Is coding a relevant metaphor for the brain?

Romain Brette[1*]

[1] Sorbonne Universités, UPMC Univ Paris 06, INSERM, CNRS, Institut de la Vision, 17 rue Moreau, 75012 Paris, France

* Correspondence: R. Brette, Institut de la Vision, 17, rue Moreau, 75012 Paris, France

romain.brette@inserm.fr

Abstract

"Neural coding" is a popular metaphor in neuroscience, where objective properties of the world are communicated to the brain in the form of spikes. Here I argue that this metaphor is often inappropriate and misleading. First, when neurons are said to encode experimental parameters, the implied communication channel consists of both the experimental and biological system. Thus, the terms "neural code" are used inappropriately when "neuroexperimental code" would be more accurate, although less insightful. Second, the brain cannot be presumed to decode neural messages into objective properties of the world, since it never gets to observe those properties. To avoid dualism, codes must relate not to external properties but to internal sensorimotor models. Because this requires structured representations, neural assemblies cannot be the basis of such codes. Third, a message is informative to the extent that the reader understands its language. But the neural code is private to the encoder since only the message is communicated: each neuron speaks its own language. It follows that in the neural coding metaphor, the brain is a Tower of Babel. Finally, the relation between input signals and actions is circular; that inputs do not preexist to outputs makes the coding paradigm problematic. I conclude that the view that spikes are messages is generally not tenable. An alternative proposition is that action potentials are actions on other neurons and the environment, and neurons interact with each other rather than exchange messages.

1. Introduction

A pervasive paradigm in neuroscience is the concept of neural coding (deCharms and Zador, 2000): the query "neural coding" on Google Scholar retrieves about 15,000 papers in the last ten years. Neural coding is a communication metaphor. An example is the Morse code (Fig. 1A), which was used to transmit texts over telegraph lines: each letter is mapped to a binary sequence (dots and dashes). In analogy, visual signals are encoded into the spike trains of retinal ganglion cells (Fig. 1B). Both the Morse code and the retinal code relate to a communication problem: to communicate text messages over telegraph lines, or to communicate visual signals from the eye to the brain. This problem has been formalized by communication theory (Shannon and Weaver, 1971), a popular tool in neuroscience (Rieke et al., 1999).

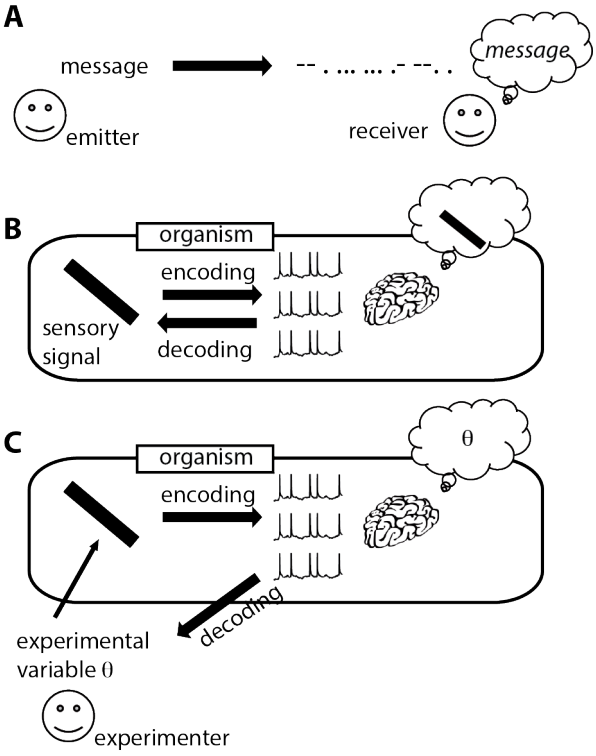

Figure 1. *The coding analogy. A, A communication channel consists of an emitter who wants to transmit some message to a receiver, in an altered form named "code" (here Morse code). The receiver knows the correspondence and can reconstruct or "decode" the original message. B, In analogy, sensory signals are encoded in the spike trains of neurons. C. In an experimental context, the emitter is the experimenter, who presents a stimulus (oriented bar) characterized by some parameter (orientation θ). The brain receives the message in the form of neural activity, from which it infers information about the stimulus (θ).*

However, the neural coding metaphor is used more broadly, in ways that are less obviously related to communication problems. For example, neurons in the primary visual cortex encode the orientation of bars in their firing rate (Fig. 1C); neurons in the auditory brainstem encode the spatial position of sounds (Ashida and Carr, 2011), and neurons in the hippocampus encode the animal's location (Moser et al., 2008). The coding metaphor applies to these situations in the sense that there is a correspondence between neural activity and some measurable property. In this sense, the metaphor applies equally well to any situation where two measurable properties co-vary. But as the linguists Lakoff and Johnson (2008) have argued, the metaphors that pervade our language are not neutral; on the contrary, they form the architecture of our conceptual system. For example, seeing the heart as a pump is quite different from seeing it as an information processing device encoding walking speed in its beat rate, two equally applicable metaphors. Is perception a communication problem? And if so, with whom?

This critique of the neural coding metaphor is articulated as follows. First, the coding metaphor is often applied in a way that the encoded property is not a sensory signal but an experimental parameter, as when interpreting tuning curves. Casting this situation into a communication problem is problematic because the communication channel is made of both the biological system and the experimental system. One may then speak of a code, but perhaps not of a *neural* code. Second, in the coding metaphor, the encoded message is information in the sense that the receiver can "decode" it into the original message, but unless we postulate the existence of a homunculus who maps neural activity to an external description of the world, this is not the situation faced by the brain. All that the nervous system ever gets to observe is the Morse code. Thus, coding usually refers to an extrinsic notion of information, while a more biologically relevant notion should be intrinsic. Finally, I will end with a more radical criticism of the neural coding metaphor, based on the circular relation between sensory signals and actions and the fact that no clear separation exists in the brain between representation and computation. This discussion will lead us to the question: are spikes messages or actions?

2. Contextual codes

2.1. The fallacy of the overwise neuron

In a reflection on theories of the brain, Francis Crick warned against the "fallacy of the overwise neuron" (Crick, 1979) that arises when the neural coding metaphor is combined with methodological reductionism. He took the example of color perception. Cones are tuned to a particular wavelength (Schnapf et al., 1987): in an experiment where light of different wavelengths is flashed, the amplitude of the transduced current varies systematically with wavelength (Fig. 2A). Thus, the current encodes wavelength, in the sense that one can recover wavelength from the magnitude of the current. However, the code is contingent on the context of the experiment. In a natural situation, the current may also vary with light intensity and therefore does not provide unambiguous information about wavelength. Indeed, animals or humans with a single functional type of cones are color blind. Thus, from the experimenter's viewpoint, it can be said that photoreceptor current encodes wavelength, yet from the viewpoint of the organism there is no usable information about color in that current. Thus, applied in this way, the coding metaphor is misleading.

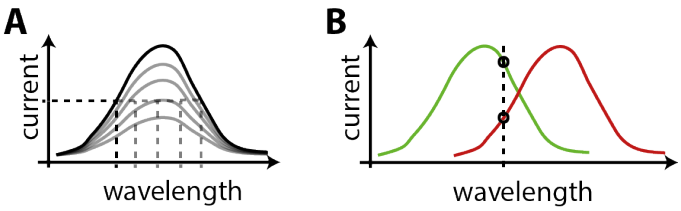

Figure 2. *Encoding wavelength of light. A, Response of a cone to flashed light as a function of wavelength (cartoon), at different intensities (grey). If intensity is fixed, wavelength can be inferred from transduced current. Otherwise, current is not informative about wavelength. B, The relative response of cones with different tunings may provide intensity-invariant information about wavelength.*

The reason is that, since the encoded property is an experimental parameter, the communication channel implicitly consists of both the biological system and the experimental system (Fig. 1C). This makes the code contextual, contingent on the specifics of the experiment. If the code is contextual, then the coding metaphor can just as well be applied to an element of context. For example, if intensity rather than wavelength were varied, then the current would encode light intensity. Thus, if it can be said that cones encode wavelength, then it can also be said that cones encode intensity, flickering frequency, and any other stimulus parameter that an experimenter can vary. Used in this sense, coding is a property not of neurons but of a particular experimental situation, which merely refers to neural sensitivity to an experimental parameter. Thus, it is misleading to speak of a neural code when one should speak of neural sensitivity (Brette, 2010).

When a message is transmitted in Morse code, the encoded message constitutes information for the receiver in the sense that the receiver can reconstruct the original message from the encoded message, a process called "decoding". If the decoder depends on the specifics of the experiment, then the message constitutes information for the experimenter, not for the organism. Thus, one can speak of a neural code for some dimension to the extent that the implied decoder is insensitive to other dimensions (but not necessarily the activity of individual neurons). For example, wavelength may be encoded in the relative activity of cones with different tunings, provided that this is invariant to light intensity (Fig. 2B).

2.2. Implications for population coding

The reductionist approach to coding is more than an inappropriate use of words. For example, in mammals, the major cue for sound localization in the horizontal plane is the difference in arrival times of the sound wave at the two ears (interaural time difference or ITD) (Fig. 3A). There are neurons in the auditory brainstem that are very sensitive to that cue (Joris et al., 1998): when a sound is played through earphones and the ITD is varied, the firing rate of those neurons changes (Fig. 3B). Thus, these neurons encode ITD in the same way as cones encode wavelength. In this sense, because the relation between ITD and firing rate is steep, a single neuron can in fact encode ITD with the same accuracy as the entire organism, as assessed by behavioral experiments (Skottun, 1998; Shackleton et al., 2003). But as we have seen, this comparison is highly misleading, because the code is contextual: playing a different sound (for example a louder or a higher sound) would also change the activity of these neurons, and therefore the firing rate of a single neuron may be informative for the experimenter, but not for the organism.

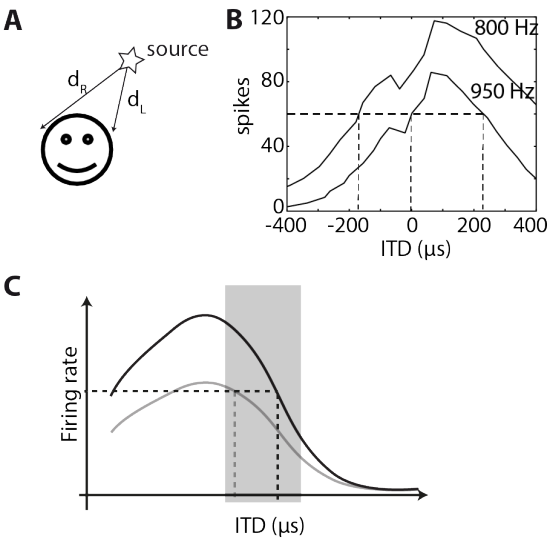

Figure 3. *Encoding sound location. A, A major cue for sound localization is the interaural time difference or ITD, d_R - d_L. B, Number of spikes in response to two binaural tones (950 Hz and 800 Hz) as a function of ITD, for the same neuron in the medial superior olive of a cat (digitized from (Yin and Chan, 1990), Fig. 10). It is possible to infer ITD from spike count if the experimental configuration (presented tone) is known, not if the sound is a priori unknown. C, If the organism lived in a world with a single sound played at different ITDs, then the best way to encode ITD would be with a neuron tuned to an ITD outside the physiological range (shaded), so that the selectivity curve is steep inside that range. However, the response of a single neuron is fundamentally ambiguous when sounds are diverse, irrespective of the steepness of the curve (selectivity curve for another sound shown in grey).*

This misuse of the coding metaphor leads to incorrect conclusions about population coding. If the confounding dimensions (level, etc) are neglected, then the best way to encode ITD is to have a steep monotonous relation between ITD and firing rate, that is, to maximize neural sensitivity to ITD (Fig. 3C).

Thus, the neuron's preferred ITD should lie outside the range of natural sounds (around ±800 μs for humans (Benichoux et al., 2016)) while the steepest slope of the selectivity curve should be inside, which leads to two symmetrical optimal selectivity curves (Harper and McAlpine, 2004). This is the concept of "slope coding", which motivates an influential model of sound localization (Grothe et al., 2010), and was meant to explain why many neurons have their preferred ITD outside the natural range (McAlpine et al., 2001) (even though many are also inside, especially in larger mammals such as cats (Goodman et al., 2013, Fig. 1a)).

Unfortunately, this conclusion is incorrect when confounding dimensions are not neglected. To encode wavelength independently of intensity, what is required is not a steep relation between wavelength and current, but cones with diverse tunings. In the same way, modeling work shows that heterogeneity of ITD tunings is crucial to raise the ambiguities due to non-spatial dimensions of sounds, so that the slope coding model cannot account for behavioral accuracy (Brette, 2010; Goodman et al., 2013).

Thus, the reductionist approach to coding leads to incorrect conclusions because it focuses exclusively on neural sensitivity. While neural sensitivity is not a sufficient property, it could be argued that it is a necessary one. Unfortunately, this is not true either. For example, no auditory nerve fiber is sensitive to ITD because their responses are exclusively monaural. Yet, the joint response of auditory nerve fibers on opposite sides is informative about ITD, in the relative timing of their spikes. In general, unless confounding dimensions are ignored, there is no direct relation between neural sensitivity and information in the activity of a neural population.

2.3. Can neurons encode experimental parameters?

For the coding metaphor to be biologically meaningful, the decoder (explicit or implied) must not be experiment-specific. An immediate implication is that the encoded property must be meaningful outside of the experimental context. Such is the case when the coding metaphor is applied to a transduction problem, for example when neural activity encodes visual signals (Laughlin, 1981). This is less straightforward when the quantity to be decoded is a variable of the experimental protocol rather than the sensory signal, as when tuning curves are measured. Crick noted that the activity of a single cone cannot encode wavelength in any meaningful sense because of confounding factors. But the more profound reason is that natural light is not monochromatic, it has a continuous spectrum. In reality, the transduced current depends on the convolution of the spectrum of incident light with the absorption spectrum of the photoreceptor. Cones do not encode wavelength because there is no such thing as the wavelength of a patch of visual scene.

Do V1 neurons encode the orientation of bars and gratings? The answer completely depends on whether the world is made of bars and gratings. If not, then the proposition must be made more precise, for example: V1 neurons encode the direction of the local gradient of the visual field. But this proposition cannot be established from a simple tuning curve experiment. It turns out that other types of experiments show that it is not true, because the response of those neurons also depends on the surround; typically, responses are suppressed when stimulus properties at the center (e.g. orientation) match those in the surround (Hubel and Wiesel, 1968; Bolz and Gilbert, 1986). Thus, it is not so clear what V1 neurons encode; perhaps how surprising a patch of visual scene is given its surround (Rao and Ballard, 1999).

Thus, the first requirement for the coding metaphor to be meaningful is that the encoded property is meaningful outside of the experimental context, whether it is the physical sensory signal or a more abstract property.

2.4. Can contextual codes make sense?

Place cells in the hippocampus encode the spatial position of the organism within a given environment (Moser et al., 2008). The code is contextual, because it is specific of each environment. However, in contrast with the color and sound localization examples, this does not make the coding metaphor irrelevant. Indeed, the issue in those examples is not so much that the code is contextual, but that the context lives outside of the organism, within the design of the experiment. A rat lives in an environment, which has some persistence. The fact that the place code is contextual implies that the activity of place cells does not provide information of the type "I am at position (48°, 2°)", but it does provide information of the type "I am in the same corner as a few minutes earlier". In addition, a key argument to justify the name of "place cell" was that the responses of those cells depend on spatial position but not on specific non-spatial features of the environment, for example light intensity - they are not merely sensitive to spatial position (O'Keefe, 1976).

In contrast, cone activity does not provide information about wavelength unless the organism knows that a monochromatic light is being presented and additionally knows how cone activity depends on wavelength. The latter point is crucial: since the organism never gets to observe stimulus wavelength but only the activity of cones, there is no way that it can infer the relation between these two observables.

In the same way, in our sound localization example, the firing rate of a single neuron provides information about the spatial position of a sound, but only if one knows which sound is being played, and what the relation between position and firing rate is for that particular sound. Thus, the code is specific of a particular sound, just as place cell activity is specific of an environment. But while a rat can explore its environment, one does not explore a sound, because a sound is transient. The initial proposition must then be qualified: firing rate encodes sound position within a given world of identical sounds. This makes the coding metaphor much less meaningful than in the case of place cells.

One may raise the following objection. The code for sound position is contextual, but it could be that context is provided by another part of the nervous system: there could be a "where" pathway and a "what" pathway, with the former making sense in the context of the latter. This hypothesis implies that there is no information *per se* in the where pathway, but only in the joint activity of the two pathways. It follows that the proposition is logically equivalent to the following: sound position is encoded in the activity of the nervous system. Indeed, consider the following proposition: sound location is encoded in the activity of the left auditory nerve (since firing rate varies with level and thus with sound position), and the context is provided by the right auditory nerve. This is true since sound location is encoded in the joint activity of the two auditory nerves, but separating code and context is arbitrary and misleading.

In summary, when the coding metaphor is applied, the dependence of the code on context must be identified. If the code depends on specifics of the experimental design, then it constitutes information for the experimenter, not for the organism. In this case, it is misleading to speak of a "neural" code because the communication channel implied by the metaphor is made of both the biological system and the experimental system.

3. Extrinsic vs. intrinsic information

3.1. The neural coding metaphor is implicitly dualist

In what sense is the neural code "information" about objective properties of the world? It is information in the sense that these properties can be inferred from neural activity. Methodologically, this inference is done by the experimenter, who confronts these properties with measurements of neural activity. But by using the terms "neural code", we imply that the brain must also do this inference. How is it possible for the nervous system to infer external properties from neural activity, if all that it ever gets to observe is that activity? In fact, what does it even mean that a neural network infers external properties (e.g. the direction of a sound source), given that those properties do not belong to the domain of neural activity?

This hints at the homunculus fallacy. In the coding paradigm, meaning is obtained at some final stage where neural activity is mapped to the initial message, where the world is described in external terms. This leads to what philosopher Daniel Dennett called the "Cartesian theater" (Dennett, 1992). Descartes suggested that the mind, which he viewed as non-material, interacted with the brain at the pineal gland. At that stage, the mind would read the activity of the nerves and interpret it in terms of the outside physical world. The "Cartesian theater" is the brain seen as a screen where the world is projected, that a homunculus (the mind) watches. Pushed to its ultimate conclusions, the coding view of perception is logically equivalent to the dualism of Descartes.

I suggest that the appeal of the coding metaphor for the brain stems from the misleading use of the word "information". In the coding metaphor, neural activity is information about the world in the sense that it is possible to reconstruct the external world from neural activity, just as it is possible to map a Morse message back to the original message. The problem is, a Morse message translated from German is not informative unless one understands German, and it is even less informative if all that the receiver ever gets to observe is Morse messages. To decipher the Enigma code during World War II, the Allies used the weather of the day as an element that they could look for in the encoded messages. Thus, they used a known element external to the messages (a "known-plaintext attack"), in addition to their prior knowledge of German.

Thus, the fundamental problem with the coding metaphor, as it applies to the brain, is that it conveys an *extrinsic* notion of information: the meaning conferred to symbols is extrinsic, given by an external observer who has access to both the external world and neural activity. This meaning is not intrinsic to the activity of the nervous system. Thus, it is not immediately obvious that this extrinsic notion of information is the right way to address the informational problems faced by the brain. This problem is known as the *symbol grounding problem*: how do spikes, the symbols of the neural code, make sense for the organism?

3.2. Information as subjective laws

How can there be any information about the world without direct access to the world? This is the question to be answered if we are to address the implicit dualism of the neural coding metaphor. It has been addressed by philosophers and psychologists. O'Regan and Noë (2001) proposed the analogy of the "villainous monster". Imagine you are exploring the sea with an underwater vessel. But a villainous monster mixes all the cables and so all the sensors and actuators are now related to the external world in a new way. How can you know anything about the world? The only way is to analyze the structure of sensor data and their relationships with actions that you can perform. If dualism is rejected, then this is the kind of information that is available to the nervous system. A salient feature of this notion of information is that, contrary to Shannon's information, it is defined not as elements of a set but as relations or logical propositions: if I do action A, then sensory property B happens; if sensory property A happens, then another property B will happen next; if I do action A in sensory context B, then C happens.

James Gibson previously developed a related psychological theory (Gibson, 1986). While criticizing the information-processing view of perception, he argued that there is information about the world present in the *invariant structure* of sensory signals: "A great many properties of the [optical] array are lawfully or regularly variant with change of observation point, and this means that in each case a property defined by the law is invariant". Clearly, the word "information" is not meant in the sense of communication theory, but rather in the sense of scientific knowledge. A set of observations and experiments provides information about the world, in the form of laws that relate observables (sensory signals) between them and with possible actions. This form of information is intrinsic; I proposed to call this set of laws the subjective physics of the world (Brette, 2016). A related view, formalized by theoretical biologist Robert Rosen (Rosen, 1985), is that biological organisms build an internal model of the world, in which the variables are sensory signals. This view addresses the symbol grounding problem by mapping sensory signals to elements of an

internal model. The signals make sense in reference to that model; they are not mapped to externally defined properties.

To be more concrete, imagine an organism with a linear retina in a world of tiny objects (Fig. 4A). The presence of an object is signaled by the firing of a retinal neuron, whose identity indicates the object's location through a "labelled-line code". However, to infer location from neural activity requires knowing the relation between object position and neuron identity, and this information is not conveyed by the flickering of the cells. Therefore, retinal activity alone does not constitute spatial information for the organism. Now suppose that the organism can move its retina laterally. It can then observe a relation between muscular state (through proprioception or afferent copy) and neural activity (Fig. 4B). Neural activity is now informative of which neuron would become active if a given action were performed. Thus, neural activity is intrinsic information to the extent that it can be mapped to a rich internal model. We note that this model still entails a very poor notion of space, with no metric structure. A richer notion would require the ability to perform different types of movements, in particular movements which translate the entire organism (see Brette (2016)). The key point in this example is that neural activity conveys the same extrinsic information about the object's location whether the organism can move or not, yet there is intrinsic information in the latter case but not in the former.

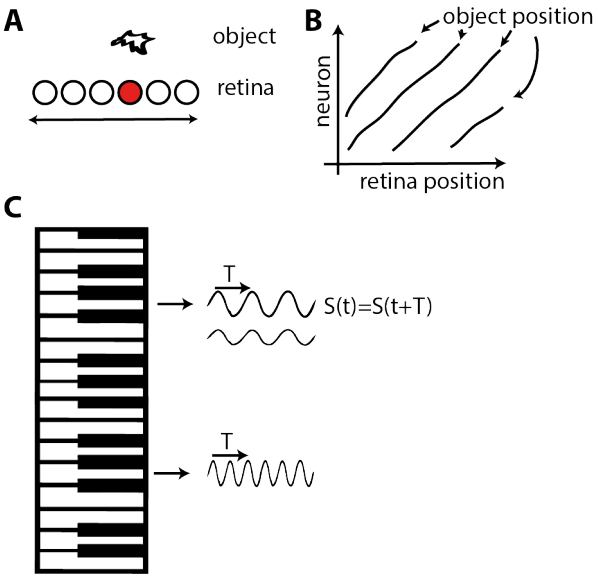

Figure 4. *Information in sensorimotor structure. A, Each neuron on a linear retina fires when an object is presented at a specific location. From the brain's viewpoint, these are blinks from different cells with no intrinsic notion of space. B, If the organism can move the retina, then object position is defined as the relation between neural activity and proprioceptive state of the retina. The firing of a cell informs the organism of which cell would fire if the retina were moved in particular directions. C, Piano keys are encoded into the acoustical signals. Beyond Shannon information, there is intrinsic information in the form of sensory laws followed by the signals. For example, the signal produced by a key is periodic, with a specific period, independently of how hard the key is struck. The same sensory law is obeyed by the signal produced by a key an octave higher, providing a natural topology.*

3.3. Information in a single stimulus

In the coding metaphor, a code constitutes information in the sense of a translation dictionary. But there is more information in the sensory signals than implied by this metaphor. Imagine a world of sounds made by various piano keys (Fig. 4C). From the coding point of view, there is extrinsic information in the acoustical signals about which piano key is played, in the sense that there is a one-to-one mapping between piano keys

and signals. From the organism's viewpoint, this information is poor since the mapping is unknown. However, there is much more information in these sensory signals than what the classical coding paradigm suggests, if we take the viewpoint of information as sensory laws. It can be said for example that the acoustical signals S(t) follow the law S(t+T) = S(t) for some delay T. The striking difference with the extrinsic notion of information conveyed by the coding metaphor is that something can be said of a *single* stimulus: a (tentative) model of a single stimulus can be built – despite the problem of induction, building a model from finite observations is tentative but possible to the extent that science is possible (see (Brette, 2016) for a discussion). This relates to another theoretical concept of information known as *Kolmogorov complexity*. Kolmogorov complexity is the size of the shortest program that produces the signal. In this case, the program could be to repeat the waveform of one period. This program is a type of information that is absent from the coding metaphor, which considers information in a set-theoretic sense: a stimulus is just one element of a set of possible stimuli.

But a stimulus is not just an element of a set: it can have structure (laws or generative program) and this structure constitutes information. In particular, the structure has predictive value: a law postulated on the basis of the beginning of a signal can be tested on the rest of the signal; or a program can produce the rest of the signal. Structure also defines relations between different stimuli. For example, consider two musical notes that differ by an octave (Fig. 4C). In the coding view, these are just two distinct labels. However, there is a relation between the two signals stemming from the fact that the period of the lower note is twice the period of the higher note. This implies that the sensory model associated with the lower note (S(t) = S(t+T)) is also valid for the higher note (but not conversely), which defines a relation of similarity between these two notes. This topology does not exist in the classical coding view, because each signal is seen as an element of an unstructured set. It turns out that these similarities in structure match perceptual similarities, with some qualifications (Zheng and Brette, 2017): two notes played by different instruments elicit the same pitch if they have the same periodicity but different spectral content.

3.4. Encoding intrinsic information

In the neural coding literature, variations of the coding paradigm have been proposed to address the question of intrinsic information. An interesting proposition is predictive information, which is the mutual (Shannon) information between the past and future of a signal (Bialek et al., 2001; Palmer et al., 2015). This should not to be confused with "predictive coding" (Rao and Ballard, 1999; Clark, 2013), which is a variation of the efficient coding paradigm where neurons encode the difference between the signal and the expectation of the signal based on other observations (either previous observations or the activity of other neurons). Suppose for example that there is some acoustical noise in addition to the musical notes. In the predictive coding paradigm, neurons encode only the difference with expectations, i.e., mostly the noise. It is considered efficient because the code is meant to stand for the original signal, which includes noise. This is paradoxical because noise cannot constitute information for the organism in any meaningful sense. This paradox occurs because the predictive coding paradigm (and more generally the efficient coding paradigm) does not distinguish between what is meaningful and what constitutes noise. Predictive information improves on this by proposing that what should be encoded is only the part of the signal that is informative about the future. In our example, predictive information would only include information about the piano key being played and not acoustical noise.

However, predictive information comes with the same limitation as more classical forms of coding in that what is encoded is not the structure of signals; the encoded messages are note labels. The prediction of the future signal can be made by the encoder, but not by those who read the encoded message. The fact that two notes separated by an octave are structurally similar is not communicated. Thus, the fundamental problem in this and more classical coding paradigms is that all the meaningful structure (the sensory model or the program) is private to the encoding process, while only the encoded message is communicated.

3.5. Representation of structure and structure of representations

Thus, for the coding metaphor to be biologically meaningful, this intrinsic information, the structure of signals, should be encoded and communicated. There are in fact a few examples in the neuroscience literature, although they are not usually cast in this way. One is the Jeffress model of ITD coding (Jeffress, 1948) (Figure 5A). In that model, neurons receive inputs from monaural neurons on the two sides, with different conduction delays. When input spikes arrive simultaneously, the neuron spikes. Thus, the neuron spikes when the two acoustical signals at the two ears are such that $S_L(t) = S_R(t-d)$, where d is the conduction delay mismatch between the two ears. Physically, this corresponds to a sound source placed at a position such that it produces an ITD equal to d. In this model, the neuron encodes a sensory law; more precisely, it indicates whether signals satisfy a particular sensory law. Its spikes cannot be used to reconstruct the original signals.

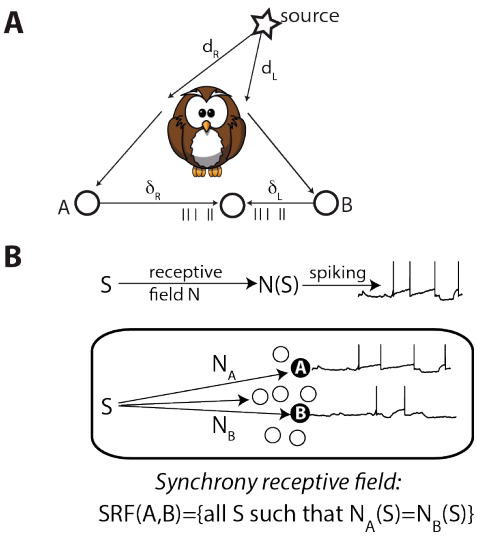

Figure 5. *Neural representation of structure (adapted from Brette (2012)). A, The Jeffress model of sound localization. The sound arrives at the two ears with delays d_L and d_R. It is then transduced into spike trains that arrive at a binaural neuron with delays δ_L and δ_R. Synchrony occurs when $d_R - d_L = \delta_L - \delta_R$, making the neuron fire. B, The synchrony receptive field. The response of a neuron to a stimulus is described as filtering of the sensory signal S through the receptive field N, followed by spiking. The synchrony receptive field of two neurons A and B with different receptive fields N_A and N_B is defined as the set of stimuli that elicit synchronous responses in these neurons.*

This model can be generalized. I recently proposed a computational theory of synchrony in sensory systems (Brette, 2012), in which synchrony reflects the structure of sensory signals (Fig. 5B). One considers two neurons A and B which convert their time-varying inputs into precisely timed spike trains, where their inputs are seen as transformed versions $N_A(S)$ and $N_B(S)$ of the stimulus S (N_A and N_B are fixed and correspond to the receptive fields of the neurons). Synchrony between A and B then reflects the sensory law $N_A(S) = N_B(S)$. The *synchrony receptive field* of A and B is the set of stimuli that elicit synchronous responses in these two neurons: it is the set of sensory signals that satisfy a particular law. A neuron receiving inputs from A and B would then fire when the stimulus is in the synchrony receptive field of A and B, that is, when it satisfies a particular sensory law. One can then see synchrony patterns or the firing of postsynaptic neurons as a code for sensory laws. This framework has been applied to pitch perception (Laudanski et al., 2014) and to sound localization in realistic environments where sounds are diffracted by the head (Goodman and Brette, 2010; Benichoux et al., 2015).

However, the concept of synchrony receptive fields captures a rather elementary notion of structure. For example, it does not capture predictive sensorimotor propositions such as: "if I do action A, then property B should be true". In fact, when perceptual information is understood as propositions about sensorimotor signals, it appears that the information to be represented by the brain is much richer than suggested by coding theory. For example, in a visual scene, there could be Paul, a person I know, wearing a new shirt, driving a car (Fig. 6). What is important here is that a scene is not just a "bag of objects": objects have relationships with each other, and there are many possible different relationships. For example there is a car and there is Paul, and Paul is in a specific relationship with the car, both a physical relationship (a particular posture within the car) and a functional relationship (driving it). Thus, a neural "code" of a scene should not only represent the individual objects but also specific relationships between these objects, possibly in a recursive way. Since phenomenal experience is at least as complex as what can be consciously described with language, neural representations of perception should be at least as structured as language: it should have syntax, as recently proposed (Buzsáki, 2010).

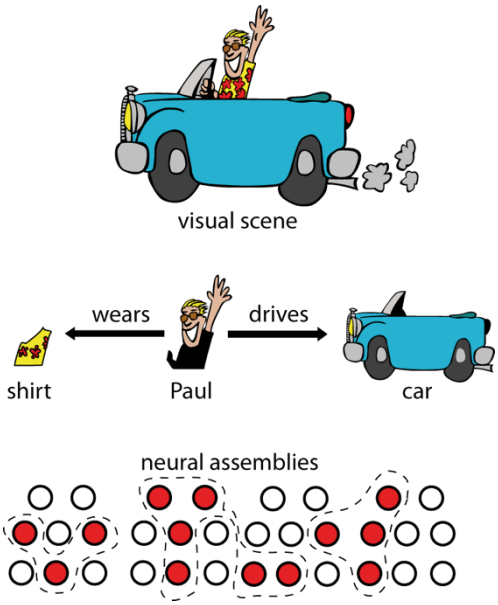

Figure 6. *Perceptual scenes are highly structured. For example, there is Paul (person I know), driving a car, and wearing a new shirt. Representing this scene by the firing of neural assemblies raises two issues: 1) it may be difficult to split active neurons into the correct assemblies (superposition catastrophe), and more importantly 2) the structure of the scene (relations shown by arrows) cannot be represented in this way.*

Unfortunately, this does not fit well with the concept of "neural assemblies", which is the mainstream assumption about how things we perceive are represented in the brain. If it is true that any given object is represented by the firing of a given assembly of neurons, then several objects should be represented by the firing of a bigger assembly of neurons, the union of all assemblies, one for each object. Several authors have noted that this may lead to the "superposition catastrophe" (von der Malsburg, 1999), i.e., there may be different sets of objects whose representations are fused into the same big assembly. Even without this segmentation problem, the representation of a scene could still be nothing else than an unstructured "bag of objects", for there are no relationships between objects in the assembly representation. Mathematically, the appropriate representation is a graph with labeled edges and vertices, not an unstructured set.

Thus there is a fundamental problem with the concept of neural assembly, which is that there is no representation of relations, only of things to be related. This remark has been made in the past, essentially in the context of the binding problem. It has led several authors to postulate that synchrony is used to bind the features of an object represented by neural firing (Singer, 1999; von der Malsburg, 1999). This avoids the superposition catastrophe because at a given time, only one object is represented by neural firing.

However, this is insufficient because only one type of relation can be represented in this way, and a symmetrical one: does Paul drive the car, or does the car run over Paul?

At this point, there is no strong alternative representational theory that would solve all these issues. The analogy of language shows nevertheless that this is possible in principle: it is conceptually possible to represent structures as complex as linguistic structure by using the order of words; in the same way, it is conceptually possible to represent sensory structure by using the order of spikes.

To conclude this part, if the neural code is to be meaningful for the organism, then what is encoded should not be objective properties of the world but internal sensorimotor models. This implies that neural representations themselves must be structured, rather than "bags of neurons".

4. Spikes: messages or actions?

4.1. The brain as a Tower of Babel

In his seminal book on vision, David Marr (1982) wrote: "If we are capable of knowing what is where in the world, our brains must somehow be capable of representing this information.". That is, if we consider that perception emerges from states of matter, then there must be a correspondence between mental representations and neural representations, states of the brain. We have argued that the kind of representations that the brain manipulates must be structured representations of internal sensorimotor models, not objective representations from the perspective of an external observer.

Despite this shift in perspective, seeing neural activity as messages remains problematic. A message is something that a reader makes sense of. But to make sense of a message requires understanding the language of the message. In the coding metaphor, that language is the mapping from original to encoded message, and it is private to the encoder, not communicated. Thus, the receiver of the message cannot be presumed to read it unless it knows that language. Mathematically, original message = encoded message + code, but only one term of this equation is communicated. If we consider in addition that those messages are coded efficiently, then each neuron uses its own private language on the basis of its personal history of inputs.

Thus, in the neural coding metaphor, the code is private to each neuron. If we follow this metaphor, this means that all neurons speak a different language, a language that might allow expressing concepts very concisely but that no one else can understand. Thus, the neural coding metaphor leads to the conclusion that the brain is a Tower of Babel.

These difficulties arise when neural activity is seen as messages. Another possible view is that neural activity is indeed activity, and action potentials are actions. Each spike has an effect on target neurons; it pushes other neurons in some direction. Does the heart send messages to the blood about how fast the organism runs, or does it move blood in the vascular system? Do neurons read each other's spike trains, or do they interact with each other?

4.2. Coding vs. acting

I will now discuss a concrete biological example where both the coding and acting metaphors can be applied. Paramecium is a unicellular organism that swims in stagnant fresh water using cilia and feeds on bacteria. It uses different kinds of sensory signals, including mechanical signals to avoid obstacles and chemical signals to localize food (Jennings, 1906). To a first approximation, it alternates between straight courses and sudden random changes in direction (Fig. 7A). It turns out that each change in direction is triggered by a spike produced by voltage-gated calcium channels (Fig. 7B) (Eckert, 1972). To find a chemical source,

Paramecium uses a simple method: when concentration decreases, the membrane is depolarized by chemical receptors and a spike is produced (with some stochasticity), triggering a change of direction (similar to chemotaxis in E. Coli) . This is of course a simplified description of Paramecium physiology and behavior, but for the sake of the argument we shall consider an organism that functions in this simple way.

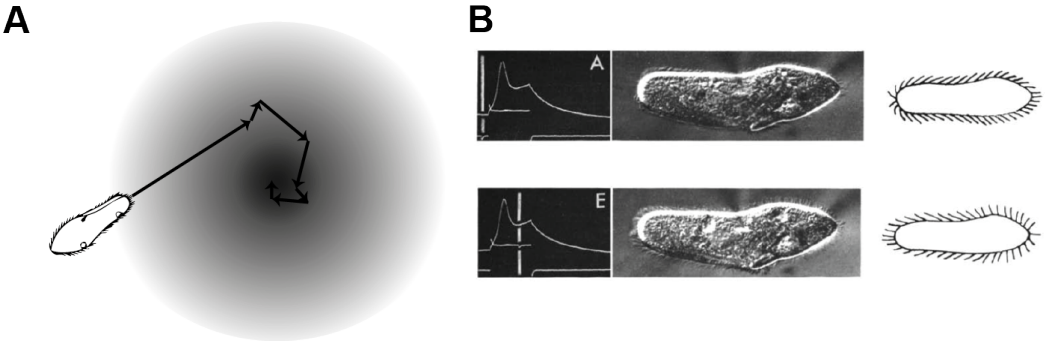

Figure 7. *Spatial cognition in Paramecium, a "swimming neuron". A, Paramecium finds a chemical source by switching to a new random direction when concentration decreases. B, Each direction change is triggered by an action potential, which transiently inverts cilia beating through a calcium pathway (adapted from (Eckert and Naitoh, 1970)).*

Thus, Paramecium is a sort of swimming neuron, where spikes can be seen both as encoding sensory signals (concentration) and as producing an action (change of direction). The same could be said of any neuron in a nervous system, except the action of a neuron is generally on other neurons and only indirectly on the environment. We can now expose the difficulties of the coding metaphor in this context.

We may argue that if the organism can navigate efficiently in its environment, then the spikes must contain information about that environment. Therefore, it makes sense to apply the coding metaphor to this situation. How should sensory signals be encoded into spikes? This question raises two issues. First, the primary goal of the organism is not to maximize information but to take appropriate action. But is coding efficiency the same as behavioral performance? In general no, because coding efficiency only depends on the sensory signals while behavioral performance depends on what you want to do: to move towards or away from a source, to look for food or to sleep, or to look for a mate. Therefore, an efficient code (in terms of signal representation) is not in general a good code (in terms of behavior).

There is a second fundamental issue. The input to the organism is not the environment but the signals it captures. It follows that sensory signals are not given *a priori*: they depend on the organism's actions, and therefore on the encoding. Thus, the very notion of coding is ill-defined because the relation between "input" and "output" is not a mapping but a circular relation. This affects the coding framework in two ways. First, if we consider organism and environment as two coupled dynamical systems, then what can be said is that the organism's action at a given instant depend on the previous sensory signals, and possibly on internal states. But the input signal is not mapped as a whole to an output signal, because the signal is not given a priori independently of the outputs. Second, one may define an efficient code on the basis of the statistics of the sensory signal, but choosing that code changes the sensory signal. Thus, it is not possible to answer the question: "what is the most efficient way to encode the sensory signals?". A possible solution would be to see it as a fixed-point problem, for example by iteratively looking for an efficient code until signal statistics are stable. But this would lead to cells looking for regions of space where sensory signals are easiest to reconstruct, with no guarantee that it matches the organism's goals. Although the circularity is particularly evident in the Paramecium example, it is a general feature of biological organisms (Ahissar and Assa, 2016); for example, visual signals both trigger and depend on eye movements.

This example illustrates the fact that spikes are not just messages. They are an integral part of the sensorimotor loop, with each spike produced by a neuron triggering changes on the neighboring network,

on the environment and ultimately on its input. Thus, failures of efficient coding may occur not because the brain is not efficient, but because the brain does not really code.

5. Conclusion

A metaphor is not true or false: it is insightful or misleading (Lakoff and Johnson, 2008). How insightful is the coding metaphor for the brain?

There are two ways in which the neural coding metaphor is used. One is when we say that neurons encode sensory signals, for example when we say that retinal spike trains encode visual signals. Discussions of efficient coding, predictive coding and predictive information belong to this category. In essence, the proposition means nothing more than there is a mapping from X to neural activity, and we are interested in properties of that mapping. Thus, it implicitly frames the problem of perception as one of representing inputs that are given to the organism. But inputs are not given to the organism. On the contrary, the organism actively captures sensory signals in order to obtain information about the environment. This makes the relation between sensory signals and neural activity circular, for any neuron along the sensorimotor loop. Consequently, normative discussions framed exclusively in terms of coding are problematic (*how should neurons encode sensory signals?*), because inputs are not independent from outputs, and coding efficiency is not *a priori* the same as behavioral performance.

Presumably, an organism is not so much interested in reconstructing the sensory signals as in obtaining information about the environment, and capturing signals is only a means to that end. Thus, a second way to use the neural coding metaphor is when we say that neurons encode some abstract property of the environment. Such propositions are highly misleading when the property is an experimental parameter, because in that case the communication channel is made of both the biological and experimental systems. If the interpretation of the neural code depends on experimental context, then the code is not neural after all, but rather neuro-experimental. This qualification certainly affects the insightfulness of the metaphor for understanding brain function.

Thus, speaking of a neural code for an abstract property is meaningful to the extent that the implied decoder is insensitive to elements of context. But who reads the neural code? If we consider that neurons encode objective properties of the world and the brain reads the code, then the neural coding metaphor is dualist. What does it mean to say that neurons encode the direction of a sound source, does the brain know about radians? To avoid the homunculus fallacy, what is to be encoded and manipulated cannot be objective properties (radians) but internal models that relate sensory signals together and with the organism's actions. It follows that a biologically relevant concept of information cannot be Shannon's unstructured notion of information (an input as an element of an unstructured set). Intrinsic information must take the form of structured models, in which not only elements but also relations between elements are represented. It follows that neural assemblies ("bags of neurons") cannot be the basis of mental representations.

Ultimately, the view that spikes are messages leads to a perplexing view of neural activity. A message is only a message for someone who knows its language. But in the neural coding metaphor, each neuron has its own language (the encoding) and only the message is communicated. Thus, the brain is a Tower of Babel. Perhaps more relevant is the view that an action potential is an action on other neurons, that producing a spike is a way of doing rather than of communicating, and that neurons interact with each other and with the environment, rather than speak to each other. After all, we do speak of neural "activity".

References

Ahissar E, Assa E (2016) Perception as a closed-loop convergence process. eLife 5.

Ashida G, Carr CE (2011) Sound localization: Jeffress and beyond. Curr Opin Neurobiol 21:745–751.

Benichoux V, Fontaine B, Karino S, Joris PX, Brette R (2015) Frequency-dependent time differences between the ears are matched in neural tuning. eLife:10.7554/eLife.06072.

Benichoux V, Rébillat M, Brette R (2016) On the variation of interaural time differences with frequency. J Acoust Soc Am 139:1810.

Bialek W, Nemenman I, Tishby N (2001) Predictability, complexity, and learning. Neural Comput 13:2409–2463.

Bolz J, Gilbert CD (1986) Generation of end-inhibition in the visual cortex via interlaminar connections. Nature 320:362–365.

Brette R (2010) On the interpretation of sensitivity analyses of neural responses. J Acoust Soc Am 128:2965–2972.

Brette R (2012) Computing with neural synchrony. PLoS Comp Biol 8:e1002561.

Brette R (2016) Subjective Physics. In: Closed loop neuroscience (El Hady A, ed), pp 146–170. Academic Press.

Buzsáki G (2010) Neural Syntax: Cell Assemblies, Synapsembles, and Readers. Neuron 68:362–385.

Clark A (2013) Whatever next? Predictive brains, situated agents, and the future of cognitive science. Behav Brain Sci 36:181–204.

Crick F (1979) Thinking about the brain. Sci Am 241:219–232.

deCharms RC, Zador A (2000) Neural representation and the cortical code. Annu Rev Neurosci 23:613–647.

Dennett DC (1992) Consciousness Explained, 1st ed. Back Bay Books.

Eckert R (1972) Bioelectric Control of Ciliary Activity. Science 176:473–481.

Eckert R, Naitoh Y (1970) Passive electrical properties of Paramecium and problems of ciliary coordination. J Gen Physiol 55:467–483.

Gibson JJ (1986) The Ecological Approach to Visual Perception. Routledge.

Goodman DF, Benichoux V, Brette R (2013) Decoding neural responses to temporal cues for sound localization. eLife 2:2:e01312.

Goodman DFM, Brette R (2010) Spike-Timing-Based Computation in Sound Localization. PLoS Comput Biol 6:e1000993.

Grothe B, Pecka M, McAlpine D (2010) Mechanisms of Sound Localization in Mammals. Physiol Rev 90:983–1012.

Harper NS, McAlpine D (2004) Optimal neural population coding of an auditory spatial cue. Nature 430:682–686.

Hubel DH, Wiesel TN (1968) Receptive fields and functional architecture of monkey striate cortex. J Physiol 195:215–243.

Jeffress LA (1948) A place theory of sound localisation. J Comp Physiol Psychol 41:35.

Jennings HS (Herbert S (1906) Behavior of the lower organisms. New York, The Columbia university press, The Macmillan company, agents; [etc., etc.]. Available at: http://archive.org/details/behavioroflowero00jenn [Accessed December 20, 2015].

Joris PX, Smith PH, Yin TC (1998) Coincidence detection in the auditory system: 50 years after Jeffress. Neuron 21:1235.

Lakoff G, Johnson M (2008) Metaphors We Live By. University of Chicago Press.

Laudanski J, Zheng Y, Brette R (2014) A structural theory of pitch. eneuro:ENEURO.0033-14.2014.

Laughlin S (1981) A simple coding procedure enhances a neuron's information capacity. Z Naturforsch [C] 36:910–912.

Marr D (1982) Vision, First Edition. W.H.Freeman & Co Ltd.

McAlpine D, Jiang D, Palmer AR (2001) A neural code for low-frequency sound localization in mammals. Nat Neurosci 4:396.

Moser EI, Kropff E, Moser M-B (2008) Place Cells, Grid Cells, and the Brain's Spatial Representation System. Annu Rev Neurosci 31:69–89.

O'Keefe J (1976) Place units in the hippocampus of the freely moving rat. Exp Neurol 51:78–109.

O'Regan JK, Noë A (2001) A Sensorimotor Account of Vision and Visual Consciousness. Behav Brain Sci 24:939–973.

Palmer SE, Marre O, Berry MJ, Bialek W (2015) Predictive information in a sensory population. Proc Natl Acad Sci U S A 112:6908–6913.

Rao RP, Ballard DH (1999) Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. Nat Neurosci 2:79–87.

Rieke F, Warland D, Steveninck R de de R van, Bialek W (1999) Spikes: Exploring the Neural Code. Cambridge, Mass.; London: A Bradford Book.

Rosen R (1985) Anticipatory Systems: Philosophical, Mathematical and Methodological Foundations, 1st edition. Oxford u.a.: Pergamon Pr.

Schnapf JL, Kraft TW, Baylor DA (1987) Spectral sensitivity of human cone photoreceptors. Nature 325:439–441.

Shackleton TM, Skottun BC, Arnott RH, Palmer AR (2003) Interaural Time Difference Discrimination Thresholds for Single Neurons in the Inferior Colliculus of Guinea Pigs. J Neurosci 23:716–724.

Shannon CE, Weaver W (1971) The Mathematical Theory of Communication. University of Illinois Press.

Singer W (1999) Neuronal synchrony: a versatile code for the definition of relations? Neuron 24:49–65.

Skottun BC (1998) Sound localization and neurons. Nature 393:531.

von der Malsburg C (1999) The what and why of binding: the modeler's perspective. Neuron 24:95–104.

Yin TC, Chan JC (1990) Interaural time sensitivity in medial superior olive of cat. J Neurophysiol 64:465–488.

Zheng Y, Brette R (2017) On the relation between pitch and level. Hear Res 348:63–69.