Digitizing Spatial Reasoning Experiment In Finding Potential Confounding Factors

Bram Pramono (7978359)
Xinhao Lan (1082620)

April 2022

Abstract

An experiment in spatial reasoning discovered that the similarity of relational structures influences the interpretation of spatial representation. That experiment had implications for analogical reasoning, diagrammatic reasoning, and language processing. We designed a digitized experiment to verify the result. Our result showed that there might be confounding factors to the original experiment, since our data differed to the original experiment.

1 Introduction

Throughout the evolution of human being, one of the key factor to our survival as a species is our ability to communicate our thoughts and exchange ideas with each other. There are two types of communications verbal and non-verbal. In verbal communication, the meanings acquired during the information exchange depend on our ability to relate words to their meaning. On the other hand, most non-verbal communication doesn't have such things that behave as words. In the studies regarding spatial reasoning, spatial representations (e.g. images, gestures, and diagrams) obtain their meaning during communication through analogical reasoning (Gattis, 2004; Cooperrider et al., 2016). In this paper, we will focus on this particular phenomenon of meaning acquirement and human reasoning when presented with spatial representations.

In our research, we wanted to explore and understand the main aspects of spatial reasoning. To do this, we chose a published work by Gattis (2004). In her research, she hypothesized that people map conceptual objects to spatial objects and conceptual relations to spatial relations. The result of her study showed that her hypothesis was indeed valid. However, as it has been extensively explained in Vasishth et al. (2018), studies in psychology could end up having significant results just by chance. Therefore, in this study, we would like to make a slight adjustments to the conducted experiment and test whether the results of our study will still follow the original hypothesis.

This paper is structured as most of empirical research papers. In section 2, the theoretical and concept background regarding spatial reasoning is explained more in detail. Since this paper replicated a published work, the details of the original study are highlighted in section 3. This section contains also the modifications we have made to the original study. The section 4 is reserved for presenting the results of the original study and ours. Following the results, we are putting forth the discussions regarding our finding. Lastly, we close this paper small words of conclusion.

2 Background

2.1 Spatial Reasoning

Spatial reasoning is considered as the core of our ability to solve problems involving spatial information, for examples in navigation and perspective taking (Schultheis and Carlson, 2013). When planning and following a route to the airport from one's house, relying only on the memory of previous journeys would not allow detour or mistake during the trip to the airport. It is necessary to solve such problems during navigation using reasoning about current position and the directions to the airport from an unidentified location. In perspective taking, people often need to imagine situations where they need to take decisions in their planning. Take an example of going to the movie with friends, choosing the best seat with the best sound and angle to watch the screen requires some spatial imagination of the object positions. To find the best seat, a reasoning process is necessary to make a judgement of the choice.

Aside from reasoning to solve problems, spatial reasoning is also often related to be a representation to pass on meaning, which is also called *spatial analogies*. With spatial analogies, people

can express different sort of information using similar spatial representations. In a study done by Cooperrider et al. (2016), people were given a short lesson of complex systems, such as stock market bubbles, and requested to explain what they have learned. The study showed that people produced spatial gestures when trying to describe complex relational structures even without the presence of spatial language in the lesson. This phenomenon brought forth the question of how spatial reasoning influence reasoning. In this paper, we are addressing this particular question by looking into the study by Gattis (2004) where she focused on the sub-question of how spatial representations acquire conceptual meaning. We are going to discuss this particular question next.

2.2 Spatial Representations and Meaning Acquirement

In her paper Gattis (2004), three core aspects that influence spatial reasoning were discussed. The first two of these aspects, namely *iconicity* and *associations*, were based on studies in the domain of semiotics, psychology, and linguistics. Iconicity has been considered to influence the reasoning process of spatial representations due to the physical resemblance of the items represented. It means that spatial representations such as maps acquire their meaning due to the forms or colors of previously known maps. However, this aspect does not seem to consider situations where the items are novel and yet to be identified or mapped to a particular concept in cognition.

The associations aspect solves this problem found in iconicity. It basically consider past experiences of objects and concepts into account when finding correspondences for an item to be identified. Associations occurs also often in language use when people are presented with analogies and metaphorical expressions. The word *more* can be associated with *up* and *feeling low* can be associated with sad moods. Associations use not only the characteristics of the items, but also te similarities between known concepts and the novel items. Though these two aspects can answer a wide range of questions regarding meaning acquirement of spatial representations, it cannot explain how adults and children without prior experience in Cartesian line graphs would consider steeper lines as faster rates of change and vice versa. People have presumably different past experiences. All these differences would lead to differences when applying associations. However, the interpretations of such spatial representations resulted in consistent outcomes.

To fill in this gap, the third aspect was introduced. The last aspect started as a hypothesis that is called the *Relational Structure Hypothesis*, which indicates that the assignment of meaning to spatial representations are influenced by two kinds of mapping, namely conceptual objects to spatial objects and conceptual relations to spatial relations. The motivation to posit such a hypothesis was grounded on the studies in analogical reasoning. In those studies, it was found that children often use objects as comparisons to explain other objects and make relational comparison as well. Examples of objects to objects mapping would be sentences such as "A tiger is like a zebra" and "Tree bark is like skin", while the example of relational mapping is "A tire is a shoe" (Gattis, 2004). These findings indicated that through analogical reasoning the attributes between objects or attributes between relations are mapped. The challenge to test such a hypothesis would be to have stimuli that are very unique and have never been encountered earlier by the participants. The following section describes in detail how such stimuli are designed.

3 Method

3.1 The Original Experiments

In the original paper, there were three experiments in total using the between-subject design. The researcher designed three conditions for each experiment, namely Subject-varying (SV), Object-varying (OV), and Relation-varying (RV). During the trial, participants must choose the sentence they think is the correct one out of two sentences. One of these sentences correspond to the Object-mapping while the other to the Relation-mapping. The experiments were all done using booklets of clipped papers with printed images and texts accompanying these images. All experiments were done during psychology classes, in which these booklets were distributed manually. It was not specified in detail how it was timed, but it was said that the participants completed the experiments between one and two minutes.

3.1.1 The First Experiment

The first experiment contained two training phases and a single test phase. The goal of this experiment was to find whether the mapping of relational structure influences the interpretation of diagrams paired with active declarative statements.

Training Phase 1
In phase 1, the participants saw two pictures along with sentences that described the meaning. On each picture, a person put forward their left or right hand. The format of the sentence for each picture was "This hand means (Animal1 or Animal2).". An example can be seen in Fig. 1 where MOUSE is mapped to Animal1 while BEAR to Animal2. This first phase is also called the initial mapping of animals to hands. In the experiment, four animals were randomized by rotating the list and mapped to the codes named Animal1 up to Animal4. These codes were then used to randomize the sentences in the next phase.

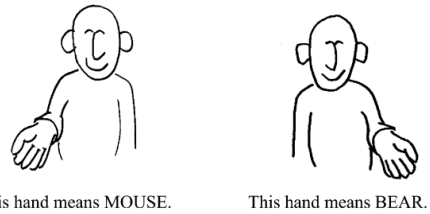

Figure 1: Training phase 1 of the first experiment (Gattis, 2004)

Training Phase 2
Participants were then shown two other drawings. The first drawing was "Right hand touches right ear", which indicated ipsilateral position of arm to body, and the second drawing was "Right hand touches left ear", which indicated contralateral position of arm to body. Below each drawing was a sentence, and the two sentences differed in either the subject (SV), the object (OV), or the action performed (RV). This between-subjects variable was the main experimental manipulation. For the SV condition, the two sentences were of the form, "This means 'Animal3 R-action Animal1' " and "This means 'Animal4 R-action Animal1.' " The relation (R-action) was either "visits" or "bites" and was counterbalanced between subjects. For the OV condition, the two sentences were of the form, "This means 'Animal1 R-action Animal3' " and "This means 'Animal1 R-action Animal4,' " and again the relation was either "visits" or "bites" and was counterbalanced across subjects. For the RV condition, the two sentences were of the form, "This means 'Animal1 R-action1 Animal3' " and "This means 'Animal1 R-action2 Animal3.' " As with the SV condition, the relations were "visits" and "bites," and the order of these two relations was counterbalanced between subjects (Fig. 2).

Subject	This means MONKEY BITES MOUSE.	This means ELEPHANT BITES MOUSE.
Object	This means MOUSE BITES MONKEY.	This means MOUSE BITES ELEPHANT.
Relation	This means MOUSE BITES MONKEY.	This means MOUSE VISITS MONKEY.

Figure 2: Training phase 2 of the first experiment (Gattis, 2004)

Test Phase
Participants were shown another drawing and needed to choose one out of two sentences that fits the picture best. The drawing is "Left hand touches left ear". Above the drawing was the question, "What does this mean?", and below the drawing were two sentences, with the instruction, "Circle the answer that fits best." For the SV condition, the two sentences were of the form, "This means 'Animal3 R-action Animal2'" or "This means 'Animal4 R-action Animal2.' " For the OV condition, the two sentences were of the form, "This means 'Animal2 R-action Animal3' or "This means 'Animal2 R-action Animal4.' " For the RV condition, the two sentences were of the form, "This means 'Animal2 R-action1 Animal3' or "This means 'Animal2 R-action2 Animal3.' " For each condition, the order of the two sentences was counterbalanced between subjects (Fig. 3).

3.1.2 The Second Experiment

The goal of this experiment was to find whether the mapping of relational structure occurs in a variety of syntactical contexts by pairing the drawings with conjunctive and disjunctive statements using the coordinating conjunctions *and* and *or*, respectively. The procedure was nearly identical to Experiment 1. For the three experimental conditions, the first animal in the statement varied for

What does this mean?

Subject	MONKEY BITES BEAR	or	ELEPHANT BITES BEAR
Object	BEAR BITES MONKEY	or	BEAR BITES ELEPHANT
Relation	BEAR BITES MONKEY	or	BEAR VISITS MONKEY

Figure 3: Test phase of the first experiment (Gattis, 2004)

SV, the second animal in the statement varied for OV, and the conjunction between them varied in RV condition. The order of these two relations was counterbalanced between subjects (Fig .4).

Subject	This means MONKEY AND MOUSE.	This means ELEPHANT AND MOUSE.
Object	This means MOUSE AND MONKEY.	This means MOUSE AND ELEPHANT.
Relation	This means MOUSE AND MONKEY.	This means MOUSE OR MONKEY.

What does this mean?

Subject	MONKEY AND BEAR	or	ELEPHANT AND BEAR
Object	BEAR AND MONKEY	or	BEAR AND ELEPHANT
Relation	BEAR AND MONKEY	or	BEAR OR MONKEY

Figure 4: Procedure of the second experiment (Gattis, 2004)

3.1.3 The Third Experiment

As for the last experiment, the goal was to find whether the mapping of relational structures depends on the explicit initial mapping. This experiment was very similar to experiment 1, the only difference was the exclusion of the first training phase in experiment 1.

3.2 Current Study

3.2.1 Modifications to The Original

We have mentioned earlier that we were replicating the study by Gattis (2004). Even though we wanted to conduct the same experiments, we had limited time and participant candidates to run our experiment on. To make sure that the power of our statistical result would come close to the original study, we made some modifications to the experiment items and conditions. That being said, we decided to run only a single experiment based on the third original experiment. There were in total four differences between our modified design and the original experiment.

Digitized form

In the original experiment, they used booklets and survey in a paper form. This form of experiment is harder to measure and control. It is harder to measure because we cannot know how long each phase of the experiment took. It is also harder to control due to the freedom people have in taking actions during the experiment. For example, people can go back and forth through the booklets before filling in the survey or choosing an answer, which might effect the end results. To improve the measurements and control of the experiment, we implemented our experiment using PCIbex (Zehr and Schwarz, 2018). By using this program, the flow of the experiment could be defined more precisely. The time of each action taken by the participant was recorded, which would give us more information about participant behaviours.

Animations instead of static images
In the original experiment, they used static images to represent gestures. However, in our experiment, we recorded videos and transformed them into animations to represent gestures. We believed that movements would give better representations of relations rather than simple ipsilateral or contralateral in static images.

Modified items
The original method considered all counterbalanced situations and recruited 194 participants for the third experiment alone. It would have been very difficult for us to achieve such a high number of participants in the short amount of time available to us. Some items randomization were applied to our experiment without having to apply the counterbalancing of random coded animals as the original experiment would have done. The reason for the counterbalancing decision had to do with our lack of experience in using PCIbex, which held us back from finding an optimal way to counterbalance the items as the original experiment. Along with this change, we also modified the verbs used in the declarative sentences from *bites* to *hugs* to soften the tone of the sentences and *visits* to *calls* just because *calls* is a more common verbs compared to *visits*.

Combined conditions
The limited number of participant candidates had also forced us to reduce the number of conditions to two. In the original paper, SV and OV were also combined in the result computation. Therefore, we believed that it should suffice to have only OV as the object-mapping condition. We chose OV instead of SV due to the commonness of Subject-Verb-Object grammar in the languages of our participant candidates. With these reduced conditions, we should have a better statistical power that would come close to the original study.

3.2.2 Procedure

In PCIbex, we made introduction, demographic profile, training, pre-test, test, and closing pages. In the demographic profile page, we asked the participants to fill in their age, gender, first language or mother tongue (L1), and their willingness to share the result of the experiment. Our experiment contained two phases, namely a training phase and a test phase. In the training phase, two animations were shown one after another with the accompanying declarative sentences. One content of the animations was "a person facing front, then slowly moving the left hand to touch the right ear", which we coded as *lh-re*. The second content of the animation was "a person facing front, then slowly moving the left hand to touch the left ear", which was coded as *lh-le*. In the OV condition, lh-re was accompanied with the text "Mouse hugs Elephant" and "Mouse hugs Monkey" for the lh-le. For the RV condition, the text was "Mouse calls Monkey" for lh-re and "Mouse hugs Monkey" for lh-le.

When the program entered the training page, the order of appearance for the animations was randomized. For each animation, the accompanying text was shown immediately, but not the button to continue. This button would appear after three seconds had passed since the animation was shown. Participants would need to press this button to show the next animation or continue to the pre-test page when no other animation were left. The pre-test page was added to notify people to be prepared for the test. In this page, a continue button needed to be pressed to start the test.

Once the participants entered the test phase, an animation would be shown and after three seconds had passed, two sentences would appear in which the participants must make their choice. Since we wanted to make sure that the stimuli to be as unique as possible, we had only a single trial for each participant. Unlike the original experiment, which allowed for two different images to have the same sentences as the answers. After the choice had been made, the participants would be thanked for their participation. The animation was selected from two animations, which were different from the ones out of the earlier phase. One content of these animations was "a person facing front, then slowly moving the right hand to touch the right ear", which was coded as *rh-re*. As you might have guessed, the content of the other animation was "a person facing front, then slowly moving the right hand to touch the left ear", which we coded as *rh-le*. In the OV condition, rh-re had two possible declarative sentences "Bear hugs Elephant" and "Bear hugs Monkey". As for rh-le, it had "Bear hugs Monkey" and "Bear hugs Elephant", which were the same sentences as rh-re in reversed order. A similar structure was applied to the RV condition. In RV, rh-re had two possible sentences "Bear hugs Monkey" and "Bear calls Monkey", while rh-le had "Bear calls Monkey" and "Bear hugs Monkey".

4 Result

In this section, we decided to only discuss the result of the third experiment from the original paper in detail. The results of the other two experiments are explained briefly. In the original experiments, data having the same sentence for the two different test images were removed. The usable data of the three experiments became 180, 178, and 190, respectively. For all three experiments, there were approximately 2/3 of the data pointing to the expected mappings following the hypothesis. In another words, the SV and OV conditions each had more than 60 participants selecting the sentence in the category Object-mapping and the other way around for RV condition where more than 60 participants selected the Relation-mapping sentence.

The result for the experiment 3 can be seen in Table 1. Even though it was not explicitly mentioned, we believed that the researcher calculated the significance of the selections distribution using the chi-squared test, because the data contained only the frequencies of selection in the two categories. From the 190 data points in experiment 3, the mapping pattern varied significantly between conditions, $\chi^2(2, N = 190) = 10.51$, $p < 0.01$.(Table 1)

Condition	Percentage of participants choosing a mapping		Number of participants
	Object-mapping	Relation-mapping	
Subject-varying	61.9	38.1	63
Object-varying	60.9	39.1	64
Relation-varying	36.5	63.5	63

Table 1: Result of original experiment (Gattis, 2004)

In our experiment, we managed to collect data from 87 participants, in which 44 participants for the RV condition and 43 for the OV condition. Since we only have a single trial for each participant, we did not need to clean up the data based on the answer. However, there were five participants who did not give the proper consent. These data points were not usable and were removed from our analysis. This left us with data from 82 participants. By applying the same kind of test as the original experiment, we found that the mapping pattern did not vary significantly between the two conditions, $\chi^2(2, N = 82) = 2.00$, $p = 0.157$.(Table 2)

Condition	Percentage of participants choosing a mapping		Number of participants
	Object-mapping	Relation-mapping	
Object-varying	76.3 (Number: 29)	23.7 (Number: 9)	38
Relation-varying	68.2 (Number: 30)	31.8 (Number: 14)	44

Table 2: Result of modification experiment

Considering the result of the modified experiment was different from the original experiment. We decide to also analyze our results based on the participants' age, gender, or native language. For the age factor, the result was almost the same as the overall result (Table 3).

Condition	Object-mapping	Relation-mapping	Number of participants
Object-varying(young)	78.1 (Number: 25)	21.9 (Number: 7)	32
Object-varying(old)	66.7 (Number: 4)	33.3 (Number: 2)	6
Relation-varying(young)	68.2 (Number: 30)	31.8 (Number: 14)	44
Relation-varying(old)	0 (Number: 0)	0 (Number: 0)	0

Table 3: Result of modification experiment based on age factor

For the gender factor, we deleted excluded two data points the one from the OV condition and the other from the RV. These two data points had neither female nor male as gender. The result had some differences compared to the previous result as in Table 4. Under the relation-varying condition, almost all female participants chose the sentence we did not anticipate. The chi-square test on the data of the female participants gave the following results $\chi^2(2, N = 46) = 4.93$, $p = 0.026$.

Condition	Object-mapping	Relation-mapping	Number of participants
Object-varying(male)	81.8 (Number: 18)	18.2 (Number: 4)	22
Object-varying(female)	66.7 (Number: 10)	33.3 (Number: 5)	15
Relation-varying(male)	45.8 (Number: 11)	54.2 (Number: 13)	24
Relation-varying(female)	94.7 (Number: 18)	5.3 (Number: 1)	19

Table 4: Result of modification experiment based on gender factor

For the native language factor, almost all of the participants in the RV condition spoke other European language, which we believed to be Dutch, and Chinese for the OV condition. We believed that this would not be a representative comparison. Therefore, we decided to exclude the analysis using first language.

5 Discussion

In this section, we are comparing the original experiment 3 with our modified experiment and found some differences in the results. In the original experiment, the participants chose the object-mapping sentence under the object-varying condition and the relation-mapping sentence under the relation-varying condition. However, in the modified experiment, the participant selections were concentrated on the object-mapping sentences for both conditions. In our experiment, under the situation of no initial mapping, the expected mapping of the relational structure hypothesis did not influence the interpretation of diagrams paired with active declarative statements.

Since the result of our experiment differed to the original one, we looked deeper into the data to find possible confounding factors. As mentioned in the previous section, we first suspected that the age or gender of our participants would be the influencing factors. It turned out that only gender indicated a significant difference. From the results, we considered that gender to be a potential confounding factor to the overall result we had.

Moreover, we communicated with some participants after the experiment and found that most people could not guess what the purpose of the experiment was. We also understood that some participants only observed changes in the touched ears in different animations, ignoring the movements of the arms and the ears. This could mean that movements were not considered as relations or that the focus when analyzing differences in animations would lie more on simpler variations such as object differences. Another possibility to explain the difference in the results would be the time factor. The original experiment was conducted using booklets with all participants completing the experiment during the same time span, which gave different individuals more time to analyze while the booklets were still being distributed to the rest of the participants. While in our experiment, the time available was less due to the forward-only action. By the time the participants reached the test phase, they might feel the need to quickly finish the experiment and just chose the simplest difference. This is like "human always don't waste extra effort doing things that could be done with less effort".

As mentioned earlier in section 3.2.2, the order of the sentences between object-mapping and relation-mapping for both conditions were fixed due to our lack of PCIbex knowledge. With the fixed order of appearance, the distribution of the selection from the two sentences would have led to a single position of the sentence, which we would expect to be the left one. However, as we have seen in our results that this was not the case. As for the RV condition, the first sentence would be the one indicating relation-mapping. Yet, most participants selected the second sentence for the object-mapping. This showed that this factor is not a confounding one.

There were possible issues mentioned in the original experiment, which we were confronted with in the modified experiment. They regarded the limitation of the generalizability of the evidence. The first issue concerned with the clarity of the task. Some participants might feel that the task did not provide a basis for an informed response, which would have led them to make a guess rather than a proper deduction in choosing the sentence. The second issue concerned with more fundamental cognitive processes. This problem was that the experiment might not test exclusively the relational structure hypothesis, but also include iconicity and associations during the reasoning process. Some flaws were also discussed earlier regarding errors that might have occurred during the encoding, retrieval, or mapping in the cognitive processes. We could not know if the participants encoded the target objects as we would have wanted. They might looked at the image and did not store all information on the image or even misunderstood the sentences. The problem with retrieval would occur during the test phase when the participants forgot what they saw during training phase. As for the mapping, it could just be the case that they mapped the information they encounter during

the training phase on some parts of the image that were not intended. For example the face of the person in the animations rather the touched ears.

Similar to the original experiment, our experiment produced a strange phenomenon. Although it was different from the original experimental result, in each condition, the probability of the participants choosing two answers was roughly around 0.66 and 0.33, in any order. We thought that it was a coincidental phenomenon. However, since our experiment were conducted with different experimental settings, it seemed doubtful that such a probability distribution would have appeared in which the result was contrasting the original one. We could not think of an answer to this, but we believed that there should have been a factor that causing this to happen.

6 Conclusion

This report showed the result of a modified experiment based on the third experiment 3, which was conducted without the initial mapping, in the original paper by Gattis (2004). The experiment was designed to verify whether the similarity of relational structures influences not just the mapping within concepts or within percepts, but also between concepts and percepts. However, in our experiment, the result did not verify this hypothesis. Contrary to the original result, the participant selections between the object-mapping sentence and relation-mapping sentence concentrated on the object-mapping one. Further research should compare the spatial reasoning experiments where the participants are provided with trials in a paper form against trials in a digitalized form. Another interesting future prespective is to investigate the behaviour between male and female participants, which showed a difference in our modified experiment.

References

Cooperrider, K., Gentner, D., and Goldin-Meadow, S. (2016). Spatial analogies pervade complex relational reasoning: Evidence from spontaneous gestures. Cognitive research: principles and implications, 1(1):1–17.

Gattis, M. (2004). Mapping relational structure in spatial reasoning. Cognitive Science, 28(4):589–610.

Schultheis, H. and Carlson, L. A. (2013). Spatial reasoning.

Vasishth, S., Mertzen, D., Jäger, L. A., and Gelman, A. (2018). The statistical significance filter leads to overoptimistic expectations of replicability. Journal of Memory and Language, 103:151–175.

Zehr, J. and Schwarz, F. (2018). PennController for Internet Based Experiments (IBEX).