of Cognitive Psychology and Ergonomics
Enschede, The Netherlands
d.s.nazareth,f.vandervelde@utwente.nl

Abstract. In collaborative modelling, a group of stakeholders construct a shared graphical representation of a system. In practice, not all stakeholders may fully comprehend the modelling language. This may reduce their participation, which results in reduced model quality. Our goal is to investigate which features of modelling languages help stakeholders contribute to collaborative modelling tasks.
Earlier research shows that iconicity, i.e. similarity between sign and object, improves understandability, and that tangibility, i.e. physical graspability of signs, improves participation of stakeholders. In this paper we report on a 2x2 factorial experiment that explores for the first time the interaction between iconicity and tangibility in the context of collaborative modelling. In this experiment, tangibility promoted equal participation, and iconicity had a beneficial impact on understandability, modelling speed and model quality. Notably, tangibility magnified the effects of iconicity. We relate these results to previous findings and interpret them in terms of existing theories.

Keywords: tangible modelling; iconicity; collaborative modelling; factorial experiment; system modelling

1 Introduction

When designing or analyzing complex systems such as urban areas, business information systems, or enterprise models in which people and machines cooperate, models of these systems or services need to be developed. To construct these models, collaboration among stakeholders with different backgrounds and expertise is needed. This puts strong constraints on the modelling language: It should be understandable by all stakeholders involved in modelling, even if they are not familiar with modelling languages, and it should promote their participation in the modelling effort.

It is well known that iconicity can enhance understandability and learnability of signs [1, 2, 15]. However, most research on iconicity did not investigate its effects in the context of collaborative modelling.

In: C. Cabanillas, S. España, S. Farshidi (eds.):
Proceedings of the ER Forum 2017 and the ER 2017 Demo track,
Valencia, Spain, November 6th-9th, 2017,
published at http://ceur-ws.org

There is also evidence that tangible modeling languages, by which we mean languages whose concepts are represented by physical, graspable tokens, such as plastic fiches or Lego pieces, have beneficial effects on collaborative modelling efforts [10, 11, 25]. This contrasts with what we call virtual languages, which consist of symbols on paper or on a screen or smartboard. In comparative studies, tangible models were produced faster and were of higher quality than virtual ones [8, 17]. Furthermore, subjects using tangible modelling tools did not divide their tasks, whereas subjects using graphical editors did, which suggests more equal participation of all modelers in tangible modelling than in virtual modelling [16, 17]. However, none of these experiments distinguish the effect of iconicity from the that of tangibility or explore the interaction of the two.

In this paper, we report on a new experiment, with a 2x2 factorial design, in which the modelling language was either tangible or virtual, and either iconic or abstract. We combine the results of this experiment with the results of previous ones, and with existing theory, in a process of analytical induction. Our experiments provide evidence that iconicity not only improves understandability, but also modelling speed and model quality and that tangibility promotes collaboration, by facilitating uniform participation of all group members. The experiments also provide preliminary evidence that tangibility magnifies the positive effects of iconicity as well as the negative effects of abstractness on understandability, modelling speed, and model quality.

2 Theoretical background

In collaborative modelling tasks, stakeholders work in a group to construct a simplified representation of an actual or potential state of affairs. We consider conceptual models, which describe social, physical and/or digital systems as a composition of concepts and relationships. These concepts may be represented by graphical signs on a screen, or by physical signs on a table or another similar surface. In the case of collaborative modelling, the representation is visible to all participants throughout the modelling effort.

Peirce defines a **sign** as "anything which is so determined by something else, called its object" [27]. A sign can be a letter, a written or spoken word, a logo, a Lego block, a diagram or anything else that refers to something beyond itself.

A sign is **iconic** if it perceptually resembles the object it represents [23]. A map can be viewed as an icon of an area, a portrait is an icon of its subject, and a model car can be viewed as an icon of an automobile. In this paper, we call a sign **abstract**[3] if it bears no likeness to the object it represents but is rather related to it arbitrarily or by some (e.g. social or legal) convention. Abstract signs therefore require a dictionary that documents the relationship between the sign vehicle and sign object, and hence are a strain on the memory compared to iconic signs. For example, a word is an abstract representation since its meaning

[3] Such a sign is called *symbolic* in semiotics, but we prefer the term *abstract* in order to emphasize the lack of resemblance (non-iconicity) rather than its reliance on an interpretative rule, as well as to avoid ambiguous use of the word *symbol* [9, p. 237]

is determined by language. Similarly, a box in a UML diagram can only be understood if one is familiar with the UML language.

In addition, we refer to a sign as **tangible** if if has physical form and can be grasped and manipulated by hand. Lego puppets, post-it notes, small scale models and Lego bricks are graspable, but a footprint in the sand or a real-sized prototype of a future home are not, even though they are physical. Tangible signs may have embedded intelligence, such as in bricks on a smart tabletop [7] or FlowBlocks, used to build models of system dynamics [37], or they may be simple non-intelligent objects, such as plastic fiches with text printed on them or 3D printed shapes. Conversely, a sign is **virtual** if it is rendered digitally on a two-dimensional display, such as a computer screen, smart board or smart table and can be only be manipulated indirectly via an input device attached to the same machine. For example, a piece of text in a graphical text editor requires a keyboard to manipulate. Similarly, an icon on a smartboard – even though it can be manipulated by hand – cannot be grasped, and is therefore virtual.

To explain some of our results, we also refer to the concepts of cognitive load and cognitive fit. The **cognitive load** of a task is the total amount of mental effort required to perform a task. The theory of cognitive load suggests that performance improves when conditions are aligned with the human cognitive architecture [28]. Miller [22] claims that the ability to remember and discriminate information can be dramatically expanded by adding dimensional stimuli (such as color, sound, material & space). This suggests that tangible signs are easier to understand than virtual signs. **Cognitive fit** is the reduction of cognitive load of problem-solving by fitting the representation of the problem to the problem itself. Vessey [33] showed that when the representation of concepts or information match a task, problem solving performance for both simple and complex tasks is drastically improved. This suggests that it is easier to construct models using iconic rather than abstract signs.

3 Related work

Bjekoviȁ highlighted the intimate relationship between enterprise modelling concepts, their signs and the community which uses them [5]. Wilmont adds that individual differences in performance on conceptual modelling tasks cannot be explained by training and experience alone and are intrinsically linked to the activation of cognitive mechanisms related to working memory, executive control and attention [35].

Fitzmaurice et al. [7] experimented with tangible user interfaces, called Bricks, which allow interacting with virtual information through an intelligent tabletop. Bricks affords synchronous manipulation, rather than through a single mouse, and has more spatial persistence than virtual signs, allowing users to make better use of spatial reasoning skills and muscle memory [7, page 447]. This suggests that in a group modelling context, tangible signs with which all participants can interact afford more equal participation compared to interaction through a single mouse-and-keyboard, and that participants will find tangible models easier to understand and remember than virtual models.

Kim & Maher [18] showed that designers building a tangible model of an office perceived more spatial relationships, and re-framed the design problem more often, than designers building a virtual model. This suggests that tangible modelling results in models of higher quality than virtual models.

Horn et al. [13] compared tangible and graphical interfaces to exhibits in a science museum and found that people were more likely to interact with a tangible interface than a graphical interface, and that the tangible interactions lasted longer. Parmar et al. [24] compared group interaction of rural women with a health information system through an iconic keyboard interface versus an iconic tangible interface, and found that interaction through the tangible interface increased product engagement and social interaction, and improved community decision-making. Zuckerman & Gal-Oz [37] studied how stakeholders built a system dynamics model using a graphical user interface to a modelling tool, and using a tool consisting of abstract tangible signs, called FlowBlocks. Tangible modelling turned out to be slower than graphical modelling, but users reported higher levels of stimulation and enjoyment with tangible than with graphical modelling, deriving partly from the physical interaction with FlowBlocks. Grosskopf et al. [10, 20] experimented with a tool for building business process models with tangible abstract elements (plastic fiches with text drawn on them) and observed that participants spent more time on modelling, and achieved more understanding of the model, than with graphical process modelling, and reported more fun building the model. These results suggest that tangible models are likely to improve participation and collaboration.

The above studies compare tangible with virtual modelling. Bakker et al. [2] were the first to investigate iconicity. They compared iconic and abstract tangible game pieces on a smart tabletop that represented a map of the game, and found that subjects preferred the iconic pieces, as it afforded better understandability.

Two experiments by the authors, looking specifically at collaborative modelling tasks, are in partial agreement with the findings listed above. In what we will refer to in this paper as *Experiment 1* [17] two groups of modelling novices built models of the physical layout and IT architecture of a university campus using tangible iconic signs or virtual abstract signs, respectively. The tangible iconic group built a model twice as fast as the virtual abstract group, with a quality twice as good. Tool satisfaction was higher for the tangible iconic group as well. In what we will refer to in this paper as *Experiment 2* [16], eight groups of modelling novices collaboratively constructed enterprise models using either tangible abstract signs or a virtual abstract signs. Again, the tangible groups produced a model of higher quality. In both Experiments 1 and 2, the virtual groups divided tasks among themselves, whereas in the tangible groups all participants had the same task, indicating a higher amount of collaboration in the tangible groups. However, agreement about the model was only slightly higher for the tangible group in the Experiment 1, and slightly lower for tangible group in Experiment 2.

All of the experiments mentioned in this Section compare various combinations of tangibility and iconicity, and it is not clear whether the effects of these two variables have always been distinguished well. To distinguish these effects, we systematically analyze these these two variables in a factorial experiment (*Experiment 3*), reported in this paper. Based on our summary of related work, we formulate two hypotheses, to be tested in the experiment:

– H1: Iconicity improves understandability, and therefore the quality of the result, based on [1, 2, 15, 17].
– H2: Tangibility improves collaboration, and therefore task efficiency, based on [7, 10, 13, 16, 24, 20, 37].

4 Experiment design

Our object of study consists of groups of people collaboratively building a model of an existing system or of a new design. Our sample consists of groups of psychology students collaboratively building a model of their university's campus and then updating that model to represent their view of the campus' future.

The small, self-selected sample makes classical inferential statistics pointless. Instead, we use a technique known in case study research as *analytical induction*, where we treat the groups as cases, and combine evidence from cases analytically to explain and generalize about observed phenomena [36].

4.1 Treatment design

We are interested in two variables: tangibility and iconicity. Therefore, we have four treatment groups, each using different representations (signs) of the same modelling language, as shown in Table 1. The underlying language used in this experiment is a simplified version of the IRENE language for modelling smart cities [31], adapted to include elements specific to university campuses. The IRENE language is designed to be used in stakeholder workshops and is therefore intended to be usable by non-technical domain experts.

Table 1: The four toolsets

	Iconic	Abstract
Tangible	Whiteboard, magnetic objects, markers	Whiteboard, magnetic cards, markers
Virtual	MS Visio, icons, lines	MS Visio, boxes, lines

After filling in a demographics questionnaire[4], participants were randomly allocated to one of four groups. Each group was taken to a separate room and

[4] The questionnaires are available in full: `https://surfdrive.surf.nl/files/index.php/s/QzscwpSO6Xf02w0`

given one of the four toolsets described in Table 1, accompanied by a document describing the semantics and syntax of each modelling element. These textual descriptions were identical for all groups.

The groups were then asked to familiarize themselves with the toolset and the descriptions, requesting clarifications if needed (*Task 0*). Once each group declared that they understood the language, they were given their first modelling task: to build a model of the current campus, as accurately and completely as possible using the tools provided (*Task 1*). The groups could take as long as they like to build this model, after which participants received individual questionnaires[4].

Each group was then asked to update the model based on how they think the campus should evolve, staying within a fixed budget (*Task 2*). After Task 2, each participant received a final questionnaire[4] with the same questions as for task 1 but with an additional question on overall enjoyment.

During the two modelling tasks (Task 1 and Task 2) the students were allowed to ask factual questions about the campus but not about the modelling language. Each task was timed and, in addition, the two modelling tasks were videotaped.

4.2 Measurement design

To evaluate H1, we need to measure understandability and model quality, and to evaluate H2, we need to measure collaboration and task efficiency. There is no established set of indicators for the variables we want to measure. Fortunately, Hornbaek's survey of usability research [14] provides common indicators for quality of outcome, task efficiency and user satisfaction. We select the ones pertaining to collaborative modelling tasks, add the ones used in our previous experiments [17, 16] and describe them in Table 2 and below.

Language understandability is evaluated by measuring the *time* taken by each group to read and declare that they understand the language, and the *number of questions* they have during this time. In addition, we measured *perceived language understandability* and *learnability* of the language, as reported in the individual questionnaires distributed after each of the modelling tasks.

Quality of the outcome (i.e. of the resulting model) is one of the three usability factors listed by Hornbaek [14]. It is commonly measured as: **semantic quality** (how well the model represents the domain) and **syntactic quality** (how well the model adheres to the prescribed syntax) [19, 30]. We operationalize semantic quality as *incorrectness* (buildings represented in the model but not present on the campus)and *incompleteness* (buildings that are present on the campus but not represented in the model). Syntactic quality is operationalized as the number of *syntactic mistakes.* These quality measurements were performed first by two of the authors independently and then discussed.

Task efficiency, Hornbaek's second usability dimension, is operationalized in terms of *time taken to complete task,* as well as several indicators of perceived effort [14]: *perceived difficulty* and *perceived time* to complete task. Both are measured via questionnaires after each modelling task.

Table 2: Measured indicators

Dependent variable	Indicator	Scale
Language understandability	Effort to understand language (Task 0)	# minutes # questions
	Perceived language understandability (Task 1 and Task 2)	1 (Easy to understand) - 5 (Difficult to understand)
	Perceived language learnability (Task 1 and Task 2)	1 (Very easy) - 5 (Very hard)
Semantic quality	Incorrectness (Task 1)	$\|M - D\|$ (model elements not in domain)
	Incompleteness (Task 1)	$\|D-M\|$ (domain elements not in model)
Syntactic quality	Deviation from syntax (Task 1)	# syntactic mistakes
Task efficiency	Time to complete task (Task 1 and Task 2)	# minutes
	Perceived difficulty (Task 1 and Task 2)	1 (Very easy) - 5 (Very difficult)
	Perceived time to complete task (Task 1 and Task 2)	1 (Very little time) - 5 (Too long)
Satisfaction	Perceived tool satisfaction (Task 1 and Task 2)	1 (Very satisfied) - 5 (Very unsatisfied)
	Perceived enjoyment (Task 1+2)	1 (Boring) - 5 (Fun)
Collaboration	Amount of discussion (Task 1 and Task 2)	# words per minute # turns per minute Coefficient of variation for words per participant
	Perceived agreement (Task 1 and Task 2)	1 (Don't agree) - 5 (Fully agree)

Hornbaek's last usability dimension is **satisfaction**. We measured *perceived tool satisfaction* and *perceived enjoyment* via the same questionnaires.

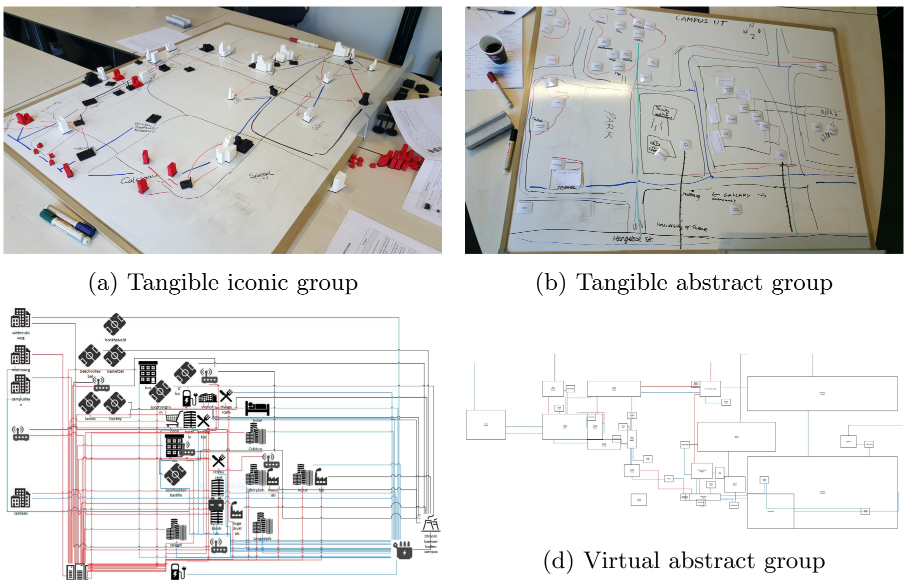
Collaboration has to do with the relative effort each participant expended in communicating with the others and resolving differences [14]. Furthermore, higher intra-group agreement is thought to be indicative of better collaboration in group problem-solving [4]. Therefore, we operationalize collaboration in terms of two indicators: amount of discussion and agreement. The *amount of discussion* is measured by (1) annotating the video recordings and aggregating the average number of words and turns per participant per minute for each group as an indicator of the individual contributions to the discussion [14, 21] and (2) by computing the coefficient of variation of words per participant as an indicator of how evenly the discussion was spread [6, 21]. Finally, the level of *perceived agreement* of each participant with their group's result is queried via questionnaires at the end Tasks 1 and 2.

5 Results and Analysis

Figure 1 shows the results of Task 1.

Table 3 contains averages of the measurements gathered via questionnaires and Table 4 contains averages of the measurements performed by us. Both are averaged per group.

We omit the results of Task 2 from this paper as this task was creative rather than descriptive in nature and therefore cannot be easily compared with previous experiments. The complete set of measurements and observations may be examined at `https://docs.google.com/spreadsheets/d/1AnTMelfLsQtGLhX136Z1Qaz50VSB3zmW36PMuBLESNg`.



(a) Tangible iconic group

(b) Tangible abstract group

(c) Virtual iconic group

(d) Virtual abstract group

Fig. 1: Models produced during Task 1

Table 3: Self-reported measurements (on task 1, unless otherwise specified)

	Difficulty	Time	Satisfaction	Learnability	Agreement	Understand-ability	Enjoyment (incl. Task 2)
TI	3.4	2.8	2.4	2.2	4.2	2	3.8
TA	3.4	3.4	3.4	2.8	3.2	3.2	4
VI	2.8	2.4	1.4	1.6	4.75	1.8	4.5
VA	3.4	2.8	2.4	2	4.6	2.2	3.6

Table 4: Objective measurements (on task 1, unless otherwise specified)

	Questions asked	Time (task 0)	Time (task 1)	Turns/ min.	Words/ min.	CV words/ part.	$\lvert M-D \rvert$	$\lvert D-M \rvert$	Syntactic mistakes
TI	0	3	33	18	105	*0.45*	2	0	1
TA	0	2	55	27	88	*0.27*	5	4	3
VI	2	3	40	34	107	*0.37*	4	3	0
VA	4	9	49	26	92	*0.67*	6	2	0

Language understandability The results show that iconicity improves understandability, and abstractness decreases it. This agrees with our expectation in H1 and can be explained by the theory of cognitive fit. The effect of tangibility on understandability is less clear: the two tangible groups reported lower understandability and learnability than the corresponding iconic groups, contradictory to the observations of Fitzmaurice[7] and Luebbe[20] that tangibility improves understandability. One explanation might be that the subjects of this experiment - psychology students - had little to no modelling experience and were unfamiliar with parts of the domain and therefore benefited from the syntactical constraints built into the software tool. This suggests that to predict the effect of tangibility on language understandability in future experiments, we may have to include the variables *experience with modelling and design* and *familiarity with the domain.*

The measurements also suggest an interaction between tangibility and iconicity: Tangible iconic signs are perceived as slightly less understandable than virtual iconic signs, but tangible abstract signs are perceived as considerably less understandable than virtual abstract signs.

Task efficiency Iconicity sped up the modelling process, and abstractness decreased it. This agrees with the theory of cognitive fit. Tangible modelling was perceived to be more difficult by our subjects than virtual modelling. This contradicts observations of earlier Experiments 1 and 2, where tangible modelling was perceived to be easier than virtual modelling [17, 16]. However, this contradiction may be explained by the flip side of our above explanation: Experiments 1 and 2 where done with students of computer science and technical management science, who have been trained in modelling and design, and felt more comfortable with the modeled domain, a smart campus and an enterprise architecture, respectively. Current measurement do not support nor rule out this explanation, and future experiments should include the variables *experience with modelling and design* and *familiarity with the domain* to test this explanation.

Tangibility magnified the effect of iconicity on task efficiency. Tangible iconic models were built faster than virtual iconic models, and tangible abstract models were built slower than virtual abstract models. This interaction may explain why in Experiment 2[17], the tangible iconic group built models twice as fast as the virtual abstract group. The interaction is also consistent with the observations of Zuckerman et al. [37] and of Lübbe [20] that tangible abstract groups took longer to build a model than a virtual abstract groups.

Quality of product Iconicity had no consistent effect on model quality but tangibility had a clear effect on syntactic quality: All tangible models contained syntactic mistakes, but none of the virtual ones did. A possible explanation for this is that our modeling tool only allows pre-defined elements and connectors, while using our tangible modeling tool arbitrary shapes could be drawn.

In terms of semantic quality however, we could not clearly separate the effects of tagibility and iconicity. Tangible iconic models were more complete and correct than any of the other models, an effect observed in both Experiments 1 and 2 [17] and in this latest, third experiment. However, tangible abstract models in this experiment had lower semantic quality than the corresponding virtual models, whereas in Experiment 2 [16] they were slightly better. A possible explanation of this apparent contradiction is the apparent uncertainty of the subjects involved in Experiment 3 in the face of the freedom afforded by our tangible modelling languages, combined with their lack of knowledge about the domain. This uncertainty may be aggravated by the preference of naive (non-technical) users to represent systems with iconic diagrams rather with abstract diagrams, compared with the preference of technical students trained in modelling and design languages, to represent systems in abstract diagrams [29].

Satisfaction Tool satisfaction was lowest for the tangible abstract group and highest for the virtual iconic group. In Experiment 2 [16] this was the reverse. This is consistent with our proposed explanations above that the subjects of our latest, third experiment were thrown off by freedom afforded by our tangible modelling tools, and find it more difficult to make abstract models than to make iconic models. Task enjoyment was highest for the virtual iconic group. This partly contradicts previous conclusions ([37, 17]), that tangible modelling is *always* more enjoyable than virtual modelling. Task enjoyment was highest for the virtual iconic group. We may again explain this in terms of lack of modelling experience and preference of naive users for guidance provided by virtual tools, and for iconic modelling tools.

Collaboration The iconic groups spoke more (words/min) than the corresponding abstract groups, despite similar or less turn-taking. Iconic groups also exhibited higher agreement. Tangibility promoted more equal participation (lower CV) for groups working with abstract signs but slightly less equal participation in iconic groups. However, closer analysis of the video revealed that the tangible iconic group contained a “silent” participant (see Figure 2) which drove up the CV. A clear effect of tangibility compared to virtuality in Experiment 3 as well as in Experiment 2[16] is that virtuality promotes task division, while tangibility does not.

This is easily explained by physical setup of the virtual modelling groups, which all used a single keyboard with mouse and confirmed by the number of words per participant (Fig. 2), which shows that the virtual groups were dominated by one or two participants. Analysis of the videos showed that these were the people grabbing the keyboard and/or mouse. Iconicity resulted in more perceived agreement than in abstract models. This effect is slight for virtual models, but is large for tangible models. So tangibility is a also a magnifier for this effect for iconic models.

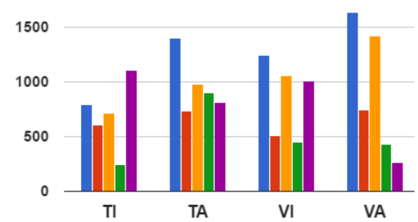


Fig. 2: Words per participant. Each bar represents a different participant.

6 Discussion

The positive effect of iconicity on understandability can be explained by the theory of cognitive fit. Improved understandability in turn explains higher modelling speed and higher agreement, assuming that modelers have similar conceptual models of the domain.

Tangibility magnifies these effects because tangible signs are graspable, allowing modelers to use their spatial reasoning skills [7]. This is in line with the cognitive theory of grounded conceptual knowledge [3]. In this view, human cognition does not initially (and primarily) develop by formal instruction but by the interactions between perception and action with which the human child explores its environment. Then, and in later life, repeated forms of these interactions result in the neural brain patterns on which conceptual knowledge, as used in language and reasoning, is based [32]. Combining graspability and iconicity would more easily reactivate these patterns, and hence the concepts they represent, thereby contributing to an increased and faster understanding.

In addition, all participants have equal access to tangible signs, and this can speed up modeling. However, this effect is moderated by other variables, such as whether the modelers are familiar with the domain, and may even be influenced by their personality - eg. willingness and ability to participate actively in a group - or higher order cognitive processes - such as relational reasoning and abstraction [35].

Note that these explanations hold even though the subjects Experiment 3 *perceived* tangible modelling as harder than virtual modelling. The increased freedom of tangible modelling made them feel uncertain. Humans tend to adapt their linguistic structure based on the context and their situational goal [12], so this can lead to more syntactical mistakes in tangible models. Still, even though they claimed to have no familiarity with the domain, the models of the iconic tangible group were semantically complete, and contained the least mistakes of the four experimental groups.

Validity Our sample of 20 volunteers was self-selected, and our four experimental groups were small. Therefore, we could not meaningfully use statistical inference so our reasoning is case-based, where we try to explain the results of a series of experiments in terms of existing theory in a process of analytical induction [26, 34]. However, as indicated at length by Znaniecki [36], who coined the term "analytical induction", this has been a common pattern of reasoning in the experimental physical sciences.

External validity is the support for our generalization to a wider population of stakeholders and domains. To this end, each the three experiments used participants from different domains (technical, management, social, respectively) and different modelling languages (TREsPASS, 4EM, IRENE, respectively). In order to maximize generalizeability, we focus on effects noticed in all three experiments to which we provide interpretations in terms of general cognitive theories which can be assumed to be valid for students and non-students alike.

7 Conclusions and Future Work

We conclude that our experiments provide support for H1 (Iconicity improves understandability), but that only provide partial support for H2 (Tangibility improves collaboration). Instead, iconicity turned out to improve group discussion and perceived agreement, which are indicators for collaboration. Tangibility mostly amplifies these effects of iconicity, but also supports more equal participation of group members than is possible with virtual models using a single mouse-and-keyboard setup.

In addition, we found that modeling experience of participants, familiarity with the domain, and personality of the participants moderate these effects. Future studies should control for these variables, as well as investigate whether other confounding factors are present in a real world setting.

The constructs outlined in this paper, and the relationships between them highlighted by hypotheses H1 and H2, together with observations of previous experiments provide support for constructing and updating theories related to group modelling of complex systems.

An open issue, that we have not investigated, is how to facilitate entering a tangible iconic model in a computer, once it has been built by a group of stakeholders. Luebbe [20] did this for tangible abstract models by camera, but for tangible iconic models intelligent tangible signs on a smart tabletop may be a more feasible option.

Another issue for future research is that complex, socio-technical systems consist of physical, social and virtual elements, not all of which can be represented in an iconic way. In these cases, models will likely consist of iconic as well as abstract signs, and the issues we have observed with abstract signs manipulated by nontechnical experts come into play.

References

1. Aversano, L., Canfora, G., Lucia, A.D., Stefanucci, S.: Understanding sql through iconic interfaces. In: Computer Software and Applications Conference, 2002. COMPSAC 2002. Proceedings. 26th Annual International. pp. 703–708 (2002)
2. Bakker, S., Vorstenbosch, D., van den Hoven, E., Hollemans, G., Bergman, T.: Tangible interaction in tabletop games: Studying iconic and symbolic play pieces. In: Proceedings of the International Conference on Advances in Computer Entertainment Technology. pp. 163–170. ACE '07, ACM, New York, NY, USA (2007)
3. Barasalou, L., Simmons, W., Barbey, A., Wilson, C.: Grounding conceptual knowledge in modality-specific systems. Trends in Cognitive Science 7(2), 84–91 (2003)
4. Barron, B.: When smart groups fail. The journal of the learning sciences 12(3), 307–359 (2003)
5. Bjeković, M., Proper, H.A., Sottet, J.S.: Embracing pragmatics. In: International Conference on Conceptual Modeling. pp. 431–444. Springer (2014)
6. Bordia, P.: Face-to-Face Versus Computer-Mediated CommunlGation:A Synthesis of the Experimental Literature. The Journal of Business Communication 34, 99–120 (1997)
7. Fitzmaurice, G.W., Ishii, H., Buxton, W.A.S.: Bricks: Laying the foundations for graspable user interfaces. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. pp. 442–449. CHI '95, ACM Press/Addison-Wesley Publishing Co., New York, NY, USA (1995)
8. Fleischmann, A., Schmidt, W., Stary, C.: Tangible or not tangible – a comparative study of interaction types for process modeling support. In: Proceedings of the 16th International Conference on Human-Computer Interaction (HCI), Part II: Advanced Interaction Modalities and Techniques. pp. 544–555. Springer (2014)
9. Frutiger, A.: Signs and symbols: their design and meaning. Van Nostrand Reinhold Company (1989)
10. Grosskopf, A., Edelman, J., Weske, M.: Tangible business process modeling - methodology and experiment design. In: 1st International Workshop on Empirical Research in Business Process Management (ER-BPM'09). pp. 53–64. Springer, Ulm, Germany (2009)
11. Heath, C., Coles-Kemp, L., Hall, P.: Logical lego? co-constructed perspectives on service design. In: Proceedings of NordDesign 2014. p. 416. Aalto Design Factory (2014)
12. Hoppenbrouwers, S., Wilmont, I.: Focused conceptualisation: framing questioning and answering in model-oriented dialogue games. In: IFIP Working Conference on The Practice of Enterprise Modeling. pp. 190–204. Springer (2010)
13. Horn, M.S., Solovey, E.T., Crouser, R.J., Jacob, R.J.: Comparing the use of tangible and graphical programming languages for informal science education. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. pp. 975–984. CHI '09, ACM, New York, NY, USA (2009)
14. Hornbæk, K.: Current practice in measuring usability: Challenges to usability studies and research. Int. J. Hum.-Comput. Stud. 64(2), 79–102 (Feb 2006)
15. Horton, W.K.: The ICON Book: Visual Symbols for Computer Systems and Documentation. John Wiley & Sons, Inc., New York, NY, USA, 1st edn. (1994)
16. Ionita, D., Kaidalova, J., Vasenev, A., Wieringa, R.: A study on tangible participative enterprise modelling. In: Proceedings of the 3rd International Workshop on Conceptual Modeling in Requirements and Business Analysis (MReBA). Springer (2016)

17. Ionita, D., Wieringa, R., Bullee, J.W., Vasenev, A.: Tangible modelling to elicit domain knowledge: An experiment and focus group. In: Proceedings of the 34th International Conference on Conceptual modelling (ER). pp. 558–565. Springer (2015), lNCS 9381
18. Kim, M., Maher, M.: The impact of tangible user interfaces on designers' spatial cognition. Design Studies 29, 222–253 (May 2008)
19. Lindland, O.I., Sindre, G., Solvberg, A.: Understanding quality in conceptual modeling. IEEE Software 11(2), 42–49 (March 1994)
20. Lübbe-Grosskopf, A.: Tangible Business Process Modeling: Design and Evaluation of a Process Model Elicitation Technique. Ph.D. thesis, Universität Potsdam (2011)
21. McQueen, R.J., Rayner, K., Kock, N.: Contribution by participants in face-to-face business meetings: Implications for collaborative technology. Journal of Systems and Information Technology 3(1), 15–34 (1999)
22. Miller, G.: The magical number seven, plus or minus two: Some limits on our capacity for processing information. The Psychological Review 63, 81–97 (1956)
23. Moody, D.: The "physics" of notations: Toward a scientific basis for constructing visual notations in software engineering. IEEE Transactions on Software Engineering 35(6), 756–779 (Nov 2009)
24. Parmar, V., Groeneveld, G., Jalote-Parmar, A., Keyson, D.: Tangible user interface for increasing social interaction among rural women. In: Proceedings of the 3rd International Conference on Tangible and Embedded Interaction. pp. 139–145. TEI '09, ACM, New York, NY, USA (2009)
25. Rettig, M.: Prototyping for tiny fingers. Commun. ACM 37(4), 21–27 (Apr 1994)
26. Robinson, W.: The logical structure of analytic induction. American Sociological Review 16(6), 812–818 (December 1951)
27. Short, T.L.: Peirce's theory of signs. Cambridge University Press (2007)
28. Sweller, J.: Cognitive load during problem solving: Effects on learning. Cognitive Science 12(2), 257–285 (1988)
29. Tabachneck-Schijf, H.J.M., Verpoorten, J.H., van de Weg, R.L.W., Wieringa, R.J.: The influence of conceptual user models on the creation and interpretation of diagrams representing reactive systems. In: Current Research in Information Sciences and Technologies. Multidisciplinary approaches to global information systems, Spain. pp. 452–456. Open Institute of Knowledge, Badajoz, Spain (2006)
30. Unhelkar, B.: Verification and Validation for Quality of UML 2.0 Models, chap. The Quality Strategy for UML, pp. 1–26. John Wiley & Sons, Inc. (2005)
31. Vasenev, A., Montoya, L., Ceccarelli, A.: A Hazus-based method for assessing robustness of electricity supply to critical smart grid consumers during flood events. In: International Conference on Availability, Reliability and Security (ARES). p. 6. IEEE (2016)
32. Velde, V.v.d.: Communication, concepts and grounding. Neural Networks 62, 112–117 (2015)
33. Vessey, I., Galleta, D.: Cognitive fit: An empirical study of information acquisition. Info. Sys. Research 2(1), 63–84 (Mar 1991)
34. Wieringa, R.: Design Science Methodology for Information Systems and Software Engineering. Springer (2014)
35. Wilmont, I., Hengeveld, S., Barendsen, E., Hoppenbrouwers, S.: Cognitive Mechanisms of Conceptual Modelling, pp. 74–87. Springer Berlin Heidelberg (2013)
36. Znaniecki, F.: The Method of Sociology. Octagon Books (1934)
37. Zuckerman, O., Gal-Oz, A.: To tui or not to tui: Evaluating performance and preference in tangible vs. graphical user interfaces. International Journal of Human-Computer Studies 71(78), 803 – 820 (2013)