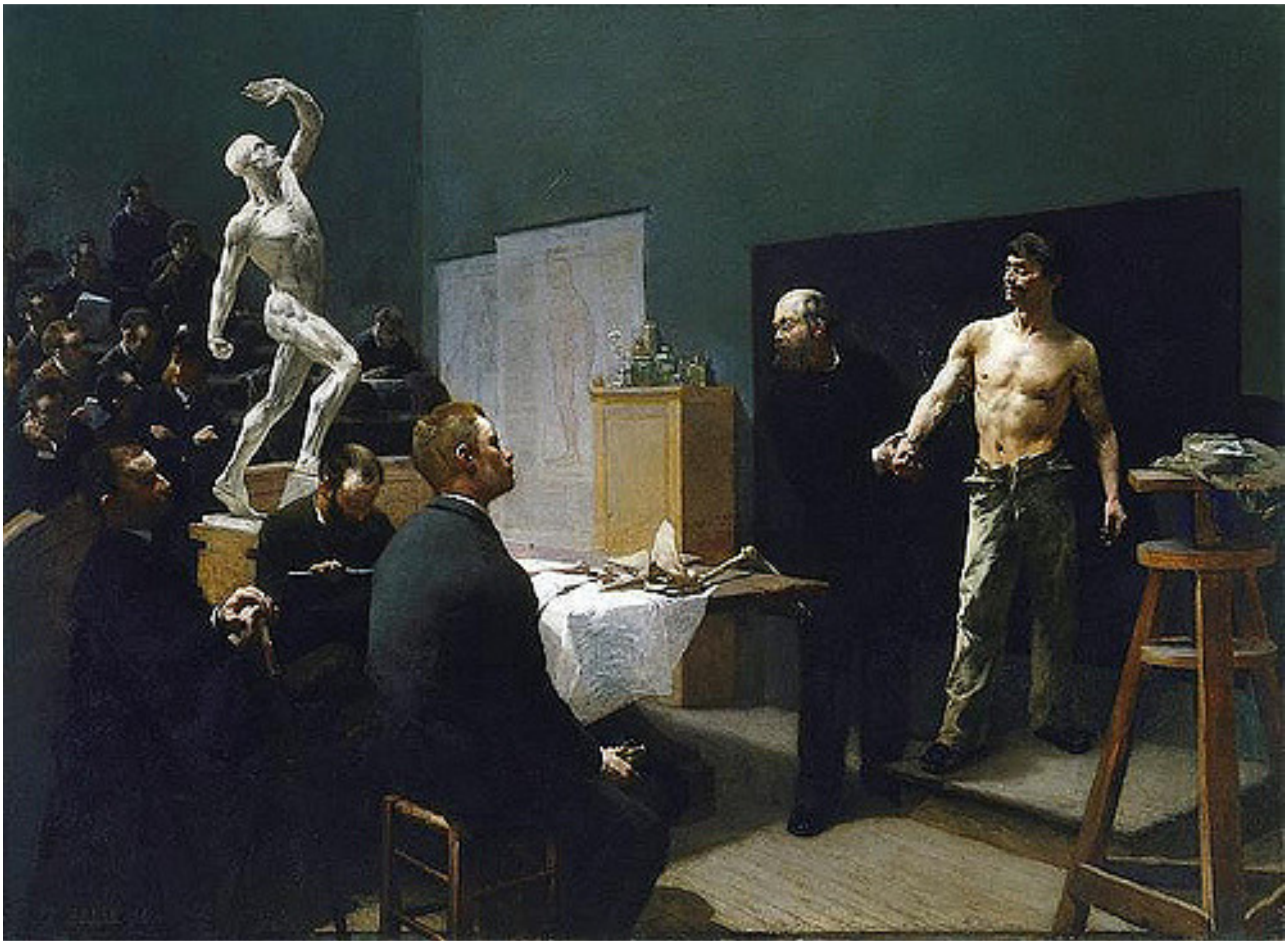
The Generality Problem for Diagrammatic Reasoning in Mathematics

One of the key problems that any theoretical account of diagrammatic reasoning in mathematics must tackle is to so-called "generality problem", i.e., "how to explain that one proof – done by referring to a particular diagram, inevitably having specific properties – can be considered as a general result." (Giardino 2017, 505). A standard response (Manders 2008, Macbeth 2010, Netz 1999, etc.) has been to draw a difference between what is particular about the diagram from what is general, so that one may argue, for example, that while what is drawn is a particular figure with specific properties, its content, i.e., what it represents is general. The idea is to draw a strong analogy between the relevant diagram and other representations in science, where particular objects are regularly used to represent general phenomena.

As early as late XVIII Century, for example, a variety of devices were used to teach anatomy: cadavers, live human models, paintings and three-dimensional models (McLachlan & Patten 2006). These three-dimensional models were sculptures representing the shape and distribution of organs in the human body. These visual scientific representations diverged substantially from the human bodies they represented. They were not made of organic materials, for example, but of stone or wood. Yet, this did not prevented them to represent flesh and blood. As long as teahcers and students were aware of which properties the (particular) model shared with its (general) object of representation, misinformation was avoided. As long as they knew that the material the model was made of was something that was not to be interpreted as part of what the model represented, for example, and that the relative distribution of the organs was, the model was epistemologically successful. A similar distinction needs to be made in the case of (at least some cases of) diagrammatic reasoning in mathematics. Thus, for example, when we use a figure like Fig. 2 in Euclid I.1. we draw a couple of closed curves and three lines to successfully represent a couple of circles and a triangle made up of their radii successfully as long as we are aware that we should not rely on our reasoning on those properties the drawing has as particular configuration of lines – for example, the size of the closed curves, the particular length of the different lines, whether they are of the same length or not, whether they are actually straight or crooked, whether the lines representing the radii actually start from the center of the curve, the inclination of the lines, etc. – and yes on those properties it shares with its general object – for example, that all the lines in the triangle have an edge on the point that represents the center of one of the circles and the other on the edge of that circle. As long as we keep both properties in check and we rely only on the second ones, our proof will be rigurous and general.



The anatomy class at the École des beaux-arts (1888) François Sallé

Just as in the case of the anatomic model, the difference is not intrinsic to the object that plays the model. It is something that emerges from the representational use we give them. For example, we can use the same male model to represent the general anatomical features of the male body in one occasion and to represent the general anatomical features of the human body in another. In each case, the epistemologically relevant features will be different. As we will see in further detail ahead, the same thing happens with diagrams: the same figure can be used to represent different geometrical facts. Nevertheless, what is important to notice right now is that the successful use of diagrams in mathematical reasoning depends not only on the objective properties of the relevant diagram, but also on how it is used and for what purposes. In this, it is not different from other tools we use. A tool is useful or not depending not only on the tool's objective properties but also on how it is used and for what purposes. Thus, if we hold that a particular way of using a tool for a given purpose explains its existence, we still need to address four substantial questions:

1. The **descritptive** question: are these tools actually used as way we say they are?
2. The **correctness** question: is such way of using the tools actually helpful for the purpose at hand?
3. The **completeness** question: what sort of purposes does using the tool this way may serve?
4. The **epistemological** question: how does the user know that the tool must be used this way for such purpose?

In the case of mathematical diagrams, the resulting questions are these:

1. The **descritptive** question: are diagrams actually used the way we say they are? Do mathematicians actually draw the line we have postulated and try to rely only on the features that are actually generalizable? In order to answer this question, we need to rely on empirical and historical data about actual mathematical practices.
2. The **correctness** question: is such way of using diagrams actually helpful for mathematical reasoning? Does this way of using diagrams reliably give us rigorous mathematical results? This question and the next are less historical and, more of a logical nature.
3. The **completeness** question: what sort of epistemic goals does using the tool this way helps us achieve? What kind of things can we prove? What kind of things can we understand?
4. The **epistemological** question: how do mathematicians know that diagrams must be used this way? How do they know which features are – to borrow Mander's terminology – "directly attributable" to what the diagram represents, and which are specific to the drawn figure?

Perhaps the best known and more developed proposal along these lines is Kenneth Mander's, based on a distinction among an geometrical object exact and co-exact properties, where only the former are generalizable, while the second are mostly byproducts of the diagram's depictive nature and have at most ergonomic value. Co-exact properties are cognitively stable in so far as they are easily and reliably graspable by mere visual inspection. For example, we have a hard time determining whether two non-overlaping angles are equal or not, and if not, which one is smaller or bigger, unless the difference is substantially large.

Thus, argues Manders, this difference cannot be co-exact and any reasoning process that relies on our visual judgement of it would be unreliable and unfit for mathematical proof. However, if one of the angles completely overlaps the other, then it becomes extremely obvious which one is bigger or smaller, or whether they are the same size. The stability of this process of extracting information visually from the diagram makes this property advatantageous on both the heuristic and epistemological senses. It is heuristically advantageous because the visual process of extracting this information is extremely easy, it is epistemologically advantageous because it is also extremely accurate: if an angle looks smaller than another, it is smaller.

Any other way the diagram may be misleading, argues Manders, must be chalked to our lack of skills in drawing a clear and adequate diagram for the relevant proof, or on inferential steps relying on the accompanying text.