Quick guide

Transformation, encoding and representation

Barbara Webb

What's the difference? Transformation is when something is turned into something else, as in the conversion of carbohydrates to sugar during digestion.

Encoding is transformation into a code that carries information and can be decoded, as in the transformation of a text message into Morse code.

The term 'representation' is used in many senses, but is generally understood as a process in which something is used to stand in for something else, as in the use of the symbol 'I' to stand for the author of this article.

So what is the (biological) problem? These terms are often conflated in neurobiology, with the firing pattern of a neuron (or across a population of neurons) interchangeably referred to as a 'representation', 'encoding' or 'transformation' of an external stimulus.

Are these terms not interchangeable? To say a neuronal firing pattern is a transformation of a stimulus places the focus on the causal relationship, like any other physiological process: how the internal state of the nervous system is changed by the external state of the world, and how this internal state mediates the production of behaviour. 'Encoding' suggests the internal state is an information-preserving mapping, involving systematic correspondence, such as maintaining the spatial layout of the receptor array, or the firing rate varying with signal amplitude. It also implies that what stimulus property is mapped to what neural property is in some sense arbitrary, as long as it is consistent, that is, can be decoded. The role of a code in a system depends on the information it carries, rather than its physical structure *per se*; the classic biological example is the genetic code.

Representation, in the sense of standing-in, suggests the neural activity can be used instead of the stimulus, indeed, even when the stimulus itself is absent. This point was central to the introduction of the notion of representation in cognitive psychology [1] — that not all behaviour could be explained by reference to current stimuli and the history of reinforcement. Direct sensing is *presentation* of the stimulus, but some cognitive capabilities depend on the ability to *re*-present (possibly in a highly abstracted form) the stimulus when it is not directly sensed.

Does it make any difference what terminology we use? Calling neural activity the 'transformation' of a stimulus is merely to say they are causally related, in some way that can be discovered. To say it 'encodes' a stimulus is a stronger empirical claim: that there is a systematic mapping between the stimulus and the neural activity; raising the question of what information is preserved and implying we can discover how it is decoded by the nervous system. To say it 'represents' a stimulus suggests that the processes leading to that neural activity have as their aim the construction of something that can stand in for the stimulus.

Can you give an example? A toad-like system for catching prey could be built using a set of vector weights to simply transform the output of an array of motion sensors to the appropriate activation of a set of tongue muscles. Real toads are more adaptive [2], combining the output of motion sensors with other information about object dimensions, motivational state, and so on, to encode the presence of prey items, not just moving objects. A super-toad could hypothetically construct an internal geometric model of the prey's trajectory and its own position in three dimensional space, such that it could plan to make any arbitrary movement and produce a tongue snap to where it expects that particular fly to be relative to itself (and notice, if it fails, that something about the representation was incorrect).

So the different terms suggest different mechanisms? Yes, so their use is not neutral, but influences experimental paradigms and the interpretation of data. There is a great deal of work on trying to 'crack' the 'neural code' [3], which assumes there is an information-preserving mapping between stimulus and neural activity.

There is also a significant amount of work on trying to find where and how the brain reconstructs its perception of the world, for example addressing the 'binding problem' [4] or looking for the neural correlates of consciousness [5]. These approaches clearly go beyond the basic assumption that we can find consistent causal relations between stimuli, brain states and behaviour — they assume there is more than mere transformation taking place.

But aren't nervous systems always doing encoding? It is always possible to treat neural activity as a code for the stimulus, to try to determine what information has been preserved. Whether this is always the most productive way to understand the function is a different question. We cannot assume that the information is actually decoded in some subsequent stage of processing. Indeed it is not entirely clear what 'decoding' means in this context. The nervous system is not trying literally to reconstruct the stimulus, in the way that DNA is used to reconstruct proteins, or a telegraph operator reconstructs a verbal message from Morse code.

What the nervous system needs to do, in general, is to transform the input into the right action. Except in the special case of imitation, the mapping from input to output is not one of identity.

And processing to reconstruct the stimulus takes us in the opposite direction from processing to perform the right action.

But generally information is preserved? As the task is to produce the right action, it should not be surprising that the 'neural code' is often not especially good at preserving information [6]. There is never a strictly isomorphic relationship between stimulus and neural response: it is typically non-linear, often non-monotonic, and may change with the contexts of the history of stimulation (adaptation and learning), of other signals and of behaviour.

It will sometimes make more sense to treat this as simple transformation, particularly in cases of basic sensorimotor control, when significant transformation may be implemented in the sensory physics, as 'matched filtering' [7]. What then becomes interesting is to discover, empirically, those systems in which the information does seem to be preserved; even more, when what is encoded is some information in the world that does not correspond to a single sensory input but seems to require the combination of several inputs to construct. An example is the recently discovered 'grid' cell system in rat hippocampus [8].

What about representation? To say the neural activity caused by a stimulus 'stands in' for it, despite being a common way of speaking, collapses the critical distinction between simple causal mediation, and the ability to use one thing instead of another. For example, in discussing the relation of distal to proximal stimuli, it does not seem sensible to say that we are using the light waves instead of the object; at the very least, this case clearly differs from when we use one visible object, such as a landmark, instead of another, such as a currently invisible goal location, to do a task.

To say that neural activity is used 'instead of' whatever caused it seems to involve a similar confusion of points of view. A neuroscientist, as an external observer of both an animal's neural activity and its stimuli, might use one to stand in for the other. The animal, however, cannot; it has access to stimuli but no access to brain states. Considered as subsystems, some parts of the brain have access to neural activity of other parts of the brain, but they do not have access to the stimuli.

So are you claiming nervous systems do not use representation? Not at all. The argument is for treating representation as a (special) function of (some) nervous systems; specifically, the ability to recreate internally something that has the same effect, or can be used in the same way, as an external situation, especially when that external situation does not currently pertain.

A good example is the notion of an emulator [9]: that nervous systems may require predictive models of the consequences of their own actions to perform successful control. A simpler example might be memories from which we can extract new information about past states of affairs to influence current action. Note not all learning — in particular not simple associative learning — allows this flexibility.

A final example might be the demonstration — yet to be provided? [10] — that rats actually use the neural encoding of space in their place-cell system to calculate novel routes to a goal. One way to characterise this idea of representation is that it requires the organism to be able to detect when misrepresentation has occurred, for example to discover (by direct sensing) that the represented state of affairs does not pertain, or that the expected result from the actions based on the representation fails to occur.

Many people use 'representation' more loosely, for example, to mean transformation or encoding; is this really a problem? Current usage results in a tendency to slide between the different meanings. Researchers start out discussing 'neural representation' as the relationship of input signals, neural activity and behavioural output (transformation), then define it as "the information the [neural] signal provides when decoded", and then draw an explicit link between this coding and the intentional content of mental states [11].

Understanding the neural mechanisms linking stimuli to behaviour is vital but in many cases will tell us nothing about mental content. Similarly, understanding mental content is a fascinating question, but is not answered by using the label 'representation' for all neural transformation. Finally, identifying cases where the nervous system needs to, and does, reconstruct a stand-in for the stimulus is a critical empirical issue and should not be trivialised by the assumption that this is the function of all neural processing.

References

1. Craik, K. (1943). The Nature of Explanation (Cambridge: Cambridge University Press).
2. Ewert, J.-P. (1997). Neural correlates of key stimulus and releasing mechanism: a case study and two concepts. Trends Neurosci. *20*, 332–339.
3. Rieke, F., Warland, D., de Ruyter van Steveninck, R., and Bialek, W. (1999). Spikes: exploring the neural code (Cambridge, MA: MIT Press).
4. Neuron special issue. (1999). Neuron, *24*(1).
5. Crick, F., and Koch, C. (1998). Consciousness and neuroscience. Cerebr. Cortex *8*, 97–107.
6. Akins, K. (1996). Of sensory systems and the "aboutness" of mental states. J. Philosoph. *93*, 337–72.
7. Wehner, R. (1987). Matched filters - neural models of the external world. J. Comp. Physiol. A *161*, 511–531.
8. Hafting, T., Fyhn, M., Molden, S., Moser, M.-B., and Moser, E.I. (2005). Microstructure of a spatial map in the entorhinal cortex. Nature *436*, 801–806.
9. Grush, R. (2004). The emulation theory of representation: motor control, imagery, and perception. Behav. Brain Sci. *27*, 377–442.
10. Muir, G.M., and Taube, J.S. (2002). The neural correlates of navigation: do head direction and place cells guide spatial behaviour? Behav. Cogn. Neurosci. Rev. *1*, 297–317.
11. deCharms, R.C., and Zador, A. (2000). Neural representation and the cortical code. Annu. Rev. Neurosci. *23*, 613–647.

School of Informatics, University of Edinburgh, JCMB Kings Buildings, Mayfield Road, Edinburgh EH9 2QL, UK.
E-mail: bwebb@inf.ed.ac.uk