Macalester Journal of Philosophy

Volume 20 | Issue 1 Article 9

6-21-2012

Spatial Information and Diagrams

Meghan Ertl-Bendickson
Macalester College

Spatial Information and Diagrams

Meghan Ertl-Bendickson

[This paper received the 2011 Jakob Laub Prize *in Philosophy.]*

Introduction

In recent times, it has become undesirable to use diagrams in logical proofs. Logical proofs, even in geometry, are ideally purely formal representations. Recent experiments by David Kirshner and David Landy, however, have shown that the way in which we physically arrange symbols on a page when we write a formula affects whether or not we compute it correctly. Specifically, we normally place multiplied (or divided) terms closer together than added (or subtracted) terms – following the order of operations. The operations which are supposed to be performed first are placed physically closer together than those which are done later (I shall refer to this as the "Rule of Spacing"). When formula are written inconsistent with this rule, people make more computational errors. Landy claims that this implies, through his "longer is larger" hypothesis and his "syntax" hypothesis, that there are diagrammatic elements to our formal representations. I argue that even if these spatial relations are diagrammatic, it is not a problem for logic the way using a conventional diagram would be. However, while I agree that these results are very important and need to be discussed, I argue that these spatial relationships are not actually diagrammatic.

Why Diagrams are Problematic for Logic and Math

Before we can examine Kirshner and Landy's results, we need to understand some background information about diagrams and why, exactly, it is no longer considered acceptable to use them in logical proofs. Diagrams were originally developed, in the times of Ancient Greece, for use in cartography and to find ways to accurately measure spaces and distances. This means that the first diagrams were meant to describe contingent, extensional properties of the real world. "Geometry as a discipline originated in the need to solve problems concerned with distances and areas in surveying and cartography. Its subject matter was therefore the physical features of the world, and the logical relationship its conclusions bore to these features was therefore contingent, akin to that of any physical theory." [1] They were used to deal with specific instances in space and time, for instance mapping a real landscape in a particular area. Geometry developed out of these issues.

However, it has since become something quite different. A critical change came when Descartes presented to us a way to describe geometric diagrams algebraically, allowing us to convert diagrams into formal representations.[2] This was beneficial to the study of geometry in a number of ways. It allowed geometry to directly profit from advances made in the rest of mathematics, so that if a new discovery were made elsewhere it could be applied to geometry, as well. It also solved the issue, which had been recognized for many, many years, that relying too completely on a diagram can cause error solely because the actual diagrams we

[1] Greaves, Mark. 2002. *The philosophical status of diagrams*. (Stanford, Calif: CSLI Publications), 77
[2] Ibid., 78

draw are fallible. No drawing of a triangle is ever going to be a perfect triangle, so basing your calculations on a specific drawing of a triangle can cause mistakes. Working instead with the algebra allows us to talk about "perfect" geometric shapes, without having to worry about whether our diagrams are accurate. Finally, though, Descartes allowed us to begin to discuss things that are not visualizable or intuitable. Geometry was no longer restricted to the domain of things that humans are capable of visualizing. We can talk, now, of 5-dimensional objects, or shapes with more sides than we can picture, etc. This final point makes it clear that geometry had begun to move away from its original purpose – the study of the real world and extensional, contingent spaces.[3]

Another shift came with the discovery of Non-Euclidean geometry. "After this discovery, it was unclear whether the theorems of geometry could even be considered to be *true* of objects of the world, let alone descriptive of their necessary properties, because of the uncertainty about the world's actual geometry." [4] Now there were actual aspects of geometry that specifically did *not* relate to our experience of the world. In fact, we were now left a little uneasy about the exact nature of our world – what kind of geometry do we actually have? We had assumed that there was only this one type of geometry based on rules which govern the real world. But now we could see that there were others, which follow different rules, leaving us unsure as to which one we actually live in. And for those types of geometry that do not represent our world, no diagram could now be of use to us.

[3] Ibid., 78
[4] Ibid., 79

Diagrams had at one point been essential to the study of geometry, but since then the development of geometry itself has tended in a direction in which diagrams can no longer be of substantive use.

Greaves discusses a number of the fundamental reasons diagrams cannot serve a real purpose in logical proofs. The first involves the "requirement of indeterminacy of interpretation."[5] Basically, diagrams inherently impose one interpretation on a problem, but there may be others. Using solely formal representations keeps us from becoming biased towards one particular interpretation. The second reason is slightly more subtle and more pertinent to our present discussion. Logicians, mathematicians, etc have wanted very much to keep psychological processes out of our rules of reasoning. "...the consensus among nineteenth-century mathematicians that proofs in any sort of mathematics be free of any dependency on facts unique to our particular psychology..."[6] Logic is meant to be objectively true, independent of particular human cognition. If the rules of logic are based on a particular human psychological process, then it functions only for human beings, not for the objective world. Further, if a rule of logic is based on a quirk of human cognition, we cannot be entirely sure it is true. We want to describe the world as it objectively is, not the world as we subjectively experience it.

The most fundamental problem for diagrams, however, has to do with a very basic assumption of logic. A logical proof is meant to be as broad as possible. A proof is not valid if it works only for one particular instance on one particular day, or if it

[5] Ibid., 80
[6] Ibid., 80

functions only for one discipline but not others. “A single fundamental principle has been at the center of the way that logicians from Aristotle to Frege have structured their accounts – namely, that the scope of a legitimate logical theory should be as broad and general as possible...logic should not be artificially limited in its domain of applicability, and thus it should attempt to model whatever is common about reasoning broadly conceived, however small that common fraction may be.” [7] We do not want one system of logic for biology, one for chemistry, and another for philosophy. Logic is meant to be a tool applied across all disciplines to make sure that all disciplines are consistent with the real world, not just with our own thoughts. Greaves calls this the *principle of maximal scope*. Diagrams, we have seen, were developed for a purpose in direct opposition to this. Diagrams were *meant* to describe specific, contingent instances, not broad axiomatic laws. This makes diagrams fundamentally at odds with the aim of logic.

Visual Elements in Formal Representations

So, we can see now why it has seemed so important to remove all aspects of diagram from our formal representations. Diagrams are contingent, so any diagrammatic element of a formal representation is a potential weakness to the proof. It is a point at which we cannot be sure the proof is following the principle of maximal scope or that it is detached from our psychological processes. Kirshner and Landy's experiments, however, highlight the possibility of just such an element. When we write a formula

[7] Ibid., 194

on the page, certainly that is a visual object. We may call it 'writing' instead of 'drawing', but we must admit that both are visually processed and involve spatial relationships on the page. So we need to clearly distinguish what makes something a formal representation on a page, and what makes it a diagram.

Landy describes two distinctions that have been made. The first is the concept of the difference between intrinsic and extrinsic representations. Diagrams are *intrinsic* representations, because the truth I am trying to show with my diagram is intrinsic to the diagram itself. I can draw a diagram illustrating that line A is longer than line B by drawing one line longer than the other – the difference in lengths of the lines is inherent to the drawing. In a formula, however, all of the symbols involved are arbitrary. The truth I am trying to show is extrinsic to the symbols I make – when I say 1+1=2, nothing about any of those squiggles on the page is inherently related to the numbers involved or the process of addition. The drawing of the lines, on the other hand, is *not* arbitrary.[8]

Another way of getting at this difference is to say that diagrams are direct representations, whereas formal representations are indirect. The formula 1+1=2 is indirect because I arrive at the truth of the statement only through knowledge of outside laws (what the symbol '1' means, what the rule of addition is, etc). But in the diagram of the lines, the truth directly shown to me through the symbols involved. I need no outside knowledge (besides knowing the definition of 'longer') to understand what is being

[8] Landy, David, and Robert L. Goldstone. 2007. "Formal notations are diagrams: Evidence from a production task". *Memory & Cognition.* 35 (8): 2033.

stated.[9] What both of these theories are getting at is the idea of arbitrariness. Formal representations are arbitrary, diagrams are not. So in order to decide whether something is diagrammatic or formal using these definitions, we have to ask whether it is arbitrary, direct, and intrinsic.

Landy aims to show that there are diagrammatic elements to formal representations by showing how the spatial relationships between our arbitrary symbols on the page reflect the processes going on in our calculations and also how making those relationships differ from our norm causes us to make errors. "…the rule system that governs the interpretation of formal systems carry functional spatial information – in other words, they are diagrammatic." [10] Before Landy published his papers, Kirshner[11] published a paper examining the curious fact that when people write out formulas, they place operands closer together or farther apart in reflection of the order of operations. So, 1+2x3=7 tends to be written 1 + 2x3 = 7, with the multiplied terms placed spatially closer together on the page than the added terms. He wished to see if this spatial grouping affected the way we compute, or in other words, if these spatial relationships inform the steps we take to solve an equation.[12]

To do this, Kirshner made a system called a Nonce Notation, which is a system of arithmetic completely divorced from any of the symbols we currently use. This Nonce Notation

[9] Ibid., 2033
[10] Ibid., 2038
[11] Kirshner, David. 1989. "The Visual Syntax of Algebra". *Journal for Research in Mathematics Education.* 20 (3): 274-287.
[12] Ibid., 287

had two difference versions. The first was “unspaced”, the second was “spaced”. The unspaced version had nothing in common with our current notation, the spaced version was exactly the same as unspaced, except following this Rule of Spacing we apparently use. So the two systems were thus:

Current	Unspaced	Spaced
a+b	aAb	a A b
a-b	aSb	a S b
axb	aMb	a M b
a/b	aDb	a D b
a^b	aEb	aEb
b	aRb	aRb

In the spaced version, the operations which are supposed to be performed first are placed closer together than those which should be performed last, just like what we tend to do when writing in our own notational system.[13]

Kirshner took a group of highschool students and first tested them on how well they understood math in our current notational system. Those who made minimal errors on the test then went on to take the same type of test, except using the Nonce Notation. The first test was unspaced, the second was spaced. These were students who understood the laws of math and the order of operations, so any mistakes they made would mostly be due to having trouble with the new notation. He compared the scores of the first, unspaced test to the scores of the spaced test and

[13] Ibid., 277

found that indeed, students did much better when the notation was spaced. Since the only difference between the two was the spacing, it had to be the spacing itself which made their scores go up.[14] This spacing, which is reflective of the order of operations, does seem to inform our calculations. It is not irrelevant.

David Landy did a series of experiments to follow through on these findings. In his first experiment, he tested how well people could judge the truth of a statement when the spacing of it was inconsistent (meaning, when the statement did not follow the Rule of Spacing). So he asked people (in his case, college students), whether a series of statements were true or false. Some were consistent (i.e., does "axb + cxd" necessarily equal "cxd + axb"? For which the answer is yes), and some were inconsistent (i.e., does "a+b x c+d" necessarily equal "c+d x a+b"? For which the answer is no). He found that people made six times as many errors when the spacing was inconsistent.[15] Inconsistent spacing apparently interferes with people's ability to judge the truth of a statement.

Next, Landy tested whether people really do consistently add these spacings to statements when they write or type them out. First he wrote out formulas in words (so, "one plus one equals two") and asked his participants to write the same formula out in symbols ("1+1=2"). He found that people did indeed place multiplied items closer together than added items.[16] Thinking perhaps this was a quirk of handwriting having something to do with the length of time it takes a person to think about the formula

[14] Ibid., 282
[15] Landy, *Formal Notations as Diagrams*, 2034
[16] Ibid., 2034

(meaning perhaps the gaps were due to a pause in thought), he tested whether the same would happen when typing on a computer. This time, participants were asked to convert English sentences into logical symbols ("if Jack is happy, then Jill is happy" would then become "A → B"). Again, however, people left spaces between groups reflective of the order of operations. So the spacing was present whether the formal sentences were handwritten or typed.[17]

Lastly, Landy tested how spacing affects people's ability to correctly solve formulae. First he had them solve simple expressions with just one operator – so, 1+1, or 2x3. Again, these were either consistently or inconsistently spaced. He found that the spacing mattered mainly for addition. For formulae where addition was the operator, when the spacing was wider than normal participants tended to overestimate, but when the spacing was narrow, they tended to underestimate (Proximity, 13). The last experiment involved compound computations, with more than one operator (i.e. 1+2x3=7). He found that inconsistent spacing led to errors in selecting the correct operation – operands placed closer together tended to be multiplied and operands placed farther apart tended to be added regardless of what the operator actually was.[18]

Landy proposed hypotheses to explain these phenomena beyond simply ascribing it to reflecting the order of operations. He wanted to say that this is not just a representation of the rule itself, but rather a spatial reflection of the cognitive processes that we use

[17] Ibid., 2036
[18] Landy D., and Goldstone R.L. 2010. "Proximity and Precedence in Arithmetic". *Quarterly Journal of Experimental Psychology.* 63 (10): 1953-1968, 18

to follow the rule. For the simple expressions, he proposed what he called the "longer is larger" hypothesis. He speculated that we all have a "mental number line" in our heads and when we do addition (but not multiplication), we start at the first number and "move" ourselves along the line the required number of steps and then see where we end up.[19] So for 1+1=2, I would start at one on my mental number line and then take one step forward. I see that I landed at two, and therefore know that the answer is two. But when spacing is abnormally wide or narrow, it influences my perception of the question so that I overestimate or underestimate the correct response, respectively. Thus the spacing of the formula on the page is a visual representation of the act of walking along my mental number line.

For the compound expressions, Landy offers a somewhat more subtle explanation. He claims that when terms are grouped closer together, it is a spatial representation of how *syntactically* bound together they are (I shall call this the "syntax" hypothesis). "...if, as we suggest, understanding formal symbol structures typically involves spatial resources, then symbolic productions might be expected to reflect syntactic structure: The less tightly two adjacent terms are bound syntactically, the farther apart they should be placed physically." [20] In the expression 1+2x3, 2 and 3 are more tightly syntactically bound than 2 and 1, so I place 2 closer to 3 than to 1 as a visual representation of that tightness.

[19] Ibid., 10
[20] Landy, *Formal Notations as Diagrams*, 2034

The Rule of Spacing and the Principle of Maximal Scope

If these are in fact diagrammatic elements in our formal representation, we have to ask what this shows. We have striven to remove diagrams from our computations and proofs because historically, diagrams were meant to represent contingent objects, relations, etc. Because they are contingent, they cannot follow the principle of maximal scope, which means whenever possible we should avoid them in order to keep our math and logic as broad as is possible. The other problem with the Rule of Spacing is that they seem to represent, according to Landy's hypotheses, our cognitive processes. We have tried hard to remove any psychological factors from math and logic, because, again, we do not want math or logic to be contingent on the human mind. Theoretically, another species ought to be able to use logic exactly the way we do. It ought not to work only for human beings.

However, we cannot just reject the Rule of Spacing *solely* because it is diagrammatic. We need to ask whether this is indeed a weakness, whether it does fall prey to the above problems. I argue that if these tendencies are diagrammatic, they do in fact still follow the principle of maximal scope exactly the same way that any arbitrary, formal representation would, and thus are not in fact a problem we ought to eliminate. These diagrams are of a different sort than, say, a drawing of a triangle. Yes, they are a reflection of the cognitive processes we use to solve equations, but so is the plus sign or the equals sign. These things are symbolic ways of communicating the steps we take to solve an equation, and if they are standardized, the way the equals sign is, we eliminate most the problems psychological interference might cause. They are not representations of contingent, extensional objects or relations in

the material world like our diagrams in cartography were. So while we have diagrammatic elements in our formal representations, it is not problematic in the same way.

Why the Rule of Spacing Does Not Yield Diagrams

I do not, however, fully support the idea that these *are* diagrammatic elements – specifically *because* of the differences between them and conventional diagrams mentioned above. Certainly they are visual and imagistic. But not all images are necessarily diagrams – all of our arbitrary symbols we use in formal notations are also imagistic in that they communicate their information visually. The distinction we have made is that diagrams are direct and intrinsic. For the "longer is larger" hypothesis, there could be ways to directly represent that. If we do perform addition by walking a "mental number line", a *direct* representation of this would involve making the spaces between symbols bigger for formulas in which the numbers involved are bigger; so we might have 1+1=2, and 3 + 5 = 8. This is a direct representation of our mental number line: we have to go further down it to get to 8 than we do to get to 2, so the formula directly represents this by spacing the numbers farther apart.

But this is not what Landy shown. In fact, what he has shown is the exact *opposite.* He proved that there is a common distance we put between the symbols, and that when that distance is inconsistent, it throws us off and we come up with the wrong answer. This may be proof that we are walking a mental number line and that that is how we do addition, but it is not proof that the Rule of Spacing is diagrammatic. We have a consistent distance, and any deviation from that distance is problematic. So while the

spatial relations of the symbols on the page are important, they are not any sort of direct representation, and thus are not diagrammatic.

For the "syntax" hypothesis, there is more of a sense in which the spatial relationships Landy found are direct. We are saying that two terms are closer together syntactically, and so we place them physically closer together on the page. This seems like a direct representation, or at least, it certainly does not seem arbitrary. However, the idea of two things being "more tightly syntactically bound"[21] is not a reference to a spatial relationship in the first place. The word "close" is misleading – we are referring here to a different kind of closeness. Saying two things are closer syntactically is different than saying Minneapolis is closer to Chicago than to Paris. There is no real physical distance involved in syntax, and there never could be, because syntax is not a physical object to begin with.

What do we mean by "syntactic closeness", then? We may say that being "more tightly bound" is referring to *temporal* distance, in that the terms are more tightly bound because they are dealt with first and are therefore temporally closer together ("tighter"), but then we are right back to referring directly to the order of operations. They are only temporally closer together because the rule of the order of operations says they should be, and if these spacings are only reflective of our rule, then they are most certainly not diagrammatic, unless we want to say that parentheses (which are also only reflective of the order of operations) are also diagrammatic. The spacing would then only be an arbitrary symbol

[21] Ibid., 2034

of the rule, the same way that addition and multiplication symbols are arbitrary symbols of their own respective rules. If we think about the order of operations and what it actually says, there is nothing about physical closeness that directly implies it the way saying a diagram represents that line A is shorter than line B because I have physically drawn line A shorter than line B. The Rule of Spacing, as a representation of the order of operations, is intuitively helpful, but not intrinsic. Again, for these spacings to be diagrammatic and not merely imagistic, they have to represent something in a direct way, and for these reasons if they are only representative of the order of operations, they do not.

Taking a step back, we have to further note that Landy has not in fact *proved* either the "longer is larger" hypothesis or the "syntax" hypothesis. He has shown that the spatial relationships between the symbols on the page affect the way we compute formulae. He has not shown *why* this is the case – that would require a whole different type of experiment. These two hypotheses might be plausible explanations, but they have not yet been proven or even strongly supported. Perhaps the Rule of Spacing *is* only a symbol of the order of operations, and thus arbitrary. Perhaps it is not indicative of some deeper cognitive process. Again, we use parentheses in algebra to help us follow the order of operations, and we do not consider those to be diagrammatic, even though they (like any other symbol, even the numbers) are visual.

The underlying point here is that just because something is visual does not mean it is diagrammatic. The requirement for something to be diagrammatic, by Landy's own standards, is that it is direct and intrinsic. In order for him to support his claim that

there are diagrammatic elements to formal representations, he needs not only to prove that the way symbols are arranged on the page affects the way we think, but also that the spatial relations involved are direct representations and not merely arbitrary symbols. Without this second step, all he has shown is that formulae are imagistic and that there is an aspect of that trait that affects the way we compute that we have not yet acknowledged.

The Import of the Data

I am not, however, dismissing the findings of Kirshner and Landy as insignificant. I believe it is still highly important to examine what their results mean. The issue I see for the disciplines of mathematics and logic is not that we have diagrams in our formal representations, but rather that we have implicit rules at play. It seems that an undiscussed rule has developed and been passed from teacher to student, and that it is powerful enough to cause people to make computational errors when it is disobeyed. Why the rule developed in the first place, which is what Landy is discussing with his two hypotheses, is an important and interesting question, but not necessarily relevant to mathematicians, logicians or philosophers. For those disciplines, the fact that the rule exists is the crux of the issue.

There are two ways we may address the Rule of Spacing: We may either actively suppress it, which requires explicit discussion of its existence and then for teachers to make certain they are not subconsciously passing it on to their students; or it needs to be defined and standardized, the same as the rule of addition or the order of operations. Without doing either of these, our psychological processes *are* interfering with our computations

in exactly the way we fear. The math or logic we do is being influenced by subconscious mental processes, and there may be differences in this from person to person. Perhaps what is "close together" or "far apart" for one person is different for another, and so when that first person writes out a formula in what they think is consistent with the Rule of Spacing, it is inconsistent for the second person, causing them to make a computational error. But if we make the rule explicit, perhaps standardize the distances between operands and particular operators, then this would hopefully minimize the interference of our own subjective psychologies.

There are, as Landy points out[22], a number of benefits to this rule, such that perhaps we ought not to bemoan its presence. The fact of the matter is that we are not purely linguistic beings. We also necessarily process information through our senses, since that is how we acquire it. This is unavoidable. For the purposes of written logic and mathematics, this means we process the information visually as well as linguistically. So incorporating visual elements into our rules might make it easier for us to process the information we are trying to convey. Particularly, when we first teach a student arithmetic, making the order of operations a spatial as well as a syntactic rule might make it easier to remember and follow. This would minimize the number of mistakes we make when computing formulae and help us learn faster.

In fact, the rule could be helpful for teachers as well as students.[23] If we had such a visual rule representing the steps we

[22] Ibid., 2038
[23] Ibid., 2038

took to compute a formula, a teacher could more easily see why a student got the wrong answer on a test or assignment. If a student writes 2 x 2+3=10, it is most likely that he or she did not follow the order of operations correctly and thus the teacher can much more easily correct and instruct him or her. On the other hand, if the student writes 2x2 + 3=10, it is of course still possible that he or she does not understand the rule of operations, but it is also more possible that there is some other error responsible. Basically, this visual rule is a way of representing the steps we took to solve an equation, the same way we use parentheses. So, it can communicate more efficiently to a teacher whether a student correctly understands the rule.

Conclusion

For many years, logicians and mathematicians have worked to remove diagrams from logical proofs and formulae for the reason that diagrams, due to the nature of their origins, do not follow the principle of maximal scope. We have drawn a strict distinction between diagrams, which are intrinsic and direct, and formal representations, which are extrinsic and indirect, or arbitrary. Kirshner and Landy, among others, have rather convincingly shown, however, that there are relevant spatial relationships to our formal representations – mainly, we tend to spatially represent the order of operations by placing physically closer together those operations which ought to be performed first. Landy explains these tendencies using what he calls the "longer is larger" hypothesis in simple expressions, and what I have called the "syntax" hypothesis in compound expressions. Because these spatial relationships so strongly affect the way we compute, Landy

claims they are diagrammatic.

I argue that if this were so, these diagrammatic elements would in fact follow the principle of maximal scope and therefore not be a problem the way a diagram of a triangle, for instance, is. I further argue, however, that even though the Rule of Spacing is visual and imagistic, it is not diagrammatic because the way it represents the information it is conveying is not direct or intrinsic. Regardless, the Rule of Spacing is currently an unacknowledged rule affecting the way we compute, which is problematic and needs to be addressed.

Bibliography

Greaves, Mark. 2002. *The Philosophical Status of Diagrams*. Stanford, Calif: CSLI Publications.

Kirshner, David. 1989. "The Visual Syntax of Algebra". *Journal for Research in Mathematics Education.* 20 (3): 274-287.

Landy, David, and Robert L. Goldstone. 2007. "Formal notations are diagrams: Evidence from a production task". *Memory & Cognition.* 35 (8): 2033.

Landy D., and Goldstone R.L. 2010. "Proximity and precedence in arithmetic". *Quarterly Journal of Experimental Psychology.* 63 (10): 1953-1968.