Activity diagrams with location context: Experimental comparison of colour and icon annotations

Sundar Gopalakrishnan *sundar.gopalakrishnan@idi.ntnu.no*
Norwegian University of Science and Technology (NTNU)
Trondheim, Norway
Guttorm Sindre *guttorm.sindre@idi.ntnu.no*
Norwegian University of Science and Technology (NTNU)
Trondheim, Norway

Abstract

In mobile information systems, the location of the user when performing tasks may be important to take into account during development. Yet, mainstream process models seldom capture this aspect. In previous papers we have evaluated analytically a number of small adaptations for showing location of actions in UML activity diagrams. Two of the most promising adaptations have also been evaluated experimentally. However, another notation alternative that got a quite positive score in the analytical evaluation has not been experimentally evaluated, namely a notation attaching symbolic icons to the activity nodes. This paper reports on an experiment comparing this notation to the most promising one from previous experiments (colour). The results showed no significant difference between the two notations, neither in the quality of answers to the experimental task, the time spent performing the task, nor in opinions about the notation alternatives, as investigated in a post-task questionnaire.
Keywords: Mobile, multi-channel, information system, process model, diagram notation, visual communication.

1. Introduction

Business process modelling languages like BPMN [1] and UML activity diagrams [2] generally do not capture the location of the activities performed, nor location-related usage context. For traditional office information systems using desktop computers (or laptops placed on desktops), it is understandable that location or related context has not been given much attention. However, the location and context of activities performed is of much higher importance in mobile information systems [3]. For instance, imagine a project developing IT support for power line maintenance workers. It would then be likely to have quite an impact on the requirements and design of the solution whether a certain information processing task was supposed to be done in the office before going out on a repair job, while driving back or forth, while walking in the terrain, or while climbing in a power mast. If the task were to be performed on foot in a dense forest, this would require different equipment and imply other usability challenges than what a desktop application is normally faced with.

This makes it interesting to investigate specialized process notations that do show location-related information in diagrams. Of course it is not necessarily a good idea to cram all kinds of information into a diagram, rather the diagrams might show the most important information and the rest could be described in prose text or tables beside the diagrams. However, there are indications that users prefer diagrams over other forms of representation [4], and the required location or context for actions is likely to be relatively more important in a mobile information system than in a non-mobile system. Work on mobile ontologies by Veijalainen [5] supports the idea of the 'where' aspect as essential in mobile processes. This motivates our quest for ways of showing location in diagrams.

In previous work [6-10] we have mainly looked at small adaptations of UML Activity Diagrams. In [9] a wide range of alternative ways of indicating location was outlined, and the alternatives were evaluated analytically using Moody's 9 principles for visual notations [11]. In [7], we compared experimentally a notation using various colours for locations to a UML baseline of simply indicating the locations by textual notes attached to the nodes. The adapted notation using colour scored better both for participant performance and opinion. In [8], acknowledging that a notation with colour might give trouble to colour-blind users, we compared this to a similar notation using b/w pattern fills instead. In this experiment, there was no significant difference in performance between the two notations, but the participant preference was still in favour of the colour notation. Colour and pattern-fills were also two of those notations that scored best in the analytical evaluation [9]. However, another adaptation which scored quite well analytically was the usage of symbolic icons to indicate location of the actions. This is a notation we have not tried experimentally in previous works. The novelty of this study is thus an experimental comparison of an icon-based and colour-based adaptation for showing location in UML Activity Diagrams. The research questions are as follows:

RQ1: Will model users perform better using a notation with colour or with icons?

RQ2: Will there be any difference in user preference for the two notations?

The rest of the paper is structured as follows: Section 2 presents related work in the area. Section 3 shows the notation alternatives to be explored. Section 4 describes the experimental design, and section 5 reports on the results. Section 6 gives a discussion, including threats to validity, whereupon section 7 concludes the paper.

2. Related Work

It is useful to differentiate between space and place when discussion the modelling of location [12-13]. “Space” describes geometrical arrangements that might structure, constrain, and enable certain forms of movement and interaction; “place” denotes the ways in which settings acquire recognizable and persistent social meaning in the course of interaction. The notations explored in this study focus on 'place'. Work on modelling of space can be found in [14].

There have been several efforts presenting adaptations of diagram notations and trying them out experimentally. Mendling et al. [15] propose to insert small icons inside each business process activity, but not to indicate location, rather the nature of the activity, where the authors identify 25 generic labels. Some examples of concepts represented using such iconic labels are "assess" (weight scales), "complete" (check mark), "decide" (question mark), "promise" (handshake), and "search" (magnifying glass) [15] (p.52). In [16] we discuss minor adaptations to BPMN diagrams similar to what we had previously proposed for UML activity diagrams, but the BPMN proposal has not been evaluated experimentally. Reijers et al. [17] explore the usage of syntax highlighting in business process models.

Ottensooser et al. [18] compare the understandability of BPMN diagrams with textual use cases, finding that the former were only understandable to trained users, while the latter were understandable to a wider range of people. La Rosa et al. [19] investigate some visualization techniques to improve the understanding of BPMN models, including usage of colour and thicker lines. The purpose is however not to introduce new types of information into the model (like location) but rather to help the user focus on important patterns in the model. Usability tests of the modifications proved successful in experiments. Figl et al. [20] study the effectiveness of various symbol for routing (e.g., forks and joins) in process models. Hanandeh et al. [21] propose an adaptation of UML for mobile systems, but focussing on state diagrams, not activity diagrams.

Most closely related to this paper is work by Baumeister et al. [22], where the authors propose some extensions to UML activity diagrams specifically targeting the modelling of mobile systems. In particular, it distinguishes between what the authors call a responsibility centred notation, where swim-lanes are used to indicate "who" performs and action, and a location centred notation, where swim-lanes indicate "where" an action is performed. The responsibility centred notation uses textual labels in action or actor nodes to indicate location. The location centred notation similarly indicates the responsible actor by textually labelled actor objects. Since responsibility centred notation is the standard usage of swim-lanes, it is location centred notation which represents the major deviation from normal practice. This resembles a notation which was briefly tried but dismissed in [6] because the need for showing responsibility by other means created a lot of extra nodes and edges in the diagram, thus creating a messy diagram that would be hard to comprehend especially for stakeholders with limited technical skills.

Saleh and El-Morr [23] propose another modification of UML activity diagrams, M-UML, to support the modelling of mobile systems. The main contributions of the work are on the conceptual level, while the modification to the notation is fairly limited, e.g., tagging mobile activities with an M and remote activities with an R. I.e., there is no detailed indication of location in terms of different places that the mobile user might be located.

Decker [24] also proposes an adapted notation of UML activity diagrams, specifically targeting access control for mobile workflows. Each activity node can be linked to one or a series of location nodes, shown by parallelograms, to indicate that it is compulsory or prohibited to perform certain activities in certain locations. [25] makes a similar proposal for BPMN diagrams, again showing locations by icons with a parallelogram shape. Thus creating separate nodes for locations, this slightly resembles the usage of UML notes to indicate locations, as investigated analytically in [6] and experimentally in [7], however with poorer user performance than an alternative approach of coloured nodes.

A systematic review by de Oca et al. [26] includes references to a number of other works that have also explored the understandability of business process models in various ways. After Moody presented his "physics" of notations [11] there have also been several evaluations of the visual syntax of various modelling languages [27], such as UML [28], BPMN [29], and i* [30-31]. To our knowledge, the current paper is however the first experimental comparison of colour and icon adaptations for showing location in UML Activity Diagrams.

3. Notation alternatives and hypotheses

Bertin [32] presents 8 different visual variables that might convey meaning in two-dimensional diagrams. These are the two planar variables horizontal position and vertical position (an object could be placed above, below, left, right of another, or also inside or partly overlapping another object, to convey various meanings), and the six retinal variables: size, brightness (="value"), texture, colour, orientation, and shape.

As Bertin points out these can be used alone or in combination to give different meanings to diagrammatic constructs. With possibility to combine follows, of course, the potential for combinatorial explosion, so there is almost no end to the various notations that could be imagined, even for a modelling language with a quite limited number of concepts. This means that not everything can be tried experimentally and common sense and analytical evaluations must be used first to have some idea what is at all worth trying. Figure 1 shows some ways of indicating locations in activity nodes. Rows 1 and 2 use existing UML syntax, but 1 at the cost of getting long activity labels and 2 at the cost of getting an extra node for the note beside each activity. The latter compared unfavourably to a notation using colour (cf. row 3) in [7]. Row 4 illustrates the basic idea of the alternative notation to be pitted against colour in this paper, namely using symbolic icons inside the activity nodes to convey the place of action. Row 5 indicates a notation that would likely be a total failure, namely using the size of the activity node to indicate the place of action.

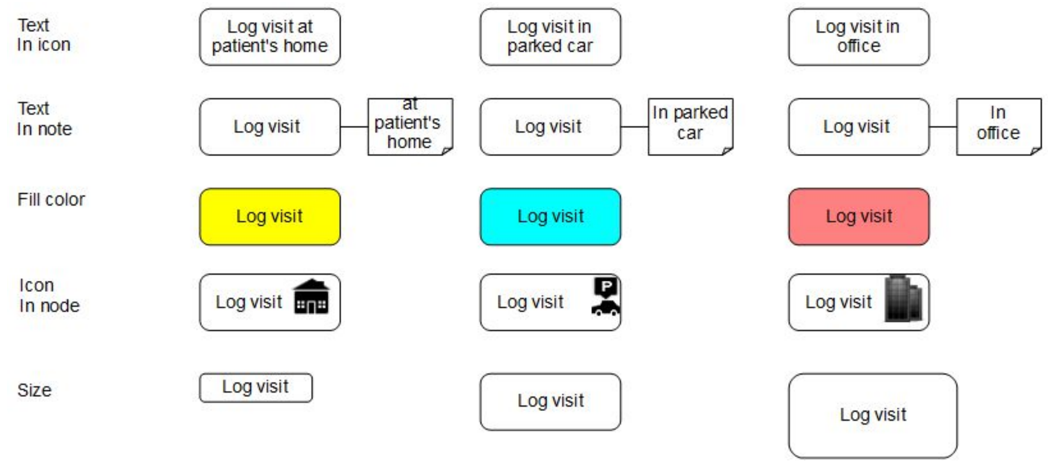

Fig. 1. Some possible ways of showing location with retinal variables. Third row (fill colour) have nodes in yellow, blue, red. The alternatives in this experiment are in rows 3 and 4.

Table 1 shows an excerpt of the analytical evaluation in [10]. That evaluation considered a bigger number of alternative notations (16), but Table 1 recapitulates the evaluation just for the five notations shown in Figure 1. Most of the criteria were based on those proposed by Moody [11]. For space reasons we cannot repeat the whole evaluation. However, it can be seen that row 5 (using size) scored extremely poorly. For instance it gets a double minus for ST (semantic transparency, one of Moody's criteria), this because changing size of nodes does not intuitively reflect different locations, people would rather think about different size, complexity or importance of activities if size was used with some meaning. Symbolic icons, on the other hand, get a double plus here: a user might intuitively understand that one activity is done in a house, another in a parked car, etc. Coloured nodes are midway between these extremes for the ST criterion, they do not intuitively hint at specific locations so the user is dependent on a legend to understand the meaning of each colour. However, colours are not misleading like size would be.

Table 22. Excerpt of analytical evaluation from [10].

Variable	SC	PD	ST	CM	VE	DC	GE	MD	Sum
Text in icon	-	-			- -	-	++	+	- 1
Text in note	-	-		- -	- -	+	++	++	0
Fill color	+	+		+	+		+		+4
Icon in node	+	++	++	+	-	-			+3
Size	+	- -	- -	+	+	-		-	- 4

All in all, colours came out better than the symbolic icons in the analytical evaluation presented in [10], but only marginally so (+4 vs. +3). Given the unavoidable subjectivity of such analytical evaluations, these results are so similar that it feels most reasonable to have hypotheses without any assumed direction in our experiment. Had we instead compared colour (or icons) with text in note, or especially with size, it would have been natural to assume the former would be better. Hence, our null hypotheses for this experiment were as follows:

- $H1_0$: There will be no difference in task score between participants using the fill colour notation alternative and participants using the icon in node alternative.
- $H2_0$: There will be no difference in time to finish the task between participants using the fill colour notation alternative and participants using the icon in node alternative.
- $H3_0$: There will be no difference in post-task opinion about the notation between participants using the fill colour alternative and participants using the icon in node alternative.

In the next section, the experiment design to investigate these hypotheses will be explained in more detail.

4. Experiment design

In modelling practice, a number of different work tasks may be related to models, such as making models, using models for discussion among stakeholders, understanding models, doing quality assurance of models, doing other work based on models (e.g., programming), etc. Since we had limited time available for our controlled experiment and preferred data that was fairly easy to analyze, we opted for an experiment task where the key challenge for the participants was to find errors in diagrams. Hence they were shown a somewhat messy prose text describing the users requirements for a system together with an activity diagram with locations added. Group A got a diagram where colour was used to indicate locations, while Group B got a diagram where icons were used for the same purpose, as shown in Figure 2.

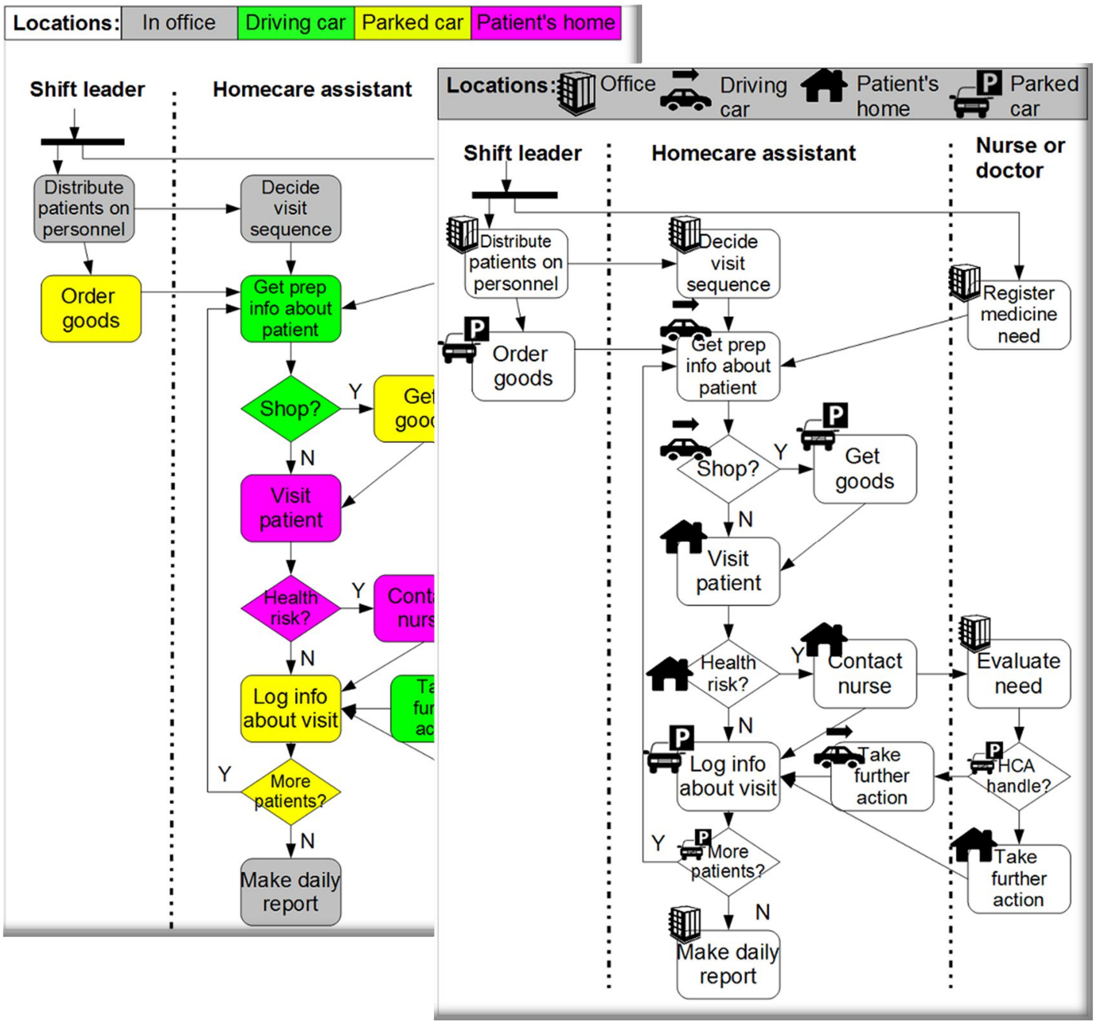

Fig. 2. Diagrams used in the experiment. Here put partly on top of each other to save space, as both were semantically the same.

A group of 47 students, all 3rd year students in a Bachelor program in Informatics, were randomly assigned to two treatment groups, resulting in 22 persons in Group A and 25 in Group B. Students performed the experiment online via the Learning Management System (LMS) that our university is currently using (*it's learning*), so the experiment questions were responded to in a questionnaire made as a web form in this LMS. As it was hard to find a time and place fitting for everybody, students were allowed to perform the experiment anywhere of their liking within a time slot of two days. This of course entails some risk due to reduced control of the participants (compared to gathering them all in one room). Birnbaum [33] discusses a number of problems that could result from studies done on the web rather than in a controlled room, such as multiple submissions, dropouts, recruitment issues, and response bias (the latter, e.g., related to user interface mechanisms used for responding). Of these, recruitment issues was not a problem in our case since the recruitment was not done directly over the web, but by contacting one particular class of Informatics students, who then did the experiment over the web. As for multiple submissions, we provided clear instructions that each person should do the test only once, and that each should do it individually - and also took care that there would be no incentives to deliver more answers (i.e., no individual bonus per answer, just the bulk reward for the entire class, as partial funding for their excursion). A further analysis of these and other threats will be made in the discussion of results towards the end of the paper.

The text they were asked to look at described stakeholder wishes for a mobile information system to support home care workers while visiting clients (typically elderly people who still leave at home but need some help, so the home care assistant will drive around visiting one after the other). It was stated in the experiment description that the text would be assumed correct, while the diagram might contain some errors. The diagrams had deliberately been seeded with some errors, typically location information which was inconsistent with what was described in the text. The task for the students was thus to identify which activities had the location correctly given, and which had errors. The students were asked to have a stopwatch running (e.g., on their mobile phone) while doing the test, so they could also report the time used on the questions. In total, the diagram had seventeen nodes, whereof wrong locations were given for four of them, while the other locations were right compared to the textual description

For each participant, the following variables could be measured:

Precision (PR): How many of the errors reported were really errors?

Recall (REC): How many of the errors did the participant find?

Time (T): How much time was used for the task? (This was self-reported by the participants in whole minutes)

Perceived Ease of Use (PEOU): Measured by a questionnaire administered just after the task. This same questionnaire also measured **Perceived Usefulness (PU)** and **Intention to Use (ITU)**.

This questionnaire was inspired by the Technology Acceptance Model (TAM) [34] and the related Method Evaluation Model (MEM) [35]. It contained 13 questions, where 5 related to PEOU, 5 for PU, and 3 for ITU. Table 2 shows the question for the colour group, the other group got identical questions except "colour" would be replaced by "icons". All questions were scored on a 5 point Likert scale from Strongly agree (5), via agree (4), neutral (3), disagree (2), and strongly disagree (1). Questions 4, 5, 6, 7, and 10 were formulated in the opposite direction of the others, these were reversed in the scoring.

Table 2. Questions for the post-task questionnaire.

1. The modified notation gave me a better understanding of where an activity or decision is performed.
2. I found it easy to master the modified activity diagram notation.
3. It was easy to find location errors in the diagrams with this notation.
4. It would have been easier to find errors if the location was given in textual notes beside each use case oval, rather than as colour.
5. The notation was hard to learn.
6. I was often confused about the meaning of colours for activities and decisions.
7. The addition of colour in the diagram made me less effective in performing the task.
8. If using activity diagrams in a future project where it is important to specify location of activities, I would consider using this notation.
9. The notation made me find location errors in the diagram more quickly than if locations had been given in the activity label (e.g., "Make daily report in office")
10. Colour in the diagram made it appear messy and hard to understand.
11. The notation was easy to remember.
12. If somebody else proposes to use this notation in a future project I am involved in and where location of activities is important, I would support that.
13. I would rather prefer to model the location of activities in some other way than by colour.

5. Results

The student responses were automatically exported from the LMS to Excel, and since the data were fairly simple to analyze, the statistical analysis was performed directly in Excel rather than moving to a dedicated statistics tool. Results for the performance of the two groups are shown in Table 3. As can be seen, on average the group using coloured activity diagrams scored slightly better than the group using activity diagrams with icons, having 0.05 higher precision and recall. However, this advantage was far too low to be significant. On the other hand, the students using activity diagrams with icons completed the task slightly faster than those getting the coloured ones (on average 1.4 minutes, i.e. 1 minute 24 seconds faster). Again, this was far from being significant (p=0.2). Significance was tested by t-tests after first running F-tests to check equality of variances.

Table 3. Performance of the two groups.

Compared variable	**Colour (N=22)**		**Icons (N=25)**		**Diffe-rence**	**Effect Size**	**Sign.? Y/N**
	Mean	**SD**	**Mean**	**SD**			
Precision	0.63	0.20	0.58	0.24	0.05	0.19	No
Recall	0.78	0.25	0.73	0.31	0.05	0.19	No
Time	16.45	6.26	15.04	5.25	-1.41	0.23	No

The time spent and the point score might somehow be related, so we also tested the correlation between these two. These correlations were positive but small: 0.17 for the Colour group (i.e., those who spent more time on the task tended to do somewhat better than those who spent less time, but not markedly so), and practically zero (0.006) for the Icons group. Hence the slightly shorter time spent by this group on average does not seem to explain the slightly poorer task score.

Table 4 shows the results from the post-task questionnaire. As can be seen, the colour variant was scored higher for Perceived Usefulness, with an effect size of 0.27 (Cohen's d), but this was not enough to be significant (p=0.08). For PEOU and ITU, results were almost equal, actually with the icons group just marginally higher on average. Since the score range was from 1 (strongly negative) to 5 (strongly positive), the overall average opinions of 3.73 and 3.63 can be said to be quite positive for both notations (3.0 would be completely neutral), though not overwhelmingly positive.

Table 4. Perceptions of the two groups.

Compared variable	**Colour (N=21)**		**Icons (N=25)**		**Diffe-rence**	**Effect Size**	**Significant? Y/N (p-value)**
	Mean	**SD**	**Mean**	**SD**			
PEOU	3.76	0.66	3.78	0.76	-0.02	0.03	No
PU	3.82	0.61	3.55	0.71	0.27	0.42	No (p=0.15)
ITU	3.49	0.63	3.53	0.62	-0.04	0.06	No
Overall	3.73	0.55	3.63	0.62	0.10	0.16	No

6. Discussion and conclusion

It could be considered somewhat disappointing that the experiment did not produce any significant result. Both for answer score, time, and preference, there were no significant differences, so null hypotheses could not be rejected. Of course, one possible interpretation of this is that the two notations investigated are about equally good. This does not have to be a bad thing − a modelling tool might then offer different notations based on user preference (e.g., a colour-blind user would prefer to have the locations indicated by icons rather than by colour) and none would be a poorer alternative than the other.

However, it could also be that the experiment was not well designed, so that it failed to test the two notation alternatives sufficiently. The task proved to be somewhat easier than we had intended, the students on average spending just a little over a quarter of an hour on the main task of the experiment (though the four slowest students, two in each group, spent half an hour), yet scoring quite well. So, this could mean that the experimental task failed to challenge the participants enough, thus they did well regardless of notation alternative and the result was no significant difference.

Threats to Validity

Internal validity: were observed effects due to the different in treatment, or something else? Since we did not observe any effect, there is little to say here, but as speculated above, it might be that we could have observed an effect if the tasks had been more challenging. The fact that time spent on each sub-task was self-reported by the students could also be seen as less reliable than having times measures by the Learning Management System. The reason for this choice was primarily that the Learning Management System in use at our university did not support the timing of tasks within a test. However, even measured tests − and even with the experiment performed in a more controlled setting − could be unreliable, e.g., some students working effectively with the task all the time, while others might lose concentration along the way and drift into other thoughts. For internal validity, however, the challenge would be if reduced reliability from self-reporting caused any bias between the two treatment groups. Since groups were randomly selected, there is no particular reason why one group should have been less accurate than the other one in reporting time spent. Related to some of the threats for online experiments discussed by Birnbaum [33] we found no sign of answer copying or multiple submissions, and there were no drop-outs during the test (i.e., no half-finished responses). Also, while there was a two-day window during which students could do the experiment (so that each could find a suitable time), all participants performed the tasks within an hour once they got started, so this did not cause a lot of variation in duration of exposure to the tasks. There could of course still be other differences impacting the results, for instance some students performing the tasks late in the evening when they were tired, while others did them at an optimal time for performance. But again, similar variation in fitness might also be the case with an experiment in a restricted one hour time-slot (e.g., some participants having slept poorly or been to a party the night before, others being fully rested). Moreover, there is no reason why this should should have hurt the performance of one random group more than the other.

Conclusion validity: Is there sufficient statistical power to back the claims? Again, with no claims made, the only interesting question in this regard is if we missed any possible observation by having too few participants? However, most comparisons were far from showing significant differences, the one which came closest, was Perceived Usefulness with $p=0.15$ (two-tailed). While it could be speculated that we might have observed a difference for this variable if the experiment had a much bigger number of participants, it seems intuitively more reasonable to interpret this difference as due to arbitrary variation in questionnaire responses, especially since the related PEOU and ITU does not show a difference in the same direction.

Construct validity: Do our measures correctly represent what we wanted to measure? We checked the participants' ability to identify errors in a model, which both requires them to understand the model and be able to compare it with some other document (in this case text).

This would seem a relevant work task, but of course not the only work task related to models. A broader experiment might also have tested the ability to develop models, discuss them with non-technical stakeholders, remember modelling languages over time, etc. A main weakness of the experiment reported here is thus that it tested only one type of task (error detection) with only a limited set of variables (precision and recall, plus time spent and the participants' opinion about the modelling language). Also, all the errors that were seeded into the models were related to the location of activities, since that was the adaptation that we wanted to test. However, in retrospect we realize that it might have been a good idea to have other errors in the diagrams too, e.g., logical errors related to branching and sequencing, as well as swim-lane placement of activities. This to check not only how good the two adaptations were for conveying locations, but also to check if the adaptations might have adverse effects on the understanding of other aspects of the diagram.

External validity: Our experiment participants were all 3rd year IT students. These are of course not representative of the full range of possible model users in a real-world IS development project. There are two main challenges: First, students may be less competent than professionals in IS development / business process modelling. Second, the students may be more competent (in understanding UML diagrams) than stakeholders whose education did not include any such modelling. For the concern about students vs. practitioners [36] several sources indicate that this need not be the most serious problem [37-39]. Since our research questions have the term "model users", not "practitioners", both students and practitioners are groups of people who use models for various purposes (e.g., learning, work). Moreover, students in the final year of their study program (like these 3rd year Bachelor students were, although some of them might continue further master studies rather than going into the workforce) would not necessarily be so different from novice practitioners. Indeed, the fact that all participants were IT students is perhaps worse concerning external validity, and it would be interesting to redo the experiment with participants with a much wider range of educational backgrounds. Yet, it must be said that the purpose of this particular experiment was not to investigate the understandability of UML Activity Diagrams as such, only the understandability of the *adaptations* (colour or icon) related to mobility. In that context, it made sense to use participants who were already familiar with standard UML diagrams, which was the case for these students.

The homogeneity of participants is not the only problem for external validity. Another problem is that tasks in controlled experiments are much shorter and simpler than work-life tasks, so that several challenges pertaining to complexity, communication and collaboration are not addressed. Also, experimental participants (whether they are students or practitioners) may not have the same motivation as workers, unless the experiment has some kind of incentive. However, in this particular experiment, the latter does not seem to have been so much of a problem, as the students did quite well, but the former seems to have been a crucial weakness, since the difference in performance between the strongest and weakest participants was fairly small.

Considerations for future work

If pursuing future work in investigating the virtues of adding location to activity diagrams in general, and comparing the usage of colours versus icons in particular, the following observations can be made: (1) The experimental tasks should be more varied, and more challenging. In the current experiment, the UML activity diagram was fairly small (17 nodes) and did not contain much complicated logic. We attempted to make the task challenging by writing the natural language text in a somewhat messy way, but apparently were not successful enough at this. Moreover, it could be questioned whether messing up the text is the right way to go. Probably it should rather be the diagrammatic model which is more complex, since it is the quality of the notation we want to test, not the students' reading abilities. (2) The diagram did not do enough to "stress" the two notation alternatives. There was only one location given for each activity, and the total number of different locations to be captured in the diagram was only four. Increasing the total number of different locations in the model might soon have caused problems, especially for the colour notation (except for the colour-blind, most people will easily distinguish between some few primary colours, but if needing a bigger number − say 15 − requiring the usage of several intermediary colours, the Perceptive Discriminability (one criterion of [11]) will sure deteriorate. Hence, if we had come up with a case with a large number of different locations, we might have found some advantage for icons that we failed to find now. On the other hand, if looking at a large zoomed-out model, colours would still be observable, while icons would be too small to be differentiated, so big models with fairly few different locations could give advantage for colour.

All in all, therefore, any future experiments with these two notational adaptations should involve a wider range of participants than the current one, and try a variety of tasks, with big models, small models, more complex models, and observe the performance of bigger tasks over some time such as in [40]. Also, it would be interesting to try several possible locations per node, although this was done in a similar experiment for use case diagrams without finding significant differences between colour and icons [41]. Moreover, even if controlled experiments can address much more complex tasks than what was done in this particular experiment, there is of course a limit to what can be tried in a controlled setting, which will still fall short of some work-tasks of industrial complexity. It is therefore important also to do case studies with a longer duration, e.g. where teams collaborate about models in fairly large projects. Moreover, adaptations of BPMN might be more interesting than UML activity diagrams, depending on usage statistics for these languages, but UML activity diagrams was chosen for this experiments in order to complete a series of previous experiments that had looked at other adaptations of UML activity diagrams.

Another question that might be explored in further research is concerning tool support. Any notational adaptation, whether it be colour or icons, and whether they are used to indicate location or some other aspect of the context of process activities, need not be something that you either have or do not have. In a tool the user could switch them on and off at will, more like a highlighting function. Hence, for future investigations a more dynamic usage of such adaptations could be more interesting than just statically adapted diagrams.

References

1. Wohed, P., van der Aalst, W. M. P., Dumas, M., ter Hofstede, A. H. M., Russell, N., On the Suitability of BPMN for Business Process Modelling. In: Dustdar, S., Fiadeiro, J., Sheth, A. (eds.) Business Process Management, pp. 161-176. Springer Berlin / Heidelberg (2006)
2. Dumas, M., ter Hofstede, A., UML Activity Diagrams as a Workflow Specification Language. In: Gogolla, M., Kobryn, C. (eds.) «UML» 2001 — The Unified Modeling Language. Modeling Languages, Concepts, and Tools, pp. 76-90. Springer Berlin / Heidelberg (2001)
3. Walderhaug, S., Stav, E., Mikalsen, M., Experiences from Model-Driven Development of Homecare Services: UML Profiles and Domain Models. In: Chaudron, M.R.V. (ed.) Model-Based Design of Trustworthy Health Information Systems (MOTHIS), workshop at MODELS'08, pp. 199-212. 2008, Springer Berlin / Heidelberg (2008)
4. Figl, K., Recker, J., Exploring cognitive style and task-specific preferences for process representations. Requir. Eng. 21(1), 63-85 (2016)
5. Veijalainen, J., Developing Mobile Ontologies; who, why, where, and how? In: Becker, C. et al. (eds.) International Conference on Mobile Data Management, pp. 398-401. IEEE (2007)
6. Gopalakrishnan, S., Sindre, G., Alternative Process Notations for Mobile Information Systems. In Popplewell, K., et al. (eds.): Proc. Int'l conference on Interoperability for Enterprise Software Applications (I-ESA'10), Coventry, UK, 13-15 Apr, (2010)
7. Gopalakrishnan, S., Krogstie, J., Sindre, G., Adapting UML Activity Diagrams for Mobile Work Process Modelling: Experimental comparison of two notation alternatives. In van Bommel, P. et al. (Eds.): Proc. 3rd IFIP WG8.1 Working Conf on

the Practice of Enterprise Modelling (PoEM2010), Delft, Netherlands, 9-10 Nov, 2010. Springer, LNBIP 68, pp. 145-161, (2010)
8. Gopalakrishnan, S., Krogstie, J., Sindre, G., Adapted UML Activity Diagrams for Mobile Work Processes: Experimental comparison of colour and pattern fills. In Halpin, T. et al. (Eds.): Proc. EMMSAD 2011, London, UK, 20-21 Jun, 2011. Springer, LNBIP vol. 81, pp. 314-331 (2011)
9. Gopalakrishnan, S., Sindre, G., Diagram Notations for Mobile Work Processes. In Johannesson, P., et al. (Eds.): The Practice of Enterprise Modelling (Proc. PoEM'11), Oslo, Norway, 2-3 Nov, 2011. Springer, LNBIP vol. 92, pp. 52-66 (2011)
10. Gopalakrishnan, S., Krogstie, J., Sindre, G., Capturing Location in Process Models: Comparing Small Adaptations of Mainstream Notation, Int J. of Info Syst Modeling and Design (IJISMD), 3(3):24-45, July-September (2012)
11. Moody, D.L., The "Physics" of Notations: Toward a Scientific Basis for Constructing Visual Notations in Software Engineering. IEEE T. Software Eng. 35(6), 756-779 (2009)
12. Dourish, P., Re-space-ing place: "place" and "space" ten years on. In: Hinds, P., Martin, D. (eds.) Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work, pp. 299-308. ACM (2006)
13. Harrison, S., Dourish, P., Re-place-ing space: the roles of place and space in collaborative systems. In: Ehrlich, K., Schmandt, C. (eds.) Proceedings of the 1996 ACM conference on Computer supported cooperative work, pp. 67-76. ACM (1996)
14. Nossum, A., Krogstie, J., Integrated Quality of Models and Quality of Maps. In: Halpin, T. et al. (eds.) Enterprise, Business-Process and Information Systems Modeling, pp. 264-276. Springer, Berlin/Heidelberg (2009)
15. Mendling, J., Recker, J., Reijers, H.A., On the Usage of Labels and Icons in Business Process Models. Int. J. Info. Syst. Mod. Design 1(2), 40-58 (2010)
16. Sindre, G., Krogstie, J., Gopalakrishnan, S.: Visually Capturing Usage Context in BPMN by Small Adaptations of Diagram Notation, Proc. 16th Working Conference on Exploring Modelling Methods for Systems Analysis and Design (EMMSAD'13), 17-18 Jun 2013, Valencia, Spain, Springer LNBIP vol. 147, pp. 324-338 (2013)
17. Reijers, H.A., Freytag, T., Mendling, J., Eckleder. A., Syntax highlighting in business process models. Decis. Support Syst. 51(3) 339-349 (2011)
18. Ottensooser, A., Fekete, A., Reijers, H.A., Mendling, J., Menictas, C., Making sense of business process descriptions: An experimental comparison of graphical and textual notations. J. Syst. Softw. 85(3) 596-606 (2012)
19. Rosa, M.L., ter Hofstede, A. H. M., Wohed, P., Reijers, H. A., Mendling, J., van der Aalst, W. M. P., Managing process model complexity via concrete syntax modifications. IEEE T. Ind. Inform. 7(2) 255-265 (2011)
20. Figl, K., Recker, J., Mendling, J., A study on the effects of routing symbol design on process model comprehension. Decis. Support Syst. 54(2) 1104-1118 (2013)
21. Hanandeh, F.A., Alsmadi, I., Al-Shannag, M.Y., Al-Daoud, E., Mobile agents modelling using UML. Int. J. Business Inform. Syst. 19(4) 419-432 (2015)
22. Baumeister, H., Koch, N., Kosiuczenko, P., Wirsing, M., Extending Activity Diagrams to Model Mobile Systems. In: Aksit, M., Mezini, M., Unland, R. (eds.) Objects, Components, Architectures, Services, and Applications for a Networked World, pp. 278-293. Springer Berlin / Heidelberg (2003)
23. Saleh, K., El-Morr, C., M-UML: an extension to UML for the modeling of mobile agent-based software systems. Inform. Software Tech. 46(4) 219-227 (2004)
24. Decker, M., Modelling Location-Aware Access Control Constraints for Mobile Workflows with UML Activity Diagrams. In: Boyeon, S., et al. (eds.) 3rd International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies. IARIA (2009).
25. Decker, M., Che, H., Oberweis, A., Stürzel, P., Vogel, M., Modeling Mobile Workflows with BPMN. In: Giaglis, G.M., Zwass, V., Chau, P., Mylonopoulos, N.

(eds.) 2010 Ninth International Conference on Mobile Business and 2010 Ninth Global Mobility Roundtable (ICMB-GMR), pp. 272-279. IEEE (2010)
26. de Oca, I.M.-M., Snoeck, M., Reijers, H.A., Rodríguez-Morffi, A., A systematic literature review of studies on business process modeling quality. Inform. Software Tech., 2015. 58: 187-205.
27. Caire, P., Genon, N., Heymans, P., Moody, D.L., Visual notation design 2.0: Towards user comprehensible requirements engineering notations. In: Gotel, O. (ed.) 21st IEEE International Requirements Engineering Conference (RE'13), pp. 115-124, IEEE (2013)
28. Moody, D., van Hillegersberg, J., Evaluating the visual syntax of UML: An analysis of the cognitive effectiveness of the UML family of diagrams. In: van den Brand, M., Gasevic, D., Gray, J. (eds.) Proceedings of Second International Conference on Software Language Engineering, pp. 16-34. Springer, Berlin / Heidelberg (2009)
29. Genon, N., Heymans, P., Amyot, D., Analysing the Cognitive Effectiveness of the BPMN 2.0 Visual Notation. In: Malloy, B., Staab, S., van den Brand, M. (eds.) Proceedings of Fourth International Conference on Software Language Engineering, pp. 377-396. Springer, Berlin / Heidelberg (2011)
30. Moody, D.L., Heymans, P., Matulevicius, R., Visual syntax does matter: improving the cognitive effectiveness of the i* visual notation. Requir. Eng. 15(2) 141-175 (2010)
31. Moody, D.L., Heymans, P., Matulevicius, R., Improving the Effectiveness of Visual Representations in Requirements Engineering: An Evaluation of i* Visual Syntax. In: Ryan, K. (ed.) 17th IEEE International Conference on Requirements Engineering, pp. 171-180. IEEE (2009)
32. Bertin, J., Semiology of Graphics: Diagrams, Networks, Maps. Madison, WI: University of Wisconsin Press (1983)
33. Birnbaum, M.H., Human Research and Data Collection via the Internet. Annu. Rev. Psychol. 55: 803-832 (2004)
34. Davis, F.D., Bagozzi, R.P., Warshaw, P.R., User Acceptance of Computer Technology: A Comparison of Two Theoretical Models. Manage. Sci., 35(8) 982-1003 (1989)
35. Moody, D.L., Dealing with Complexity: A Practical Method for Representing Large Entity Relationship Models (PhD Thesis), Dept Of Information Systems, University of Melbourne (2001)
36. Daun, M., Salmon, A., Bandyszak, T., Weyer, T., Common Threats and Mitigation Strategies in Requirements Engineering Experiments with Student Participants. In Requirements Engineering: Foundation for Software Quality, pp. 269-285, Springer, Berlin / Heidelberg, (2016)
37. Berander, P., Using students as subjects in requirements prioritization. In: Juristo, N., Shull, F., Zelkowitz, M. (eds.) Proc. of 2004 ACM-IEEE Int'l Symposium on Empirical Software Engineering, pp. 167-176. IEEE (2004)
38. Runeson, P., Using Students as Experiment Subjects – An Analysis on Graduate and Freshmen Student Data. In: Linkman, S. (ed.) Proc. 7th Int'l Conf. on Empirical Assessment & Evaluation in Software Engineering (EASE'03), pp. 95-102. Keele University, Staffordshire, UK (2003)
39. Carver, J., Jaccheri, L., Morasca, S., Shull, F., Issues in using students in empirical studies in software engineering education. In: Berry, M., Harrison, W. (eds.) Proc. of Ninth Int'l Software Metrics Symposium, pp. 239-249, IEEE (2003)
40. Pinggera, J., et al., Styles in business process modeling: an exploration and a model. Softw. Syst. Model., 14(3) 1055-1080 (2015)
41. Gopalakrishnan, S., Sindre, G., Use case diagrams for mobile and multi-channel information systems: Experimental comparison of colour and icon annotations, Proc. 21st Working Conference on Exploring Modelling Methods for Systems Analysis and Design (EMMSAD'16), 13-14 Jun 2016, Ljubljana, Slovenia, Springer LNBIP, vol. 248, pp. 479-493, (2016)