Original Paper

Adaptive Behavior

Adaptive Behavior
0(0) 1–9

DOI: 10.1177/1059712313483144
adb.sagepub.com
SAGE

Towards an enactive account of action: speaking and joint speaking as exemplary domains

Fred Cummins

Abstract

Sense-making, within enactive theories, provides a novel way of understanding how a comprehensible and manageable world arises for a subject. Elaboration of the concept of sense-making allows a fundamental reframing of the notion of perception that does not rely on the pick up of information about a pre-given world. In rejecting the notion of the subject as an input/output system, it is also necessary to reframe the scientific account of skilled action. Taking speech as an exemplary domain, I here present the outline of an enactive account of skilled action that is continuous with the concept of sense-making. Extending this account to the rich domain of joint or synchronous speaking allows many of the principal themes of the emerging enactive account to be considered as they relate to a familiar and important human practice.

Keywords

AQ1

1 Introduction

The enactive approach to understanding the relation between experience and the world is motivated, in part, by a desire to avoid dualist mediated epistemologies of the kind so aptly caricatured in the film *Being John Malkovich*, in which an inner subject peers out at, and tries to make sense of, an external world. Doing so demands that we question our linguistic habits that have a tendency to return to the language of "inner" experience contrasted with "outer" world.

The language of mediated experience of the world runs very deep indeed. This is the "picture" that Wittgenstein says "held us captive" (Wittgenstein, 1973, §115), and, despite the best efforts of Gibson, Merleau-Ponty, Heidegger, Varela and many others, it continues to lurk as the framework within which conventional cognitive science is couched. It pre-dates Descartes, and it insinuated itself into the emerging discipline of psychology at a very early stage, in the form of the unifying concept of the reflex arc, with the world providing input at one end, action appearing at the other, and the subject presumed to lurk in between. John Dewey lamented the reliance on this linear throughput system thus:

> "[T]he reflex arc idea, as commonly employed, is defective in that it assumes sensory stimulus and motor response as distinct psychical existences, while in reality they are always inside a coordination and have their significance purely from the part played in maintaining or reconstituting the coordination" (Dewey, 1896, p. 360).

Dewey objects to treating the subject as an input/output system, a common assumption of behaviorism, classical cognitivism and latter-day Bayesian accounts, and in doing so, he points to one source of the problem. If we adopt a view of the person as an input/output system, with knowledge of the world coming in through the senses, and action on the world as the other end of the chain, we are *already* committed to the obligatory separation of the experiential domain of the subject from the common world we inhabit. Indeed, once we adopt a realist stance that treats the world as having an intrinsic being independent of any observer, we *necessarily* arrive at some form of meditational epistemology. This is of no concern for many accounts we may wish to develop, but for the proper treatment of perception and action we need some alternative. The answer, it seems, is not some kind of anti-realism, but to learn to think relationally, and to see how both perception and

UCD School of Computer Science and Informatics, University College Dublin, Dublin, Republic of Ireland.

Corresponding author:
Fred Cummins, UCD School of Computer Science and Informatics, University College Dublin, Dublin, Republic of Ireland.
Email: fred.cummins@ucd.ie

action arise in the relation between an organism and its world.

Among the many theoretical developments that are of help in developing an enactive epistemology is the account of active perception developed primarily in the work of Kevin O'Regan and Alva Noë, most notably in their well-known 2001 paper (O'Regan & Noë, 2001). Although their subsequent elaboration of the theory first presented in this foundational paper has diverged somewhat, their theoretical account of the skill of vision as the mastery of sensorimotor contingencies, first introduced in 2001, serves as a useful landmark. Perception here is re-conceived as an exploratory activity, and seeing (for this is an account of visual perception) is understood as the mastery of the sensorimotor contingencies that inhere in the activity. Such sensorimotor contingencies arise in the relation between the organism and its surround. A strong source of convergent support for this approach to perception more generally comes from work in the field of sensory substitution (Rita, 1972), where perception with the aid of a novel organism–world interface is possible only after learning how movement and the attendant sensory flux co-vary in a lawful way. These ideas are also found in J. J. Gibson's exploration of the lawful covariance of movement and optic flow (Gibson, 1966).

Yet the language we inherit trips us up at every opportunity. The work referenced above is cast as pertaining to perception, as if perception were a distinguishable activity, or a separable facet of experience, that is distinct from acting in and on the world. The term "sense-making" has found some currency as a way of getting away from the inherited set of associations that come with the term "perception" (Di Paolo, 2005; Froese & Di Paolo, 2011). This serves to emphasize the active engagement through which a comprehensible and manageable world arises for a subject, and helps to prevent unwanted associations with such notions as sensations and representations.

This shift of focus could go further. The role of the sensory modalities and exploratory action in making the world both comprehensible and manageable can be extended to provide a corrective to the other end of the discredited reflex arc: action. In this paper, I will propose a somewhat novel framework within which the skilled activity of speaking may be viewed. As with the shift from a representational account of perception to an enactive account of sense-making, I shall try to reframe the discussion of behavior, moving from the notion of "motor control" to a radically different, coordinative, view of what skilled action is. Just as the two ends of the reflex arc are not really separate, so my account of skilled action will be, in fact, an account of sense-making. And just as the enactive account of perception based upon sensorimotor contingencies has clear pre-cursors in the work of Gibson and the ecological approach to perception, so my account of action is built upon the theory of coordination dynamics most closely associated with Scott Kelso (Kelso, 1995), although space constraints will prohibit dense referencing.

2 Motivating the sensorimotor coordination

Here is Dewey again, insisting on framing "perception" within an active framework, which he dubs, in a surprisingly contemporary tone, a sensorimotor coordination:

> "Upon analysis, we find that we begin not with a sensory stimulus, but with a sensori-motor coördination, the optical-ocular, and that in a certain sense it is the movement which is primary, and the sensation which is secondary, the movement of body, head and eye muscles determining the quality of what is experienced. In other words, the real beginning is with the act of seeing; it is looking, and not a sensation of light. The sensory quale gives the value of the act, just as the movement furnishes its mechanism and control, but both sensation and movement lie inside, not outside the act." (Dewey, 1896, p. 358–359).

The sensorimotor coordination is the framework within which movement and sensory change co-occur. This general account applies whether we describe the act as perceptual exploration of the world, as in enactive accounts of perception, or as the exhibition of a skilled action, as in speaking. The former reveals how the world becomes comprehensible, the latter, how it becomes manageable. Their union, we might describe as sense-making.

In skilled action, there is a necessary and lawful co-variation of movement and sensory change. Let us take a simple example. Standing still is, perhaps, the simplest skilled "action" we can identify. The boundary conditions of this skill require that the relation between the torso and the surrounding environment be relatively invariant. If the person should lean too far forward or backwards, a correction is required. Considering only the visual modality, we can see that any such deviation from the desired position necessarily goes hand in hand with a characteristic change in the distribution of patterned light falling on the retina, as shown in Figure 1. Lee and colleagues demonstrated this necessary relation clearly in the swaying room paradigm (Lee & Aronson, 1974), in which a room is prepared without a floor. The room can be moved as a whole backwards and forwards. Small sinusoidal oscillation of the room generates a pattern on the retina that otherwise would normally come from swaying of the torso. The result is an automatic correction by the subject, so that he sways with the same frequency and approximately in phase with the room. A sudden jolt of the room can knock a small child over.

In describing skilled action, we thus need to examine the lawful co-variation of movement and sensory

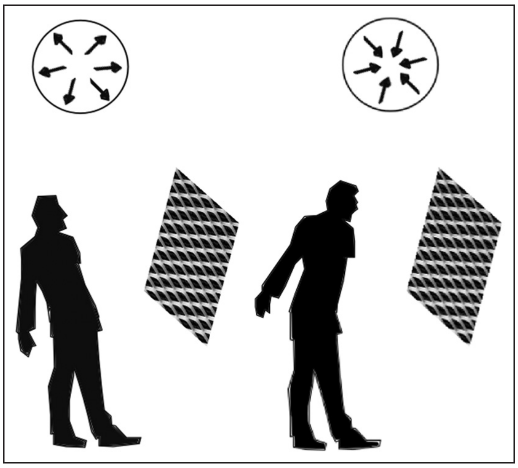

Figure 1. As a subject leans forwards or backwards, there is necessarily a concomitant expansion or contraction of the textured pattern of illumination on the retina (top).

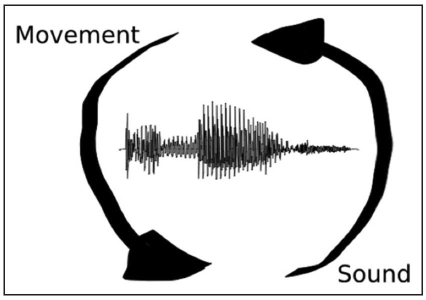

Figure 2. Speaking represented as a sensorimotor coordination, with mutually supporting motor and sensory arcs.

change, and also the boundary conditions that serve to define and delimit the act, making it a felicitous exercise of a skill, and not a random movement.

3 The act of speaking

We turn now to the skill that is speaking. The origin of the account of speech to be developed here lies in a very simple phenomenological observation. Its very obviousness may underlie the failure of theoretical accounts of speech to acknowledge it at all. Being trained as a phonetician, I failed to notice it for very many years, as the conventional and institutionalized language of speech "production", by a "speech production system" became second nature. One is taught at an early stage that speaking is a process by which some non-speech thought is encoded, first into movements, then into an acoustic signal, to be received by a listener for decoding. This entirely conventional view treats speech as a product, and the movements of the articulators as the production plant.

What this picture hides is striking: speech sound does not happen after the movements of speech articulation. Movement and sound co-occur. Always. In every single utterance I or anybody else has ever made, speech sound and the movements of speech co-occur. They are not serially ordered as first movement, then sound. Hopefully this is obvious. I rather fear it is so obvious as to invite summary dismissal for lack of novelty. But I wish to suggest that it opens up a novel perspective on speech, and then by extension, on a whole range of allied phenomena.

Let us first use this observation to characterize speech as illustrated in Figure 2. On this view, speech may be described as *the constrained co-occurrence of movement and sound*. This is illustrated in the figure as a single coordination (á la Dewey), which is shown as mutually supporting sensory and motor arcs. The notion of a superordinate coordination entails that the movement and sound are at no time independent of one another. The constraints, or boundary conditions, are not shown, and indeed, they will, in general, be complex. Indeed, collectively, they are the definition of a spoken language.[1]

With this starting point, we can already look anew at some familiar phenomena that arise in speaking. We first consider speaking under conditions of delayed auditory feedback (DAF). If a feedback delay of about one-fifth of a second is artificially imposed on a speaker, the constrained co-occurrence of speech and movement is shattered, and speaking rapidly becomes difficult, if not impossible (Yates, 1963). This can be accounted for under conventional accounts by interposing some complex cognitive architecture that constantly monitors speech output and that feeds back into a corrective process with an appropriate time delay (Howell, 2002). Under the present construal of speech, however, a more direct account suggests itself. To eliminate the sensory arc of the sensorimotor coordination that is the act of speaking, is to remove the very constitutive conditions of speaking itself.

Interestingly, experienced simultaneous translators can learn the skill of listening to one sound stream in one language, while speaking in another language. Such skilled speakers have also been found to be relatively unaffected by DAF (Fabbro & Darò, 1995). This observation suggests that we might more carefully characterize speech as the constrained co-occurrence of movement and sensory flux, where that sensory flux includes both sound and proprioception (and kinesthetic awareness). In the exceptional case of simultaneous translators, the sensory arc of the coordination that is speaking is then provided by proprioception/kinaesthesia alone.

Mediational epistemologies emphasize the separation of the sensory modalities, and consider knowledge as constructed from signals arising in distinct input channels. The dynamic approach to be outlined here takes the act as a whole as a unit that is defined and delimited by constrained motor and sensory covariance. The dissection of the act into separate forms of covariance distributed across the modalities is not prior to the act, but is an intellectual elaboration of the sensory psychologist, taking apart that which is originally whole (Skarda, 1999).

4 Synchronous speaking

We now extend this description of speaking to the case of joint, or synchronous speech, as illustrated in Figure 3. We first circumscribe the phenomenon to be studied: joint speaking, where multiple people say the same thing at the same time, occurs in many situations such as classrooms, temples, courtrooms and football stadia. An experimental analogue of this, called synchronous speaking, has been introduced by the present author (Cummins, 2003). In this laboratory task, subjects are presented with short, novel texts. Following an initial silent reading, they are asked to read in synchrony with one another after a go signal from the experimenter. Subjects typically have little or no difficulty complying with this task specification, and the synchrony exhibited is remarkably tight, with a mean asynchrony of approximately 40 ms throughout the phrase, rising to about 60 ms at phrase onset after a pause. This degree of tight coupling across speakers is achieved without substantial practice, and does not improve greatly with practice. It is particularly striking in light of the capacity of the voice for expressive variation, as speech is very highly malleable as a function of context and purpose. The domain of synchronous speech provides an interesting challenge within which we can contrast meditational, representational accounts and novel dynamical ones (Cummins, 2009, 2011).

For each of the speakers in the dyadic coordination, speaking is, as before, the constrained covariation of movement and sensory flux, but in this case, the sensory arc of the coordination is a summation of an endogenous component and an exogenous one. The former comprises both sound and proprioception from the speaker herself, while the latter is provided by the sound produced by the second speaker. The sensorimotor coordination is thus not entirely individualistic, and speaking in this task is not entirely attributable to one individual or to the other, but to the dyad. As each speaker is embroiled in a coordinative act that includes the speech of the other, the two speakers are literally coupled, forming a superordinate dyadic system, with no individual locus of control.

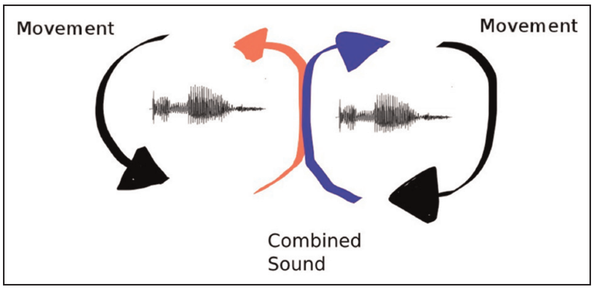

Figure 3. Synchronous speaking represented as a dyadic sensorimotor coordination, with a mutual sensory arc.

In order to further motivate this perspective on synchronous speaking, we briefly contrast this dynamical/coordinative description of the act with a description couched in individualistic, information processing terms (Figure 4), framed by the notion of control, rather than coordination. To describe the situation in which two speakers speak in synchrony, we might observe that each is both speaking, and monitoring the speech of the other speaker. Within a control-based paradigm, speaking is understood as originating in a central executive who issues a motor command, The motor command is copied to a predictive forward model, so that the sensory consequences that actually occur can be compared with those that are predicted to occur. This representationally voracious scenario rapidly becomes intractable when the prediction needs to include another speaker, who is in turn predicting oneself. Nothing here is logically impossible, and it is not necessary, or even plausible, that an internal predictive model should be very high fidelity. There are representational accounts available that rely less heavily on the complex machinery I here evoke, for example Bayesian accounts which seek to minimize the complexity of the prediction, or perceptual control theory (PCT), which would regard each speaker as a negative feedback system based on a hypothetical internal variable (Bourbon, 1995; Köording & Wolpert, 2004). But any such representational characterization of the act of speaking in synchrony must surely appear as massively less parsimonious than the suggested dynamical and coordinative description (Figure 3), while also failing to capture qualitatively the manifest coupling among speakers.

It might reasonably be objected that many facets of the manifest behavior that are best described within representational approaches are not available within the proposed enactive view. This is probably correct. There is value to regarding each speaker as a closed domain, interacting with a separate world, of which the co-speaker is a part. Enactive accounts are still in their infancy, and it is not clear that they will ever provide a substitute for extant accounts of, for example, canonical linguistic structure. The account suggested here is intended to be complementary to such accounts, not to replace them, with the motivation that only by eschewing the subject–object dichotomy does the reality of the dyadic system become approachable (Dale, Dietrich, & Chemero, 2009). It is only in this fashion that the

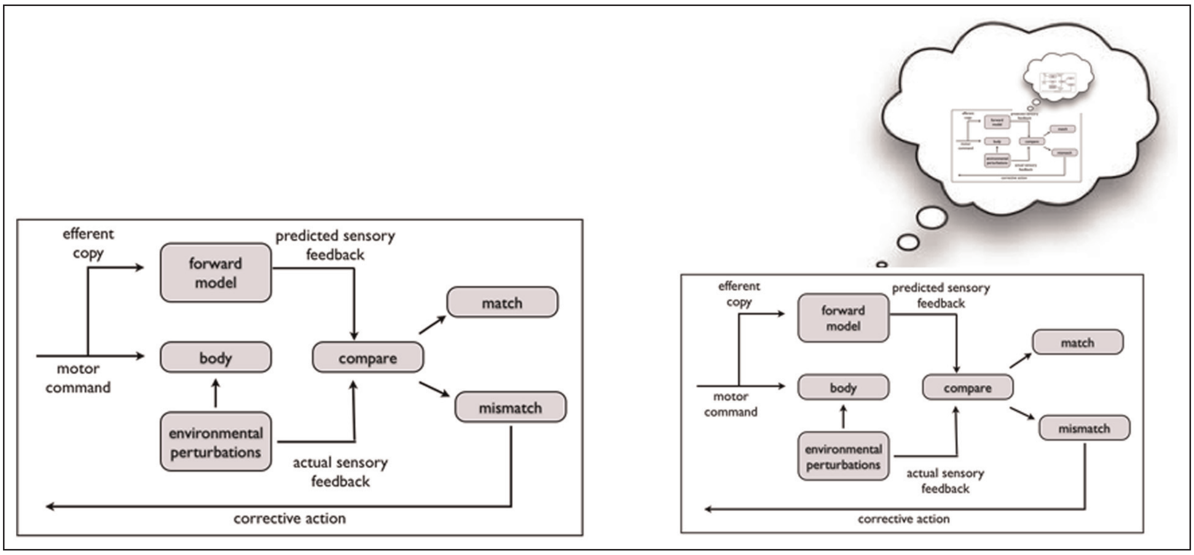

Figure 4. Predictive description of synchronous speaking. Left: speaking is just another skilled action, in which the sensory consequences of present movement are monitored and compared with predictions. Right: Prediction of prediction of prediction of ….

characteristics of the joint behavior that are properly located at the level of the dyad, rather than the individual, can be adequately acknowledged.

The parsimony of the dynamical account is only one factor to take into consideration when contrasting the approaches. Another is its predictive value and ability to account for details of the phenomenon under study.

The dynamical account regards the two speakers as coupled, and the basis for the coupling is the shared sensory arc of the sensorimotor coordination that is the act of speaking. This creates a superordinate domain of organization at the level of the dyad. When two systems, each with their own intrinsic self-sustaining dynamical behavior are coupled, the resulting coupled system has system level properties that are not derivable from the mere conjunction of the component systems (Kelso, 1995; Pikovsky, Rosenblum, & Kurths, 2001). In particular, the conditions under which coupling can arise constitute boundary conditions for the dyadic system. Once those conditions are not present, there will be a qualitative change from a single superordinate system to two decoupled component systems. This can be illustrated graphically by a familiar analogy. Consider runners within a three-legged race. Two component systems (the runners) are physically coupled by having their medial legs tied together. Running is still possible, but only under constraints that arise at the dyadic level. A misstep, under such constrained circumstances, can rapidly bring the coupling to an abrupt end as the components fall down. In an entirely analogous fashion, a speech error that arises in the course of synchronous speaking often has the result that both speakers abruptly and simultaneously stop speaking. Other outcomes are possible: if the coupling is no longer given, one speaker may plough on with no regard for the co-speaker. But abrupt and simultaneous cessation is a frequent occurrence. This constitutes an error type that is specific to the condition of speaking in synchrony, and that provides highly specific information about the kind of phenomenon under observation. If we view the speakers as coupled through the felicitous co-variation of movement and sound, a speech error by one speaker has the necessary consequence for both speakers that the constrained co-variation of movement and sound is no longer given. This may thus bring about the dissolution of the dyadic level of organization. In contrast, a representational account that views the two speakers as essentially separate systems, will require a great deal of *ad hoc* machinery to account for the occurrence of the simultaneous cessation of speech. Both accounts are incomplete, but only the dynamical account accords a reality to the coupling between speakers.

5 Escaping the obligatory view of the person as a linear throughput system

Dewey objected to the poverty of the view of the person as driven by the reflex arc in the following terms:

> "[Proponents of the arc fail to see] that the arc of which it talks is virtually a circuit, a continual reconstitution, it breaks continuity and leaves us nothing but a series of jerks, the origin of each jerk to be sought outside the process of experience itself, in either an external pressure of 'environment,' or else in an unaccountable spontaneous

variation from within the 'soul' or the 'organism.'" (Dewey, 1896, p. 360)

and later

"What we have is a circuit, not an arc or broken segment of a circle." (Dewey, 1896, p. 363).

But his objections have not, so far, been given due consideration. No wonder, for the distinction between Dewey's relational conception of organism and world, and the conventional perception → cognition → action series is not a small one. The two accounts are built upon two entirely different, and incompatible, metaphysical foundations. The latter, linear throughput model is familiar, to the point of invisibility. Despite the claims of cognitive science practitioners to distance themselves from Cartesian dualism, it is resolutely Cartesian in structure and spirit. It distinguishes between the mental and the physical, which has the significant advantage that it allows us to coordinate our joint activities with respect to an "external" world that has properties considered to be independent of any experiencer. It supports efficient causal explanation in which the behavior of an organism is the final product of a long process that begins with the recovery or construction of a representation of a pre-existing world. But it comes at the cost of having an unbridgeable metaphysical divide between experience and the world. This is an appropriate metaphysics for building houses and ordering pizza, but it has severe limitations when tasked with describing or explaining the relationship between the knower and the known, or the dancer and the dance.

Dewey, in contrast, anticipates the perspective shift that comes with recognizing that the world and the subject co-arise in the recurrent interactions between an organism or system and its environment. This is the approach introduced perhaps most visibly in The Embodied Mind (Varela, Thompson, & Rosch, 1992), but also foreshadowed in the early work of Merleau-Ponty, who argued in the spirit of Dewey (compare this with the above-mentioned quote from (Dewey, 1896, p. 358–359)):

"The organism cannot properly be compared to a keyboard on which the external stimuli would play and in which their proper form would be delineated for the simple reason that the organism contributes to the constitution of that form … 'The properties of the object and the intentions of the subject … are not only intermingled; they also constitute a new whole.' When the eye and the ear follow an animal in flight, it is impossible to say 'which started first' in the exchange of stimuli and responses. Since all the movements of the organism are always conditioned by external influences, one can, if one wishes, readily treat behavior as an effect of the milieu. But in the same way, since all the stimulations which the organism receives have in turn been possible only by its preceding movements which have culminated in exposing the receptor organ to the external influences, once could also say that the behavior is the first cause of all the stimulations." (Merleau-Ponty, 1963, p. 13).

The reciprocity of perception and action is obscured in a perception-then-cognition-then-action framework. How shall we then talk of the activity of a being that closes the cycle, and does not hide the subject between input and output?

The enactive concept of sense-making offers a chance to develop a vocabulary that finds widespread acceptance, and that allows many details traditionally covered under either "perception" or "action" to be revisited. Indeed, as di Paolo has pointed out (Di Paolo, 2005), the term points both to the sense made, and to the activity that is necessary for this to happen. A similar re-alignment of vocabulary underlies Gibson's 1966 book title *The Senses Considered as Perceptual Systems*, which strove to argue that the meaning that sense-making activity gives rise to can come about precisely because organism–environment relations are systematic and subject to natural law.[2]

6 Perception, action, and sense-making

The term "sense-making" has been employed in several ways within the enactive literature, broadly construed. In the context of "perception", sense-making has usefully steered the discussion away from the passive uptake of features of objects and events in the "outside" world and towards the process of active exploration and inquiry that characterizes the directed interaction between agent and world, giving rise to an interpretation of the world for a subject. I take this to be the way in which "sense-making" is interpreted when used with respect to the theory of sensorimotor correspondences most closely associated with O'Regan and Noë, although the term does not appear in their 2001 paper (O'Regan & Noë, 2001).

In the foundational literature on the concept of autopoiesis, sense-making plays a more fundamental role. Sense-making has been presented as constitutive of cognition. There is some divergence of views here. Thompson (2004) makes the identification that "Living entails sense-making, which equals cognition". Di Paolo chooses to tie the concept of sense-making to the notion of adaptivity, or active homeostasis, whereby an organism will preferentially seek interactions with the environment that lead it away from potential danger to its systematic integrity (Di Paolo, 2005), thereby allowing the emergence of graded norms. Irrespective of the difference in emphasis here, in each of these contexts sense-making is more than "figuring stuff out". It is the process by which a world of significance to the subject arises, encounters are meaningful and experience becomes intentional (Varela et al., 1992).

Bridging the gap between the received psychological ontology (including the concepts of perception and action) and the emerging vocabulary of the enactive approach is not straightforward. It is not possible to simply map from the notions of autonomy, sense-making and autopoiesis to the central concepts within the ruthlessly individualistic models of contemporary cognitive psychology. If one starts with the foundational sense of sense-making and develops the notion towards a substitute for "perception", one must ask what is this "mastery of sensorimotor contingency" (SMC) that is required to allow a world of teapots (objects) and car crashes (events) to emerge? The answer is not, and cannot be, confined to the biological individual. The mastery of SMC entails norms of behavior and conventions of interpretation under which objects and events emerge laden with significance: there are no teapots that are not simultaneously property, imbued with functional significance, monetary and aesthetic value, etc. There are no car crashes that are not fraught with consequences. It is not possible to peel away a purely perceptual veneer to the encounter with a teapot or a crash. Consider even the difference in the form of tactile exploration you might engage in with a Ming vase on the one hand and a toy plastic teapot on the other. Here, the theory of SMC needs further development in order to improve upon the toothless notion of perception without inherent significance for the isolated Cartesian subject.

In the domain of action, or skilled behavior, we have a similar debt, although the steps leading from an account of sense-making to the conventional descriptive ontology of behavior are somewhat different. The constraints that bound behavior, distinguishing it from mere movement, likewise cannot be said to lie entirely within the individual. In learning to speak, one is learning to constrain the co-occurrence of sound and movement such that the sounds produced function as speech in the speaker's community. These constraints are best expressed in patterns of activity distributed widely over communities, and not lodged within individual speakers. A successful syllable is one that functions as a syllable in use.

7 Joint speaking and participatory sense-making

When we turn from the speech activity of an individual and now consider joint speaking, where multiple speakers exhibit a great deal of very tight coordination, we stumble upon a domain within which very many of the principal themes of the emerging enactive account come into focus. At a relatively mechanistic level, we find the constrained linkage of movement and sensory flux, but within a collective domain, demonstrating that the stability of form that is so characteristic of skilled action is not restricted to the activity of a single individual. This stability, demonstrated in that the same functional activity can be repeated again and again, is the principal evidence for the mastery of sensorimotor contingencies in skilled engagement with the world. It is not confined to the individual. The coupling that is evidenced by the collective cessation of speaking on the occasion of a speech error demonstrates that the sensorimotor relation is being regulated at the level of the dyad, rather than the individual. This may provide an operationalized example of participatory sense making, in the sense of De Jaegher and Di Paolo (2007) who defined it thus:

> "Social interaction is the regulated coupling between at least two autonomous agents, where the regulation is aimed at aspects of the coupling itself so that it constitutes an emergent autonomous organization in the domain of relational dynamics, without destroying in the process the autonomy of the agents involved (though the latter's scope can be augmented or reduced)." (De Jaegher & Di Paolo, 2007, p. 493).

In the empirical study of joint speaking, e.g. using the laboratory tool of synchronous speech, we have a unique opportunity to observe, influence and manipulate the coupling that obtains between speakers. For example, in a recent study, we varied the relative loudness of the speaker's own voice in comparison with that of the co-speaker (Cummins, Li, &Wang, 2013). In both English and Chinese, we found that an increase in the relative loudness of the co-speaker resulted in a greater degree of synchrony among speakers, suggesting that this manipulation allowed direct regulation of the coupling strength between them. We also induced speech errors by having occasional mis-matches in the texts being read by pairs of speakers. This produced somewhat different effects in the two languages. In English, a stronger coupling made the dyad more susceptible to speech errors. In Chinese, no such susceptibility was observed. We interpret this difference in the languages in conjunction with another observation: in Chinese, the prosody of synchronous speech was somewhat altered, such that the individual syllables were more prominent and regular than when spoken by a single individual. In English, no such prosodic alteration has been observed. This suggests that the greater prominence of the syllable in synchronous Chinese provides a degree of coordinative stability that is unavailable in English.

The syllable is often regarded as a coordination among consonantal and vocalic gestures (Browman & Goldstein, 2000). We note that the syllable functions somewhat differently in the two languages: in English, syllables are highly variable both in the consonantal sequences allowed, and in the strength of articulation of the individual vowels. As a result, syllabification of continuous English speech is often ambiguous or even impossible. It is thus not clear that English speakers have the option of enhancing syllabic coordination. In Chinese, syllables are simpler in their segmental make up, and it is normally unproblematic to identify individual syllables, even in continuous speech. By enhancing the prominence of the syllable, a degree of stability appears to be available to the speakers that supports synchronization, and simultaneously provides a degree of resistance to external perturbation, such as that occasioned by inducing speech errors. There is thus an interplay between the coordinative structures that characterize each language, and their manifestation in speaking jointly. Further work is being done to extend these observations.

So much for the making of sense; what about the sense that is thereby made? Joint speaking occurs in many and diverse social contexts. Many of these are contexts accorded a great degree of collective significance: group prayer and recitals of oaths and pledges are two clear examples. We might add to them the spontaneous expression of group purpose in the synchronized chanting of demonstrators during protests. Clearly the purpose and significance of joint speaking is interestingly different from that of speech produced by one person at a time. As De Jaegher and Di Paolo point out, the significance of collective activity is by no means always positively valenced. Thus, we find the exploitation of the transcendent significance of joint speaking exploited in the service of propaganda, as with chanting and synchronized activity at mass political rallies on all sides of the political spectrum.

Accounts of social interaction and empathy that are couched in the individualistic language of most current contemporary cognitive neuroscience inevitably lean on the notion of one system trying to figure out the other. This is true of both simulation theory and Theory of Mind theory, and of most attempts to include a "mirror system" into accounts of collective behavior. Although there is great variety in such accounts, they all share the *Being John Malkovich* problem of having a subject look out at a profoundly estranged and separate world. In acknowledging the possibility of the emergence of group intentions through highly coordinated collective activity, we see that we have available to us now a radically different form of subjectivity. Joint speaking provides an empirical domain in which we can compare and contrast the individual and joint activities at both a mechanistic level, and at the level of significance for the participants. It provides us with a foothold in developing an enactive theory of action.

Acknowledgements

Thanks are due to Marek McGann for numerous conversations around these topics, and to both editors and reviewers for their comments which helped to improve the manuscript.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Notes

1. Consideration of how the constraints that characterize skilled action such as speech are acquired would take us far beyond the present topic, but the present account must be acknowledged to be incomplete without some such story.
2. For the sake of completeness, we might note that von Uexkïll's notion of the Umwelt similarly collapses both the world of meaning and sense (Merkwelt) and the world of effective action (Wirkwelt) (Von Uexküll, 1992).

References

Bourbon, W. T. (1995). Perceptual control theory. In *Comparative approaches to cognitive science* (151–172). MIT Press.

Browman, C., & Goldstein, L. (2000). Competing constraints on intergestural coordination and self-organization of phonological structures. *Les Cahiers de l'ICP. Bulletin de la communication parlée*, *5*, 25–34.

Cummins, F. (2003). Practice and performance in speech produced synchronously. *Journal of Phonetics*, *31*(2), 139–148.

Cummins, F. (2009). Rhythm as entrainment: The case of synchronous speech. *Journal of Phonetics*, *37*(1), 16–28.

Cummins, F. (2011). Periodic and aperiodic synchronization in skilled action. *Frontiers in Human Neuroscience*, 5(170).

Cummins, F., Li, C., & Wang, B. (2013). Coupling among speakers during synchronous speaking in English and Mandarin. *Journal of Phonetics*. (Submitted)

Dale, R., Dietrich, E., & Chemero, A. (2009). Explanatory pluralism in cognitive science. *Cognitive Science*, *33*(5), 739–742.

De Jaegher, H., & Di Paolo, E. (2007). Participatory sense-making. *Phenomenology and the Cognitive Sciences*, *6*(4), 485–507.

Dewey, J. (1896). The reflex arc concept in psychology. *Psychological Review*, *3*(4), 357.

Di Paolo, E. (2005). Autopoiesis, adaptivity, teleology, agency. *Phenomenology and the Cognitive Sciences*, *4*(4), 429–452.

Fabbro, F., & Darò, V. (1995). Delayed auditory feedback in polyglot simultaneous interpreters. *Brain and language*, *48*(3), 309–319.

Froese, T., & Di Paolo, E. A. (2011). The enactive approach: theoretical sketches from cell to society. *Pragmatics & Cognition*, *19*(1), 1–36.

Gibson, J. (1966). *The senses considered as perceptual systems*. Houghton Mifflin.

Howell, P. (2002). The EXPLAN theory of fluency control applied to the treatment of stuttering. *Amsterdam Studies in the Theory and History of Linguistic Science Series 4*, 95–118.

Kelso, J. A. S. (1995). *Dynamic patterns*. Cambridge, MA: MIT Press.

Körding, K. P., & Wolpert, D. M. (2004). Bayesian integration in sensorimotor learning. *Nature*, *427*(6971), 244–247.

Lee, D., & Aronson, E. (1974). Visual proprioceptive control of standing in human infants. Attention, *Perception, & Psychophysics*, *15*(3), 529–532.

Merleau-Ponty, M. (1963). *The structure of behavior (translation alden fisher)*. Boston: Beacon Press. (Published in French, 1942)
O'Regan, J., & Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences*, *24*(5), 939–972.
Pikovsky, A., Rosenblum, M., & Kurths, J. (2001). *Synchronization: A universal concept in nonlinear sciences*. CUP.
Rita, P. Bach-y. (1972). *Brain mechanisms in sensory substitution*. Academic Press New York.
Skarda, C. (1999). The perceptual form of life. *Journal of Consciousness Studies*, *6*(11-12), 11–12.
Thompson, E. (2004). Life and mind: From autopoiesis to neurophenomenology. a tribute to francisco varela. *Phenomenology and the Cognitive Sciences*, *3*(4), 381–398.
Varela, F., Thompson, E., & Rosch, E. (1992). *The embodied mind: Cognitive science and human experience*. MIT press.
Von Uexküll, J. (1992). A stroll through the worlds of animals and men: A picture book of invisible worlds. *Semiotica*, *89*(4), 319–391.
Wittgenstein, L. (1973). *Philosophical investigations: The english text of the third edition*. Macmillan (New York).
Yates, A. (1963). Delayed auditory feedback. *Psychological Bulletin; Psychological Bulletin*, *60*(3), 213.

Author biography

Fred Cummins is co-director of the Cognitive Science Programme at University College Dublin. He obtained a PhD with joint major in Linguistics and Cognitive Science at Indiana University in 1997. Following postdoctoral positions at Northwestern University (Evanston, IL) and IDSIA (Lugano, CH), he returned to Ireland. His research has focused on dynamical and embodied approaches to coordination in speech. His current area of interest is the domain of joint speaking, both in the wild and in the laboratory.