Representing subset relations with tree diagrams or unit squares?
Katharina Böcherer-Linder, Andreas Eichler, Markus Vogel

To cite this version:
Katharina Böcherer-Linder, Andreas Eichler, Markus Vogel. Representing subset relations with tree diagrams or unit squares?. CERME 10, Feb 2017, Dublin, Ireland. hal-01950553

HAL Id: hal-01950553
https://hal.science/hal-01950553
Submitted on 10 Dec 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers. L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Representing subset relations with tree diagrams or unit squares?

Katharina Böcherer-Linder¹, Andreas Eichler² and Markus Vogel³

¹ University of Education Freiburg, Germany; katharina.boechererlinder@ph-freiburg.de
² University of Kassel, Germany; eichler@mathematik.uni-kassel.de
³ University of Education Heidelberg; vogel@ph-heidelberg.de

In this paper, we refer to the efficiency of different visualizations for mathematical problem solving. Particularly, we investigate how set relations that are potentially important in probability are made transparent by two different visualizations, i.e. the tree diagram and the unit square. In this paper, we use these two visualizations as representations of statistical information. First, we analyze theoretically the quality of visualizing set relations by tree diagrams and unit squares. Second, we briefly report a published study with students of mathematics education (n = 148) where the unit square outperformed the tree diagram when the perception of subset relations was regarded. A main focus of this paper is a replication of the aforementioned study with n = 58 undergraduate students. Finally, we discuss the significance of our results, specifically for the teaching and learning of conditional probabilities.

Keywords: Representation of statistical information, set relations, tree diagram, unit square.

Introduction

In mathematics education, it is widely accepted that representations and visualizations could have a considerable impact on students’ learning. For example, Duval claims that visualizations of mathematical concepts are “at the core of understanding in mathematics” (Duval, 2002, p. 312). However, research in mathematics education and cognitive psychology gave evidence that visualization does not necessarily foster students’ understanding. For this reason, a crucial question in research in mathematics education is to identify the one of potentially different visualizations that is most efficient referring to students’ learning. We refer to this question concerning competing visualizations aiming to support students’ learning in statistics and probability, but also in fractions. The transparency of set relations plays a crucial role in probability (Böcherer-Linder & Eichler, 2017), but is also important in many other domains of mathematics education, such as the teaching and learning of fractions.

Theoretical framework

In the context of Bayesian reasoning, research in cognitive psychology has shown that the transparency of set relations in the visualization of statistical information impacts on the performance in tasks concerning the Bayes’ rule: “any manipulation that increases the transparency of the nested-sets relation should increase correct responding” (Sloman, Over, Slovak, & Stibel, 2003, p. 302).
Proponents of this point of view, called *nested-sets account*, attribute the difficulties of Bayesian reasoning to the fact that some sets of events are nested (Lesage, Navarrete, & Neys, 2013; Sloman et al., 2003). For the example of a medical diagnosis, Figure 1 illustrates the nested-sets situation. “Transparency of set relations” means that it is easy to see, how many elements are in the sets and how the sets relate. The cognitive model into which the *nested-sets account* has been incorporated is the dual process theory (Barbey & Sloman, 2007): Representing the statistical information in a standard probability format (without visualization) obscures the nested-sets structure of the problem and, therefore, triggers the associative system which may lead to biases. Representing the statistical information with natural frequencies and / or appropriate visualizations in contrast trigger the rule-based system because nested sets relations are made transparent, enabling people to reason consciously and according to the logic of set inclusion (Barbey & Sloman, 2007). For the design of effective visualizations, proponents of the *nested-sets account* claim that visualizations are helpful to the extent that they make the nested set structure of the problem transparent (Barbey & Sloman, 2007; Sloman et al., 2003). There are different competing visualizations that claim to visualize efficiently set relations or situations that necessitates applying Bayes’ rule and it is an open question which visualizations are the most efficient and which properties explain these visualizations’ efficiencies.

In this paper, we investigate how set relations are made transparent by two competing different visualizations, i.e. the tree diagram and the unit square. We use these two visualizations as representations of statistical information (Venn diagrams are not considered in this paper because we focus on the visualization of statistical information and Venn diagrams are pure set representations but not representations of statistical information). First, we analyze theoretically the quality of visualizing set relations by tree diagrams and unit squares. Second, we briefly refer to a published study (Böcherer-Linder & Eichler, 2017; n = 148 undergraduate students) where we investigate whether the tree diagram or the unit square is more efficient to support the perception of subset relations. Since we used a new approach to explain the effectiveness of visualizations of the Bayes’ rule, we conducted a replication study (n = 58 undergraduate students) which results are in the main focus of this paper. Finally, we discuss the significance of our results, specifically for the teaching and learning of conditional probabilities.
Visualizing set relations

“A flower girl is selling red and white roses and carnations.” We use this situation as an example to illustrate how the tree diagram and the unit square visualize set relations. In this situation, we have sets (for example the set of all roses) and subsets (for example the subset of all red roses) and subset relations (for example the red flowers among the roses). If we attribute some numerical values to the number of roses and carnations, the situation can be visualized by showing absolute numbers in the tree diagram and the unit square:

![Tree Diagram and Unit Square](image)

**Figure 2: Representing statistical information with the tree diagram and the unit square**

Both, the tree diagram and the unit square can be seen as nested-sets representations. In the tree diagram, the logical relations between sets and subsets are visualized by lines. The subsets are on a lower level than the sets in the tree and the branches connect the subsets with the sets. For example the subset “red roses” is on a lower level than the set “roses” and the branch connecting “red roses” and “roses” visualizes the relation between both sets. The tree implies a hierarchical structure which means that subsets are always on a lower level than sets. Therefore, only those subset relations that are in line with the hierarchy are salient. For example, the subset relation of “roses among the red flowers” where the “roses” are the subset and all the “red flowers” are the including set is not visualized by a branch in the tree and thus, is not transparent.

In the unit square, subset relations are visualized by areas being embedded in other areas. For example the subset of “white roses” is represented by a partial area of the rectangle that represents all roses. In contrast to the tree diagram, the unit square implies no hierarchy. That means that subset relations can be grasped vertically (e.g. “white roses among the roses”) as well as horizontally (e.g. “roses among the red flowers”). Therefore, all subset relations that are possible in this situation are transparent in the unit square.

Because of these differences in the properties of the two visualizations, we expected a difference between the tree diagram and the unit square when the perception of different subset relations is regarded. Therefore we hypothesized:

If the subset-relation is not in line with the hierarchy of the tree diagram, the unit square is more efficient to make the subset-relation transparent (*hypothesis 1*). If the subset-relation is in line with
the hierarchy of the tree diagram, there is no significant difference between the unit square and the tree diagram (*hypothesis 2*).

**Method**

The method in our first study and the replication study was the same. In the first study, we administered a questionnaire to 148 students who were enrolled in a course of mathematics education. In the second study, the test was administered to 58 students who were also enrolled in a course of mathematics education. In both studies, we asked the students (among other questions concerning conditional probabilities) in one task that we indicate below to calculate proportions and to indicate the result in form of fractions:

\[
\text{proportion} = \frac{\#\text{subset}}{\#\text{set}}
\]

In this way, we could analyze if the right subsets and right sets have been grasped from the visualization. The questionnaire had two versions, one showing tree diagrams, the other showing unit squares to represent the statistical information. The rest of the test-items remained constant and the participants were randomly assigned to one of the two groups. Thus, any potential difference in the results could directly be attributed to the influence of the visualizations.

To assess the influence of representation on the perception of subset relations we designed test-items that each addressed structurally different subset relations. In Figure 3, we show the questions that were accompanied by either the tree diagram or the unit square shown above. Note that the item (d) addresses a subset relation that is not in line with the hierarchy of the tree diagram and therefore a higher performance for the unit square was expected. The items a, b, c and e address subset relations that are in line with the hierarchy of the tree diagram and therefore no significant difference between the tree diagram and the unit square was expected. We rated correct answers with 1 and incorrect answers with 0.

Flowers: A flower girl is selling red and white roses and carnations. Altogether, she has 120 flowers. Calculate the following proportions. Indicate the results in form of fractions.

The proportion of

a) red carnations among all carnations.

b) white roses among all flowers.

c) white flowers among all flowers.

d) carnations among the red flowers.

e) roses among all flowers.

**Figure 3: Items to assess the perception of subset relations**

**Results**

Figure 4 shows on the left side the results that we reported in Böcherer-Linder and Eichler (2017). As we hypothesized in this study, the unit square \((M = 0.66, SD = 0.44)\) was more efficient than the tree diagram \((M = 0.38, SD = 0.48)\) for the item (d) that addressed a subset relation that is not in line with the hierarchy of the tree diagram. The difference for the item (d) was significant \((t (146) = 3.579,\)
$p < .001$) with an effect size of $d = .58$. In the replication study there is again a significant difference referring the item (d) that addressed a subset relation that is not in line with the hierarchy of the tree diagram (unit square: $M = 0.63$, $SD = 0.43$; tree: $M = 0.31$, $SD = 0.39$; $t(54) = 3.101$, $p<0.01$, $d = .82$).

Thus, we replicated our result referring hypothesis 1. Further, it is interesting that the ratios of correct answers are very similar in the original study and the replication study. However, there are also unexpected differences for the items (b) and (c).

In our first study, we investigated also the differences in the other four items that addressed subset relations within the hierarchy of the tree diagram. For these items a $t$-test for the accumulated score referring to these four items ($\alpha = .739$) yielded no significant difference between the tree diagram ($M = 3.46$, $SD = 1.023$) and the unit square ($M = 3.46$, $SD = 1.036$), $t(146) = 0.000$, $p = 1.000$. In the same way none of the items yielded a significant difference between the unit square and the tree diagram when investigated individually. Thus, there was no reason to reject our hypothesis 2 in our first study. However, we could not replicate these results in the second study. When the items were regarded individually, a $t$-test yielded significant differences for item (b) ($p<0.05$) and (c) ($p<0.01$). Also, the accumulated score referring to all four items ($\alpha = .634$) yielded a significant difference between the tree diagram ($M = 3.01$, $SD = 1.11$) and the unit square ($M = 3.67$, $SD = 0.52$), $t(54) = 2.762$, $p = 0.008$.

In the first study, the mean values of correct answers for the tree diagram were almost equal for all of the four items a, b, c and e addressing subset relations that were in line with the hierarchy of the tree diagram (88%, 85%, 86%, 86%). In contrast, in the second study for every item differences appeared. However, the performance for item (d) was lower for both visualizations. This might indicate that the subset relation (d) is more difficult to perceive than the other subset relations and that the visualization with the unit square is more helpful in this case.
Discussion of the results

For the situation of the flower girl, our results show a very clear effect in favor of the unit square. This was the case in the first study and, with very similar results, in our replication study. Nevertheless, we suggest for future research to prove this effect also for other contexts. In another study with 143 students of electrical engineering, we replicated the effect for subset relations that are not in line with the hierarchy of the tree diagram for two more different contexts. It is further desirable to investigate the effect of those subset relations used in the items a, b, c and e in more depth. This is especially the case since our results for this kind of items seems to be ambiguous. The replication study yielded significant differences referring to the efficiency of the tree diagram and the unit square in supporting students solutions in tasks where the subset relations are in line with the hierarchy of the tree diagram. Although this result was not expected and we further hypothesize that this result will not be replicated in further studies, it agrees with our overall hypothesis, i.e. the supremacy of the unit square to visualize situations in which the Bayes’ rule has to be applied.

There are further aspects that that could be investigated in more detail. For example, there is the question of the order in the sequence in the tree diagram. It might be interesting to study the effect of the transposed order (roses / carnation on a lower level than red / white) and to compare it with a rotated unit square (roses / carnation arranged vertically and red / white arranged horizontally). This setting could be clarified if the hierarchy of the tree actually is the reason for the results in our study.

Moreover for the context of Bayesian reasoning, the results of Binder et al. (2015, p.6) suggest an advantage of the 2×2-table compared to the tree diagram, although no statistical difference between 2×2-tables and tree diagrams was reported. Thus, it is an open question if 2×2-tables are equally efficient than unit squares to make subset relations transparent or if there is an additional effect of the unit square due to the redundant geometrical and numerical representation.

Finally, there are further possibilities for visualizing set relations. One of these possibilities that was used in mathematics teaching is the double tree (Wassner, 2004). Thus, it could be interesting if a specific version of the tree diagram is able to decrease the weakness of the tree diagram to identify appropriately set relations.

Implications

The main result of our research seems nearly trivial: Visualizations have to visualize the main aspects of a mathematical concept if they aim to support students’ understanding of this concept. Accordingly, a subset relation must be transparent when the aim of the visualization is to represent subset relations. However, it is by no means at all trivial to identify the crucial aspects of a mathematical concept. Actually, the tree diagram is very prominent in statistics education research (Veaux, Velleman, & Bock, 2012) and also psychological research (Binder, Krauss, & Bruckmaier, 2015) for visualizing Bayesian situation that necessitates applying Bayes’ rule. However, our research gave evidence that – compared to the unit square - the tree diagram is not efficient to visualize the set relation that is crucial to understand the structure of a Bayesian situation since it requires a subset relation that is not in line with the hierarchy of the tree diagram.

Our results have firstly some consequences if statistics education is regarded. Sloman et al. (2003) expressed that bringing out nested set structure has been identified as being important for the improvement in Bayesian reasoning tasks. Thus, restricted to teaching and learning probability, our
results imply to reconsider the role of the tree diagram to support students’ learning referring to probability and Bayesian reasoning. This would be a considerable shift in statistics education (c.f. e.g. (Gigerenzer, 2014; Wassner & Martignon, 2002). A little bit more globally, it could be considered if proportions, and in particular proportions of proportions could be appropriately visualized by a unit square to emphasize the connection between proportions of proportions and conditional probabilities. Thus, the unit square could potentially be understood as a visual connection between fractions and probabilities.

More generally, our results imply to focus the discussion of visualizations on the structure of visualization and on its relation to the structure of the represented mathematical concept. While the superiority of visualizations is a consensus in mathematics education as we outlined in the introduction, it is a crucial objective to find out which visualization best fits to a mathematical concept, especially in situations where several competing visualizations exist as it is the case for Bayesian reasoning situations. For example, Binder et al. (2015) show that the tree diagram supports Bayesian reasoning compared to pure symbolic representation, whereas our results imply that the required subset relation is not transparent in the tree diagram. Indeed, in recent research, the unit square outperformed the tree diagram in Bayesian reasoning tasks (Böcherer-Linder & Eichler, 2017; Böcherer-Linder, Eichler & Vogel, in press). Therefore an ongoing task of educational research should be to precisely identify the relation of a visualization and its structure and the mathematical concept and its structure. One main message of our paper is that this relation is not sufficiently investigated, but could considerably impact on students’ learning.

**References**


