THE SYMBOLIC LANGUAGE OF SUBSTANCES AND MOLECULES: NOISE OR HARMONY?

J.D Bradley
RADMASTE, Wits School of Education,
University of the Witwatersrand, Johannesburg, South Africa
Email: john.bradley@wits.ac.za

ABSTRACT

Most substances are given names and formulae based upon knowledge of their molecules. However for substances most commonly met in elementary chemistry courses, especially inorganic substances, this is often not the case. The potential noise is amplified further when dealing with chemical reaction equations. It is argued that since the names and formulae given to substances and their molecules are often the same, we should give more attention to the use of state descriptors and symbols for substances and use the word molecules when we mean to refer to them. *[African Journal of Chemical Education—AJCE 6(2), July 2016]*

INTRODUCTION

In a previous paper [1] we have reviewed the language of teaching and learning in the context of the macroscopic and microscopic (or sub-microscopic) levels of concept descriptions. A case was made for developing a macro-micro dictionary that would juxtapose the two levels of description. These two levels also correspond closely to the molar and molecular levels identified by Jensen [2].

Johnstone [3] drew attention to three ways in which we communicate chemical concepts, two of which are the macroscopic and microscopic ways. The third way he called symbolic, and in this paper we examine how this relates to the other two. We address the question of the extent to which the relationships support the macroscopic-microscopic descriptions, developing ideas that were first presented in 2008 [4].

In so doing we shall particularly focus upon chemical formulae and names and chemical reaction equations. Do these harmonise with our descriptive language or do they add to the noise? The level of discussion is intended to be suited to those teaching chemistry at the secondary and tertiary levels and should be read with this purpose in mind.

NAMES AND FORMULAE OF SUBSTANCES AND MOLECULES

Before Lavoisier's time, substances were given idiosyncratic names and hieroglyphic-like symbols [5]. All of this changed in the wake of Lavoisier's definition of elements (1789). In an obvious development based upon identifying names of elements, he proposed that the names of compounds should reflect the names of the elements they comprised. Hence for example, copper oxide, sodium chloride, etc: the binary naming system, later extended to include common groups like sulfate and hence, copper sulfate etc. After some years the publication of Dalton's atomic

theory and the determination of atomic weights by him and others and then the introduction of element symbols by Berzelius, led to the appearance of formulae, such as CuO and NaCl. Although the first of these could be called copper oxide, when another oxide was identified as Cu2O, the need for a somewhat more sophisticated naming became apparent. The names cuprous oxide and cupric oxide were created for Cu2O and CuO, respectively. A similar situation arose with the formulae CO and CO2. In these two cases the names adopted were treated differently: carbon monoxide and carbon dioxide. The latter names are systematic in the sense that the names directly reflect the atomic composition. In the case of the copper oxides, this is not so; the terms cuprous and cupric require special interpretation. Another approach (using the concept of oxidation state) was also suggested: thus copper (I) oxide instead of cuprous oxide and copper (II) oxide instead of cupric oxide. All of this can be learned but it adds to the burden of new learners, who in the early stages of learning are rather unlikely to understand the oxidation state concept. Furthermore the temptation to misunderstand what the Roman numbers are saying is strong: the (II) does not imply two atoms of copper! The more systematic naming of the two oxides as dicopper oxide and copper oxide, has a more direct relationship with the atomic composition. Unfortunately these names are rarely used. We thus find today the persistence of names of many common substances that are essentially macroscopic in nature (belief in atoms not required!), whilst the formulae they have are microscopic in nature (reflecting atomic composition).

This short discussion draws attention to just a small part of the confusion that may be created with names and formulae. It is as if the older days of idiosyncratic naming retain their appeal, except for those who feel that oxidation states must be learned before we know what they are.

A similar problem arises when naming a series of compounds with the formulae NaCl, CaCl2, AlCl3, SiCl4, etc. The commonly-used names of these compounds are sodium chloride, calcium chloride, aluminum chloride, silicon tetrachloride. Frustrating for the beginner is the failure to use names like calcium dichloride and aluminum trichloride, which reflect directly the chemical formula in each case. Experienced chemistry teachers may say that a name like calcium chloride is what everyone uses and there is no ambiguity because there is only one chloride: this of course is true, but again this usage adds to the burden of comprehension for beginners. There is much evidence of the reality of this burden: for example, Bradley and Steenberg found [6] that learners had greater difficulty in encoding from names to formulae, than the reverse decoding: learners resorted to creating their own chemical symbols and combining atoms and groups in 1:1 ratio by default.

Although Lavoisier started the systematic naming of substances and Berzelius developed systematic formulae thereafter [7], we seem to have failed to develop and use what they began in a completely logical way. In some cases (as exemplified above) we find that whereas the formula of a compound reflects its atomic composition (a microscopic description), the name essentially reflects elemental composition (a macroscopic description). Trying to bridge this gap by drawing upon other concepts like oxidation state seems perverse.

Another aspect of the problem is the common way in which the binary nomenclature is introduced in secondary schools. This is almost exclusively done by listing ions with their formulae and charges and then explaining the balancing or matching of charges in constructing the formula of a substance. Thus, even before the concept of ionic bonding, this naming approach leads learners to a conviction that ionic substances predominate and that entities (atoms and groups of atoms) like Cl, OH, SO4, are always negatively charged. Hence alcohols are a source of hydroxide ions

and when a chlorine molecule dissociates it must yield chloride ions! In fact such entities have a valency, a concept which is not associated with any charge or model of bonding, and this would be a better characteristic to emphasize instead of charge [8].

All of the foregoing relates to the "classical" world of inorganic chemistry, which is a major part of secondary school chemistry curricula. The naming is described as binary nomenclature. Somewhat later in the development of chemistry, organic compounds became better understood and presented more serious challenges of communication. Although non-systematic names were used initially (and a few remain as "trivial" names), they were quickly and comprehensively overcome by the systematic naming developed and maintained by IUPAC. Predominantly a substitutional nomenclature was adopted and widely used at all levels. Although this does require the learning of a system, the system is logically related to the atomic composition of the compounds. Thus, for example, the name chloropropane directly informs that one chlorine atom has replaced one hydrogen atom in the propane molecule. Indeed it allows one to go even further in that it can give structural information about the molecules, by the use of locants together with a simple numbering scheme, eg 2-chloropane. This caters for the prevalence of structural isomers in a natural way too, as in 1-chloropropane and 2-chloropropane.

Thus we can see that in organic chemistry both the name and the formula apply to the substance and the molecules of the substance. This is an admirably simple situation, although it makes it easier to confuse the macro and micro descriptions! The only salvation then is to add the word "molecule", when the name or formula is used with reference to that entity rather than the substance.

EMPIRICAL AND MOLECULAR FORMULAE

Before the concept of molecules had been fully agreed upon, chemists were very happy to be able to determine the atomic proportions of a compound and to represent this information with a formula. This sort of formula is called an empirical formula. With the growing awareness of molecules, it became evident that whilst the atomic proportions must apply to the molecules as much as to the compound, the actual numbers of atoms per molecule was a different and very important matter. This too could be represented by a formula, called a molecular formula. Thus hydrogen peroxide (or more systematically, dihydrogen peroxide) has a molecular formula H_2O_2, indicating that the molecule comprises two hydrogen and two oxygen atoms. This is not revealed in the empirical formula of the compound, which is HO. The molecular formula is far more informative than the empirical: however both formulae symbolically represent both substance and molecules.

Turning then to such familiar compounds as sodium chloride, NaCl, here we find the formula is an empirical one, not a molecular one. Nevertheless the nature of the molecules of solid sodium chloride is known: they are formed of huge networks of atoms and bonds and have an indefinite size. Hence an appropriate molecular formula could be $(NaCl)_n$. Once again, the adoption of a molecular formula, is very informative. Furthermore, it makes sense of the fact that solid sodium chloride has a high melting point, unlike hydrogen chloride, HCl, which is a gas! (It should be noted that nothing explicit need be said about the nature of the bonding or even the geometric arrangement of the atoms in the lattice.)

Somewhat similar issues are met with in respect of the formulae of elements. Thus single atoms are quite rare (Noble Gases) and we find molecules of elements from the simplest (eg Cl_2) to the most complex (eg C_n). In common usage unfortunately, solid carbon is usually given an

empirical formula, that is C, whilst chlorine is always given a molecular formula. The lack of consistency again creates potential problems for learners; for example why is Cl2 a gas whilst C is a high melting solid? The macroscopic observable (mp in this case) bears no relationship to the commonly used formulae (microscopic), when one is empirical and the other molecular!

Jensen [9][10] has drawn attention to the lack of explanation by textbook authors for the formulae used in representing solid substances and to the inadequacy of the descriptors empirical and molecular (suggesting they be replaced by "relative" and "absolute", respectively).

In conclusion, if we want to make meaning of the macro/micro/symbolic relationships as envisaged by Johnstone, amongst the things we need to undertake is a study of how our symbols and formulae work to make the relationship effective or otherwise [11]. Molecular formulae can serve the purpose much better than empirical formulae, and can help to make sense of the macro-micro relationship. However to achieve this we must be more consistent and logical in their use.

THE STATES OF SUBSTANCES AND MOLECULES

The states (or states of aggregation) of substances are often represented by symbols. The more familiar of these include (s) for solid, (l) for liquid, (g) for gas, and (aq) for aqueous solution. These symbols are appended to the formula of the substance, as for example, HCl(g), NaCl (s), C6H6(l) and CuSO4(aq).

A molecule does not have a state. A set or collection of molecules (a chemical species [12]) does. Hence symbolically a molecule of hydrogen chloride, for example, is the same, regardless of its environment. This is not to ignore that in aqueous solution, molecules of hydrogen chloride are scarce because of their strong tendency to transfer a hydron to a water molecule. Thus when

symbolically representing a molecule of hydrogen chloride, the formula HCl is correct, regardless of its environment. Similar remarks would apply to benzene molecules, C6H6.

In the case of a solid substance like sodium chloride, a suitable molecular formula would be (NaCl)n. In molten sodium chloride it is unclear what the nature of the microscopic entities is. In the gas state there are molecules of formula NaCl.

CHANGES OF SUBSTANCES AND MOLECULES

Equations are often used to represent changes of substances and molecules. Word equations use the names of substances and not formulae, and are often used when introducing the macroscopic features of chemical reactions. However, changes of substances and molecules are most frequently represented symbolically. When a single substance is involved the change may be a change of state:

$$HCl(l) \;=\; HCl(g)$$
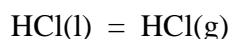
where the equal sign means there is a specific stoichiometric relationship between the initial state (represented on the left) and the final state (represented on the right). The symbolic representation of the sublimation of a substance like sodium chloride however needs careful thought: consistently one may use empirical formulae -
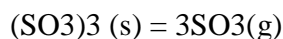
$$NaCl(s) = NaCl\,(g)$$
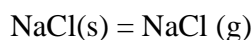
or molecular formulae -

$$(NaCl)n(s) = nNaCl\,(g)$$
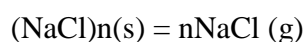
Changes of this type are not limited to "salts"; for example, using molecular formulae to symbolically represent the sublimation of solid sulfur trioxide:

$$(SO3)3 \ (s) = 3SO3(g)$$

and similarly for the depolymerisation of many organic addition polymers.

These last two examples are chemical reactions because the molecules undergo bond changes. To refer to these as physical changes because they merely involve a change of state, would be misleading [13]. It is another instance of the persistence of macro-descriptions of changes when the micro-descriptions show them no longer to be suitable.

To represent symbolically the molecular changes requires that we use the molecular formulae without the state descriptors. In all other respects, the equations remain the same.

The same considerations apply to representing more conventional chemical reactions involving more than one reactant, such as:

$$C(s) \ + \ O2(g) \ = \ CO2(g)$$
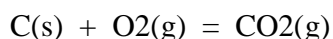
This balanced chemical equation uses a mixture of empirical and molecular formulae. More consistent is the following use of molecular formulae:

$$Cn(s) \ + \ nO2 \ (g) = \ nCO2(g)$$
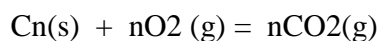
More familiar differences in representation are found in reference to elements like sulfur where different texts use the empirical formula, S, or the molecular formula, S8.

The use of mixed equations, that is ones in which a mixture of empirical and molecular formulae is used, is surely adding to the confusion and should be avoided.

Chemical reactions in aqueous solution often present further challenges to helpful communication. At the macro-level it is surely simple as well as correct to write, for example:

$$NaOH(aq) + HCl(aq) = NaCl(aq) + H2O(l)$$
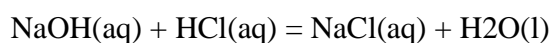
This summarizes the important stoichiometric relationships upon which a variety of quantitative problems may be set, but its relationship with how we mostly think of the situation is tenuous. The adoption of the symbolism NaOH(aq), etc, is convenient but may be said to mislead beginners. The same reaction is sometimes written for more advanced learners in ionic terms as:

$$Na^+(aq) + OH^-(aq) + H^+(aq) + Cl^-(aq) = Na^+(aq) + Cl^-(aq) + H_2O(l)$$
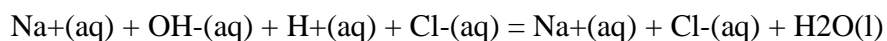
Use of the state descriptor (aq) suited to macro-descriptions means this is a hybrid symbolism, which may be justified by our uncertainty regarding the molecular composition of the hydrated ions. A strict micro symbolism would surely require formulae of the type $Na(H_2O)_x^+$, which shows the entity is a molecule with a charge (it is an ion that is a molecule) rather than showing the atomic ion apparently floating in a continuous "sea". Although the value of "x" may be uncertain in the case of Na+ for many others it is established (eg for Cu2+ it is 4).
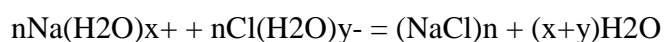
When the water is largely evaporated from such a reaction mixture, sodium chloride crystallizes. This is represented at the macro-level as:

$$NaCl(aq) = NaCl(s)$$
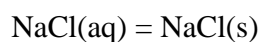
or in hybrid form as:

$$Na^+(aq) + Cl^-(aq) = NaCl(s)$$
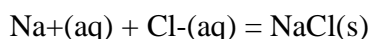
At the micro-level this becomes more explicit as:

$$nNa(H_2O)_x^+ + nCl(H_2O)_y^- = (NaCl)_n + (x+y)H_2O$$

Although we can claim to represent a reaction at the micro-level by using appropriate formulae, in general the complete representation of a chemical reaction at the micro-level would require knowledge of the reaction mechanism. When the mechanism is unknown or is not appropriate for the educational context, then it must be stressed that equations that show suitable

molecular formulae are not necessarily full micro-representations, because they do not show the mechanism.

The conceptual havoc that may be created by ill-considered descriptions is well illustrated by the teaching and learning of the Brønsted theory of acid-base reactions. This theory involves the concept of proton transfer (better called hydron transfer) and results in such mind-blowing assertions as water is an acid or water is a base (when in fact water is neutral!), supported by symbolic equations. Or again that "acid + base = base + acid", so what about neutralization and salt formation and all the traditional concepts that children learn from earlier grades? As Barke and Harsch have recently pointed out [14], the root of the conceptual upset lies in the switch from substance language and symbols to molecular language and symbols.

STRUCTURAL FORMULAE OF MOLECULES

Part of the symbolic language associated with molecules is their structural formulae. There is nothing equivalent in the language associated with substances. Structural formulae show explicitly the existence of chemical bonds between the atoms of a molecule. These formulae must be consistent with the molecular formulae but do not demand any electronic details. Apart from showing the connectivity between atoms they may also show 3-D spatial relationships. Understanding what structural formulae show (and do not show) is an important stepping stone towards discussion of chemical bonding, which opens up the electronic level as described by Jensen [2].

In organic chemistry the naming of molecules is directly based upon knowledge of their structure, so naming links with structural formula rather than molecular formula. Nevertheless, molecules and substances have the same name.

In inorganic chemistry structural formulae are less pervasive. Indeed it is normal for structural formulae to first make their appearance in a chemistry curriculum within the context of theories of chemical bonding and the 3-D spatial geometry adopted by molecules (VSEPR theory). Consequently there is very little of a factual knowledge base to such teaching and learning in inorganic chemistry. Even a comparatively simple, yet familiar molecule such as H_2SO_4 is a structural mystery to many teachers and learners. The bonding of the two hydrogen atoms to two of the oxygen atoms comes as a revelation!

Chemical bonds between atoms are shown in structural formulae usually by means of simple lines between the atomic symbols. In simple cases these lines may be equated with pairs of electrons between two atoms. However it is not necessary to accept this as a rigid requirement. Nor need there be concerns about the extent to which the bond is polar or approaches the ionic condition. As the IUPAC definition puts it [12]:

"When forces acting between two atoms or groups of atoms lead to the formation of a stable independent molecular entity, a chemical bond is considered to exist between these atoms or groups."

Thus less simple cases are amenable to the language of structural formulae, with for example, instances of delocalized bonds being represented with dotted lines.

Care must be taken to avoid confusion with lines which aim to highlight geometrical relationships. This is a problem commonly encountered in textbooks that show the types of spatial arrangements associated with descriptors such as tetrahedral or octahedral. For example, the octahedral SF_6 molecule may be shown with bonds between the F atoms as well as between the F and S atoms! Clearly bond lines and geometry lines should not be in the same picture without very distinctive coding differences.

NOISE OR HARMONY?

The preceding sections have argued that, in the symbolic language of substances and molecules, we hear something like an orchestra practicing. There are times when harmony is apparent and others when it is noise that we hear. As Johnstone has argued in another connection, noise is to be avoided if learning is to be facilitated [15]. We may also note that, although little explicit reference to misconceptions has been made above, it is implicit that symbolic language problems we have identified help to promote them.

To summarize we have found:

1. Substances have names and formulae and these may or may not be systematically related. In addition both empirical and molecular formulae are used for substances, and often without any indication as to which. Much of this noise originates from inorganic substances of the simpler and longest known type. The names of these "old" inorganic substances are often still rooted in the simple pre-Daltonian (macroscopic) language whereas the formulae are mostly post-Daltonian (microscopic). It is particularly unfortunate that the old names (eg calcium chloride) masquerade as systematic. These substances are the ones predominantly encountered by beginning learners of chemistry, thus adding to their difficulty in making sense of the subject.

2. Organic substances and "newer" inorganic substances generally have names that reflect not only the molecular formula, but the structural formula. Clearly the pressure to adopt a universal and systematic nomenclature was felt early in organic chemical history as the enormity and diversity of the organic chemistry "jungle" became evident. Use of structural formulae entails being explicit about the bonds in the molecules. This harmonious situation is in sharp contrast with the "old" inorganic chemistry where structural formulae are rather

rarely encountered (except for the case of water!). In our experience teachers are often very surprised to see a structural formula for something as simple as sulfuric acid, and there is much evidence that learners are uncertain which atoms are bonded together when presented with simple formulae, like H2SO4.

3. The fact that substances are mostly named in a manner determined by their molecular formulae, potentially constitutes a core conflict between describing a substance and describing the molecules of the substance. The simplest way to help distinguish the meaning is to use state symbols with substance formulae. This can be reinforced by appending the word molecule(s) when using formulae and names for these entities. Perhaps in this simple way noise may be transformed into harmony.

4. The potential confusion in understanding of substances and molecules due to the noise in some areas, extends into symbolic representations of physical and chemical change. We may even find equations in which a mix of empirical and molecular formulae is used without the distinction being noted. Of course it is often the case in an educational context, that methods for obtaining correct answers to stoichiometric calculations are the principal concern rather than comprehension of the chemical events [16].

Thus the relation of the symbolic corner of Johnstone's triangle to the macroscopic and sub-microscopic corners is clear in principle, but not in practice. As Dori and Hamieri [17] argue, the symbols are the basic language that mediates the interrelationship of the other two, and without paying explicit attention to this, students are handicapped. Any macro-micro dictionary, as proposed previously [1], must therefore pay attention to the need for greater symbolic clarity.

REFERENCES

1. Bradley, J.D. (2016). What are the molecules doing? AJCE, 6(2), 33-53.
2. Jensen, W.B. (1998). Logic, history and the chemistry textbook. I Does chemistry have a logical structure? II Can we unmuddle the chemistry textbook? J. Chem. Educ., 75, 879-887, 961-969.
3. Johnstone, A.H. (1991). Why is science difficult to learn? Things are seldom what they seem. Journal of Computer Assisted Learning, 7, 75-83.
4. Bradley, J.D. (2008). Abs 20th ICCE, Mauritius, 11.
5. Kolb, D (1978). The chemical formula, Part 1: Development. J. Chem. Educ., 55, 44-47.
6. Bradley, J.D. and Steenberg, E (2006). Symbolic language in chemistry – a new look at an old problem. Proc 19th ICCE, Seoul, 140.
7. Sutton, M. (2008). A clash of symbols. Chem. World, 5 (11), 56-60.
8. Nelson, P.G. (1997). Valency. J. Chem. Educ., 74, 465-470.
9. Jensen, W.B. (1977). Chemical symbolism and the solid state. J. Chem. Educ., 54, 277-279.
10. Jensen, W.B. (2004). Empirical formulas and the solid state: a proposal. J. Chem. Educ., 81, 1772-1774.
11. Treagust, D.F., Chittleborough, G., & Mamiala, T.L. (2003). The role of submicroscopic and symbolic representations in chemical explanations. Int. Journal of Science Education, 25, 1353-1368.
12. IUPAC Compendium of Chemical Terminology, Electronic Version. http://goldbook.iupac.org (version 2.3.3 accessed 25 February 2016).
13. Taber, K (2002). Chemical Misconceptions – prevention, diagnosis and cure, Volume 1: theoretical background; Royal Society of Chemistry, London, pp 97-100.
14. Barke, H-D., and Harsch, N (2016). Donor-acceptor reactions: Good bye to the laboratory jargon. AJCE, 6 (1), 17-30.
15. Johnstone, A.H., Sleet, R.J. and Vianna, J.F. (1994). An information processing model of learning: its application to an undergraduate laboratory course in chemistry. Stud. Higher Educ., 19 (1), 77-87.
16. Nurrenbern, S.C. and Pickering, M (1987). Concept learning versus problem-solving: is there a difference? J. Chem. Educ., 64, 508-510.
17. Dori, Y.J. and Hameiri, M. (2003). Multidimensional analysis system for quantitative chemistry problems: symbol, macro, micro, and process aspects. J. Res. Sci. Teach., 40, 278-302.