# 1.7 Critical Review of the Notion of Physical Quantity - Experimental Error 

"Die Welt ist meine Vorstellung:" - dies ist die Wahrheit, welche in Beziehung auf jedes lebende und erkennende Wesen gilt.<br>Arthur Schopenhauer: Die Welt als Wille und Vorstellung.

Our notion of physical quantity still needs some substantial corrections. First of all it is not quite true that the definition of a quantity can be based on a single experimental rule of comparison and a single rule of sum of values. For example, in the case of spatial distance we defined the equality of values by fitting the spikes of a compass into pairs of points. But if we would like to talk about the distance of Earth and Sun no one would be able to fit the spikes of a compass to Earth and Sun. Also in the case of mass the desire to consider the masses of these celestial objects forces us to substitute the original experimental rule by some different rule. So in general one has to consider several rules each one with a specific domain of applicability. But whenever the domains have common elements the rules should define the same structure on theses intersection sets.
Another point that needs a critical review is far more difficult to mend. We invented a nice notion; but do physical quantities actually exist? - Well as mentioned at the end of section 1.6, with masses our students successfully verified the properties of equivalence, the compatibility of the sum rule and equivalence, associative and commutate property of the sum. But that was done with a very limited set of objects. In fact subsequently these students received another task and they discovered that things are not so easy. They were asked to separate 10 heaps of sand, each heap with a mass equal to the mass unit (the mass of the nut of Figure 1.6.5). Next they located the amount of 5 heaps in each cup of the balance. According to the rules, the balance should remain in horizontal equilibrium. But that did not occur! A slight deviation from the horizontal orientation could be seen.
One possible attitude towards this finding would be to abandon the notion of physical quantity altogether. But this radical position would not be fruitful. It is far better to admit that the experimental methods that are used to verify equality of values and to build the sum of values have some imperfections. So for example in the case of the balance, the instrument might have a slight asymmetry and the observer may not perceive small deviations from the horizontal equilibrium. We shall imagine perfect experimental procedures which result in the mathematical structure discussed in the pervious section. For every object $A$ in the domain of a given quantity $Q$ we imagine that a measurement with these ideal procedures would yield some value $Q_{A}{ }^{\text {TRUE }}$. One supposes that these "true" values satisfy all the mathematical conditions that were elaborated in section 1.6. We shall refer to these conditions as conditions of selfconsistency of the quantity. The true value is a mere idea of our mind, but part of that idea is that it is a real property of the concrete object $A$. It is supposed to exist, but we shall never get hold of it. Nevertheless it is a useful concept. It is useful provided one is able to estimate how much the actually measured values may differ from the true values. Such estimate is called an estimate of experimental uncertainty. A little later we shall discuss how such estimates are established.

Let us see how the practical application of the notion of true value works. Suppose we have some object $A$ in the domain of certain quantities $Q$ and $P$. And let us assume that we have discovered some physical law that relates the values of the quantities so that the values of $P$ are a function of the values of $Q$. So we have a known mapping $F: V_{Q} \rightarrow V_{P}$. Now we measure $Q$ and obtain a value $Q_{A}{ }^{\text {MEASURED }}$. With the help of an estimate of experimental uncertainty we come to the conclusion that the true value is an element of some set $S$ of values around $Q_{A}^{\text {MEASURED }}$. So supposedly we know that $Q_{A}^{\text {true }} \in S$, where the set $S$ is determined by the experimental result $Q_{A}^{\text {MEASURED }}$ and the estimate of experimental uncertainty. Then the physical law $F$ permits to affirm that the true value of $P$ is an element of the image set $F(S)$. Now suppose we also know how to estimate the experimental uncertainties of measurements of $P$. Then we may predict that a measurement of $P$ will yield a value that is in a set slightly larger than $F(S)$, (enlarged according to the uncertainty estimate of measurements of $P$ ). So after all, we get a prediction that the result of some real measurement of $P$ will fall into a certain set. Figure 1.7.1 illustrates this situation with a graphical representation similar to the one of example 2 of the relation " $<$ " (see section 1.4). What matters is that real results permit a prediction of other real results. So one might be inclined to cut the imagined true values out of quantitative sciences and
 formulate everything only in terms of the effectively measured values. However, this would turn our lives quite difficult, because the mathematical structure that we introduced in the previous section can only be assumed for the imagined true values.

Figure 1.7.1 True and measured values of quantities related by some function $F . S=$ set of possible true values of $Q . F(S)=$ set of possible true values of $P$. The set of possible outcomes of a measurement of $P$ will be a little bit larger than $F(S)$.
In the case of one dimensional quantities one calls the set $S$ the interval of experimental uncertainty. In most cases the estimate of uncertainty is symmetric so that the measured value $Q_{A}^{\text {MEASURED }}$ lies right in the center of the interval $S$. In this case one writes the statement $Q_{A}{ }^{\text {TRUE }} \in S$ in the form

$$
\begin{equation*}
Q_{A}=(a \pm b) \mathrm{U} \tag{7.1}
\end{equation*}
$$

where $a$ is a number, $b$ is a positive number and U is some unit of the quantity $Q$. The index "TRUE" is no longer written. $a \mathrm{U}$ is the measured value and $b \mathrm{U}$ is called the experimental uncertainty. So for example, a mass of a given body that was determined with the help of a balance capable of measurements of typical precision of 2 grams could have the following appearance:

$$
\begin{equation*}
m_{A}=(3.720 \pm 0.002) \mathrm{Kg} \tag{7.2}
\end{equation*}
$$

Note that we wrote the number 3.720 with an extra 0 at the end. As far as the numerical value is concerned one has $3.720=3.72$. But the extra digit expresses that we have some knowledge concerning that ultimate digit.
In many scientific journals it is custom to write formulas such as (7.2) without parenthesis $m_{A}=3.720 \pm 0.002 \mathrm{Kg}$. This may be justified by saying that it would be
impossible to interpret the right hand side erroneously as $3.720 \pm(0.002 \mathrm{Kg})$ because one cannot sum numbers and mass values. This is not quite true. In principle one can define such sum. In linear algebra such type of sum is called a direct sum and it leads to a new two-dimensional quantity. Also things become even more dubious, when the unit involves some numeric power of 10 . It is not at all seldom to find formulas such as $m_{A}=3.720 \pm 0.002 \times 10^{3} \mathrm{~g}$. In this case the lack of parenthesis is definitely inappropriate because this formula could be confused with $m_{A}=(3.720 \pm 2.000) \mathrm{g}$. In the present book we shall apply the rule "a product binds stronger than a sum" also in the case of products of values and numbers and therefore we shall write always a parenthesis like in formula (7.2).
Let us come back to the example of figure 1.7.1. Instead of taking the prediction as relation of really measured values one often mentions the predicted true value of $P$ explicitly. That means the theoretical prediction has the form $P_{A}^{\text {TRUE }} \in F(S)$, but usually the superscript "True" is not written. The set $F(S)$ is the uncertainty interval of the theoretical prediction based on the experimental input data $Q_{A}{ }^{\text {MEASURED }}$ and the uncertainty estimate of this input data. Then one measures $P$ and finds a result

$$
\begin{equation*}
P_{A}=(p \pm r) \mathrm{V} \tag{7.3}
\end{equation*}
$$

where V is a unit of $P$. Now if the interval of uncertainty of that measurement has a nonempty intersection with the interval $F(S)$ of uncertainty of the theoretical prediction we may say that the experimental outcome turned out to be compatible with the theoretical prediction. In principle one should abandon the theoretical description if the intersection is empty. However this decision is not always the most appropriate one. As we shall see anon, the estimate of experimental uncertainty is difficult and one may not be absolutely sure about these estimates. Therefore one may not want to abandon a theoretical description right away if the intersection of the uncertainty intervals is empty but the discrepancy of prediction and experimental outcome is not much larger that the intervals of uncertainty. Figure 1.7.2 illustrates these situations:


Fig. 1.7.2 Comparison of a theoretical prediction and an experimental outcome.

It remains to discuss the important question how the estimates of experimental uncertainty are made. In principle, what determines the estimated intervals are tests of self-consistency like the one made by the students with the help of 10 heaps of sand.

One performs measurements in order to verify the self consistency rules that were explained in section 1.6 and looks what is the minimal size of the uncertainty intervals
so that the assumption that the true values obey the self-consistency rules does not lead to contradictions. If one then makes measurements in some experiment with experimental techniques that are similar to the ones used in the tests of self-consistency one will use these intervals as estimates of the experimental uncertainty. However the use of intervals determined in self-consistency tests in an experiment that uses similar techniques is not a completely save procedure. This is evident because in apparently equal repetitions of experiments the results usually differ slightly. Therefore the uncertainty intervals that were determined in specific tests of self-consistency give only a rough estimate of the possible experimental error that may occur in a real experiment.

These procedures do not look clean at all. Many people call physics, chemistry and other similar sciences the "exact sciences". Some people state that this name is due to the fact that these sciences use mathematics. This does not seam to be a good argument, because astrology also uses mathematics and this is not a science at all. Some people say they are exact, not because there are no errors involved, but because the errors are judged. Well without doubt, it is a merit to judge the errors. But this judgment is just the most inexact thing in these sciences. It might be the best thing to abandon the name "exact sciences" altogether. What remains astonishing is that these sciences are so successful despite these imperfections.
To end this section we shall report on the tasks that our first-year students received at the end of their first physics lesson. They were asked to measure the size of the small deviations from the ideal behavior of the balance with a somewhat different method of measurement. They compensated the deviation by putting an appropriate piece of paper steamer (the kind that is used in carnival) on one side of the balance to reach horizontal equilibrium. After that they spooled down a much longer piece of that ribbon on one plate of the balance until the balance got into horizontal equilibrium with a mass unit (the nut) on the other side. So the long stripe of ribbon was known to have the mass of one mass unit. Then they measured the lengths of both pieces with the help of a long ruler and determined the mass deviation from the ideal behavior applying a simple rule of three.

Finally the students received the following homework:
Exercise: We measured the length of paper ribbons with a ruler. Well, everybody knows how to use a ruler. Also every monkey knows how to open and to eat a banana, but the monkey does not know what he is doing! Scientists should not use rulers the way monkeys use bananas! So what is a ruler? A ruler is a large collection of distance values that helps to measure distances without the need to apply the experimental rules of sum of distance. We need only compare distance values and the formation of the sum of values has already been taken care of by the manufacturer of the ruler. But that presupposes that the manufacturer knew how to form the sum of distance values. One needs a definition of the sum of distance values: So there are your tasks:
(a) First define an order relation of distance values.
(b) Then define the sum of distance values.
(c) Finally write what is wrong with the definition: Well, in order to build the sum of two distance values I express the values that I want to add in terms of a unit and then I simply add them: $5 \mathrm{~cm}+2 \mathrm{~cm}=7 \mathrm{~cm}$.

