

# Philosophical Psychology

## Lecture 03: On the Motor Theory of Speech Perception

Stephen A. Butterfill  
<s.butterfill@warwick.ac.uk>

October 20, 2016

title-slide

### A Question about Experiences of (Speech) Actions

What do we experience when we encounter others' actions? One hypothesis (the Indirect Hypothesis) says that such experiences are all experiences of bodily configurations, of joint displacements and of effects characteristic of particular actions. Another hypothesis (the Direct Hypothesis) says that in observing an action we sometimes experience not only bodily configurations and joint displacements and their sensory effects but also the action as directed to a particular outcome.

---

slide-6 I suppose that speech actions are actions, so I will start with those.

categorical\_perception\_speech

### Speech Perception

\*\*\*

slide-8 In speaking we produce an overlapping sequence of articulatory gestures, which are motor actions involving coordinated movements of the lips, tongue, velum and larynx. These gestures are the units in terms of which we plan utterances (Browman and Goldstein 1992; Goldstein, Fowler, et al. 2003).

These are the actions I want to focus on first in thinking about what we experience when we encounter others' actions.

slide-9 ;t 'Trajectory of lower lip in [abe] as measured by tracking infraredLED placed on subject's lower lip'

‘Not every utterance of word transcribed with /b/ will display exactly the trajectory of Fig. 1.: the trajectory will vary with vowel context, syllable position, stress, speaking rate and speaker. We must, therefore, ultimately characterise /b/ as a family of patterns of lip movement’ (Browman and Goldstein 1986, p. 224)

**slide-11** A schematic spectrogram for a synthetic sound which is normally perceived as [ra]. The horizontal axis represents time, the vertical frequency. A schematic spectrogram for [la].

**slide-12** In the middle you see the *base*, i.e. the part of the spectrogram common to [ra] and [la]. This is played to one ear

Below you see the transitions, , i.e. the parts of the spectrogram that differ between [ra] and [la]. When played in isolation these sound like a chirp. When played at the same time as the base but in the other ear, subjects hear a chirp and a [ra] or a [la] depending on which transition is played.

How do we know that the same stimuli may be processed by different perceptual systems concurrently—for instance, how do we know that speech and auditory processing are distinct? A phenomenon called “duplex” perception demonstrates their distinctness occurs in. Artificial speech-like stimuli for two syllables, [ra] and [la], are generated. The acoustic signals for each syllable is artificially broken up into two parts, the “base” and “transition” (see Fig. \*\*\* below). The syllables have the same “base” but differ in the “transition”. When the “transition” is played alone it sounds like a chirp and quite unlike anything we normally hear in speech. Duplex perception occurs when the base and transition are played together but in separate ears. In this case, subjects hear both the chirp that they hear when the transition is played in isolation, and the syllable [la] or [ra]. Which syllable they hear depends on which transition is played, so speech processing must have combined the base and transition. By contrast, auditory processing must have failed to combine them because otherwise the chirp would not have been heard. In this case, then, the perception resulting from the duplex presentation involves simultaneous auditory and speech recognition processes. This shows that auditory and speech processing are distinct perceptual processes.

The duplex case is unusual. We can’t normally hear the chirps we make in speaking because speech processing inhibits this level of auditory processing. But plainly speech is subject to some auditory processing for we can hear extra-linguistic qualities of speech; some of these provide cues to emotional state, gender and class. Perception of these extra-linguistic qualities enables us to distinguish stimuli within a category. As already mentioned, this is a problem for Repp’s operational definition. Our ability to discriminate stimuli is the product of both categorical speech processing and non-

categorical auditory processing. If we want to get at the essence of categorical perception it seems there is no alternative but to appeal to particular perceptual processes rather than behaviours.

Source: (Liberman et al. 1981)

**slide-13** Here are 12 speech-like sounds. Acoustically each differs from its neighbours no more than any other does.

**slide-14** They would be labelled differently

**slide-15** And within a label they are relatively hard to discriminate whereas ...

**slide-16** Discriminating acoustically no less similar stimuli that are given different labels is easier (faster and more accurate).

This is categorical perception: speed and accuracy maps onto labelling ...

**slide-17** Categorical perception of mating calls and perhaps other acoustic signals is widespread in non-human animals including monkeys, mice and chinchillas (Ehret 1987; Kuhl 1987), and is even found in cognitively unsophisticated animals such as frogs (Baugh, Akre and Ryan 2008) and crickets (Wytenbach, May and Hoy 1996).

**slide-19** the location of the category boundaries changes depending on contextual factors such as the speaker's dialect,<sup>22</sup> or the rate at which the speaker talks;<sup>23</sup> both factors dramatically affect which sounds are produced.

**slide-20** This means that in two different contexts, different stimuli may result in the same perceptions, and the same stimulus may result in different perceptions.

**slide-23** co-articulation, the fact that phonic gestures overlap (this is what makes talking fast).

**slide-24** 1. Speech perception is categorical

2. The category boundaries correspond (imperfectly but robustly) to differences in articulatory gestures

3. The best explanation of (2) involves the hypothesis that the objects of speech perception are articulatory gestures

*Articulatory Gesture:* In speaking we produce an overlapping sequence of articulatory gestures, which are motor actions involving coordinated movements of the lips, tongue, velum and larynx. These gestures are the units in terms of which we plan utterances (Browman and Goldstein 1992; Goldstein, Fowler, et al. 2003).

**slide-33** The difference in differences ...

Here is the argument. Consider two sequences of sensory encounters: (a) a sequence of sensory encounters with two phonetic events that do not differ with respect to category (both are realisations of /d/, say), and (b) a sequence of sensory encounters with two phonetic events that do so differ (one is a realisation of /d/ the other of /g/, say).<sup>11</sup> Let the events encountered in the first sequence differ from each other acoustically in the same way and by the same amount as the events encountered in the second sequence differ from each other. (That it is possible to find two such pairs of events follows from the fact that we enjoy categorical perception of speech.) The two sequences are depicted in Fig. 3. Now:

**slide-38** The fourth step in this argument, (4), needs some filling in. How would the thesis that categorical perception of speech is a form of perceptual experience explain the difference in differences in phenomenal character? If the thesis is true, the first sequence of sensory encounters, (a), involves two perceptual experiences as of a single phoneme whereas the second sequence of encounters, (b), involves perceptual experiences as of different phonemes.<sup>12</sup> Let us assume (not very controversially) that perceptual experiences have phenomenal characters and that which phenomenal character a perceptual experience has depends in part on what it is as of.<sup>13</sup> It follows that differences in what perceptual experiences are as of can explain differences in the phenomenal characters of those perceptual experiences. In particular, if it is a fact that (b) involves perceptual experiences as of different things whereas (a) does not, this could explain why the sensory encounters in (b) differ in phenomenal character in a way that the sensory encounters in (a) do not.

**slide-39** Recall the question ...

**slide-40** It looks like the Direct Hypothesis wins, or at least that we must reject the Indirect Hypothesis.

There's just one little problem ...

**slide-41** How could the objects of categorical perception of speech be articulatory gestures?

The puzzle here is simple. Categorical perception of speech happens rapidly, and goal-directed actions are complex. How can something so complex be computed so quickly?

**slide-42** 'Humans [can] understand speech delivered at a rate of 20 to 30 ... phonemes per second' (Devlin 2006)

Before facing this problem directly, I want to think about action more generally ...

`double_life_motor_representation`

## The Double Life of Motor Representation

Motor representations live a kind of double life. Although paradigmatically involved in performing actions, they also occur when merely observing others act and sometimes influence thoughts about the goals of observed actions. Further, these influences are content-respecting: what you think about an action sometimes depends in part on how that action is represented motorically in you.

**slide-44** Suppose you are reaching for, grasping, transporting and then placing a pen. Performing even relatively simple action sequences like this involves satisfying many constraints that cannot normally be satisfied by explicit practical reasoning, especially if performance is to be rapid and fluent. Rather, such performances require motor representations. These representations are paradigmatically involved in preparing, executing and monitoring actions.<sup>1</sup> But they also live a double life. Motor representations concerning a particular type of action are involved not only in performing an action of that type but also sometimes in observing one. That is, if you were to observe Ayesha reach for, grasp, transport and then place a pen, motor representations would occur in you much like those that would also occur in you if it were you—not Ayesha—who was doing this.

Converging evidence for this assertion comes from a variety of methods and measures ...

**slide-45** Single cell recordings in nonhuman primates show that, for each of several types of action, there are populations of neurons that discharge both when an action of this type is performed and when one is observed (di Pellegrino et al. 1992; Gallese et al. 1996; Fogassi et al. 2005).

**slide-46** This is the performance data.

Note that the neurons are firing before the distinctive part of the action has occurred: that is, the peak is between movement onset and the monkey first touching the object to be grasped.

---

<sup>1</sup> See Wolpert et al. (1995); Miall and Wolpert (1996); Jeannerod (1998); Zhang and Rosenbaum (2007). Note that motor representations sometimes occur in an agent who has prepared an action and is required (as it turns out) not to perform it: although she has prevented herself from acting, motor representations specifying the action persist, perhaps because they are necessary for monitoring whether prevention has succeeded (Bonini et al. 2014).

Now let's compare performance with observation.

**slide-47** '(A) Congruence between the visual and the motor response of a mirror neuron. Unit 169 has a stronger discharge during grasping to eat than during grasping to place, both when the action is executed and when it is observed. Conventions as in Fig. 1. (B) Population-averaged responses during motor and visual tasks (12).'

**slide-48** 'word listening produces a phoneme specific activation of speech motor centres' (Fadiga et al. 2002)

'Phonemes that require in production a strong activation of tongue muscles, automatically produce, when heard, an activation of the listener's motor centres controlling tongue muscles.' (Fadiga et al. 2002)

**slide-49** Good, but this stops short of showing that the motor activations actually facilitate speech recognition ...

How did they reach these conclusions?

**slide-51** Involves tongue

**slide-52** No tongue required

Given TMS to motor cortex to amplify activity. Prediction: MEP in tongue muscle stronger for 'rr' than 'ff'.

**slide-54** Behaviour: interference effects (ovalization)

**slide-60** Puzzle 1

What are those motor representations doing here?

But there is also another puzzle ...

**slide-62** 'We recorded proactive eye movements while participants observed an actor grasping small or large objects. The participants' right hand either freely rested on the table or held with a suitable grip a large or a small object, respectively. Proactivity of gaze behaviour significantly decreased when participants observed the actor reaching her target with a grip that was incompatible with respect to that used by them to hold the object in their own hand.'

Follow ups: tie hands; TMS (impair)

**slide-64** Ability to perform actions: the 180 degree swing is standard walk, whereas 225 and 270 degree swings are not standard but can be trained.

Visual discrimination task: 'The visual recognition experiment was based on a forced-choice paradigm. In each trial, two point-light stimuli consisting of a total of nine dots were presented successively, at two different positions on

the screen ... Participants had to respond whether both stimuli represented the same gait pattern. Four cycles of each gait pattern were presented, each cycle lasting for about 1.2 s. The start position within the gait cycle was randomized across trials.'

Then training while BLINDFOLDED.

Then visual recognition task again.

'One possible explanation of the observed motor- visual transfer is that the participants might have picked up the rhythm that characterizes the trained motor pattern, but not necessarily the details of the learned body movement. To rule out this possibility, we performed a control experiment in which the motor training was replaced by purely visual training.'

**slide-65** Visual performance correlates with motor performance after (but not before) training. (They also found no correlation with 225 degrees, which was not trained.)

This effect is perhaps surprising given that your judgement ultimately rests on purely visual information (this is the point of the lights) whereas nothing could be seen during the training.

What explains this difference in judgement before and after training? Training of this kind typically alters the way things are represented motorically (Calvo-Merino et al. 2006). For this reason, the increase in the probability of making accurate judgements about the goals of others' actions is plausibly a consequence of differences in motor representations in the observer.

**slide-66** task: hear sound, identify one of four pictures.

Twenty-eight left-hemisphere-damaged patients with or without limb and/or buccofacial apraxia and seven right-hemisphere-damaged patients with no apraxia were asked to match sounds evoking human-related actions or nonhuman action sounds with specific visual pictures.

'In the novel sound-picture matching test used in this study, each patient was asked to listen to a sound and then choose from among four pictures the one corresponding to the heard sound. The sounds used included limb-related action sounds (LRAS), buccofacial-related action sounds (BRAS) [ten sounds were transitive, i.e., object related (e.g., inflating a balloon) and ten were intransitive, i.e., non object related (e.g., coughing). ], and non-human action-related sounds (NHARS) [e.g. sea waves, breaking]'

**slide-69** Beautiful results!

Key: limb-related action sounds (LRAS), buccofacial-related action sounds (BRAS), and non-human action-related sounds (NHARS)

A+ apraxia; AB+ : buccofacial apraxia; AL+ limb apraxia; LBD: left brain damage; RBD: right brain damage

**slide-70** ‘Double TMS pulses were applied just prior to stimuli presentation to selectively prime the cortical activity specifically in the lip (LipM1) or tongue (TongueM1) area’ (D’Ausilio et al. 2009, p. 381)

‘We hypothesized that focal stimulation would facilitate the perception of the concordant phonemes ([d] and [t] with TMS to TongueM1), but that there would be inhibition of perception of the discordant items ([b] and [p] in this case). Behavioral effects were measured via reaction times (RTs) and error rates.’ (D’Ausilio et al. 2009, p. 382)

**slide-71** ‘Effect of TMS on RTs show a double dissociation between stimulation site (TongueM1 and LipM1) and discrimination performance between class of stimuli (dental and labial). The y axis represents the amount of RT change induced by the TMS stimulation. Bars depict SEM. Asterisks indicate significance ( $p < 0.05$ ) at the post-hoc (Newman-Keuls) comparison.’ (D’Ausilio et al. 2009)

**slide-72** Puzzle 1

What are those motor representations doing here?

Motor representations concerning the goals of observed actions sometimes facilitate the identification of goals.

Question

How?

Question: How is it that motor representations concerning the goals of observed actions sometimes facilitate identification of goals?

**slide-75** Recall two facts mentioned in discussing speech perception ...

**slide-76** This makes sense given that, as D’Ausilio et al. (2009) argue, motor processes facilitate the identification of articulatory gestures. After all, motor processes and auditory processes are likely distinct.

**slide-77** This also makes sense: given that there’s a link between observing an articulatory gesture and representing it motorically, it seems that the motor representation of the articulatory gesture could be what speech perception aims at.

(I’m ignoring a lot of complexity about how, given enough context, motor impairments can have very little impact on speech perception. Maybe not all speech perception depends on recovering articulatory gestures, and maybe there are multiple kinds of process that enable us to comprehend what someone is saying.)



slide-78 Also recall the question

How could the objects of categorical perception of speech be articulatory gestures?

The puzzle here is simple. Categorical perception of speech happens rapidly, and goal-directed actions are complex. How can something so complex be computed so quickly?

Put it another way: how is it that the articulatory gesture is identified?

‘Humans [can] understand speech delivered at a rate of 20 to 30 ... phonemes per second’ (Devlin 2006)

Nothing we’ve yet covered enables us to answer this question.

teleological\_stance

## The Teleological Stance

The Teleological Stance (Gergeley and Csibra , 1995) provides a computational theory of pure goal ascription. Pure goal ascription is the process of identifying goals to which another’s actions are directed independently of any knowledge, or beliefs about, the intentions or other mental states of an agent.

slide-81 To illustrate that there’s really a question here (How is it that motor representations concerning the goals of observed actions sometimes influence thoughts about them?), consider this about MT:

‘it is not clear how these [production-perception] links or MNs [mirror neurons] would solve the problems of mapping variable acoustics to linguistic representations, which first motivated MT [the motor theory of speech perception].’ (Holt and Lotto 2014, p. 205)

slide-82 To solve this problem, we need to start with an account of pure goal ascription.

An account of pure goal ascription is an account of how you could in principle infer facts about the goals to which actions are directed from facts about joint displacements, bodily configurations and their effects (e.g. sounds). Such an account is a computational theory of pure goal ascription.

Emphasize: in the case of speech, the problem is to get from acoustic and visual information to articulatory gestures.

Emphasize: in the case of actions, the problem is to get from bodily configurations and joint displacements to goals to which actions are directed.

**slide-86** Some ants harvest plant hair and fungus in order to build traps to capture large insects; once captured, many worker ants sting the large insects, transport them and carve them up (Dejean et al. 2005).

We can think of the ants' behaviour as goal-directed without also thinking of it as involving intention.

**slide-87** As this illustrates, some actions involving are purposive in the sense that

**slide-88** among all their actual and possible consequences,

**slide-89** there are outcomes to which they are directed

**slide-90** In such cases we can say that the actions are clearly purposive.

**slide-91** The standard answer to this question involves intention.

**slide-92** An intention (1) specifies an outcome,

**slide-93** (2) coordinates the one or several activities which comprise the action;

and (3) coordinate these activities in a way that would normally facilitate the outcome's occurrence.

What binds particular component actions together into larger purposive actions? It is the fact that these actions are all parts of plans involving a single intention. What singles out an actual or possible outcome as one to which the component actions are collectively directed? It is the fact that this outcome is represented by the intention.

So the intention is what binds component actions together into purposive actions and links the action taken as a whole to the outcomes to which they are directed.

**slide-94** It is important to see that the third item—representing the directedness—is necessary.

This is quite simple but very important, so let me slowly explain why goal ascription requires representing the directedness of an action to an outcome.

**slide-95** Imagine two people, Ayesha and Beatrice, who each intend to break an egg. Acting on her intention, Ayesha breaks her egg. But Beatrice accidentally drops her egg while carrying it to the kitchen. So Ayesha and Beatrice perform visually similar actions which result in the same type of outcome, the breaking of an egg; but Beatrice's action is not directed to the outcome of her action whereas Ayesha's is.

**slide-96** Goal ascription requires the ability to distinguish between Ayesha's action and Beatrice's action. This requires representing not only actions and outcomes but also the directedness of actions to outcomes.

This is why I say that goal ascription requires representing the directedness of an action to an outcome, and not just representing the action and the outcome.

**slide-98** Why not define  $R$  in terms of teleological function? This would enable us to meet the first condition but not the second. How could we tell whether an action happens because it brought about a particular outcome in the past? This might be done with insects. But it can't so easily be done with primates, who have a much broader repertoire of actions.

**slide-102** How about taking  $R$  to be causation? That is, how about defining  $R(a, G)$  as  $a$  causes  $G$ ? This proposal does not meet the first criterion, (1), above. We can see this by mentioning two problems.

[\*Might skip over-generate and discuss that as a problem for Rationality/Efficiency] First problem: actions typically have side-effects which are not goals. For example, suppose that I walk over here with the goal of being next to you. This action has lots of side-effects:

- I will be at this location.
- I will expend some energy.
- I will be further away from the front

These are all causal consequence of my action. But they are not goals to which my action is directed. So this version of  $R$  will massively over-generate goals.

Second problem: actions can fail. [...] So this version of  $R$  will under-generate goals.

**slide-107** 'an action can be explained by a goal state if, and only if, it is seen as the most justifiable action towards that goal state that is available within the constraints of reality' (Csibra and Gergely 1998, p. 255)

**slide-108** A goal is an outcome to which an action is directed. A goal-state is a representation of the outcome in virtue of which the action is directed to that outcome. So an intention is a goal state. By contrast, a goal is not a mental state at all. In order for this to be about \*pure\* goal ascription, we need to ignore the Csibra and Gergely's odd choice of terminology.

**slide-115** We start with the assumption that we know the event is an action.

**slide-118** Why normally? Because of the 'seen as'.

**slide-123** What does it mean to say that one means is better than another? There are different respects in which one action can be better than another as a means to some realising some outcome; for example, one action can require less effort than another, or one action be a more reliable way to bring the outcome about than another.

An action of type  $a'$  is a *better* means of realising outcome  $G$  in a given situation than an action of type  $a$  if, for instance, actions of type  $a'$  normally involve less effort than actions of type  $a$  in situations with the salient features of this situation and everything else is equal; or if, for example, actions of type  $a'$  are normally more likely to realise outcome  $G$  than actions of type  $a$  in situations with the salient features of this situation and everything else is equal.

**slide-125** Any objections?

I have an objection. Consider a case in which I perform an action directed to the outcome of pouring some hot tea into a mug. Could this pattern of inference imply that the outcome be the goal of my action? Only if it also implies that moving my elbow is a goal of my action as well. And pouring some liquid. And moving air in a certain way. And ...

How can we avoid this objection?

**slide-127** Doesn't this conflict with the aim of explaining \*pure\* behaviour reading? Not if desirable is understood as something objective. [explain]

**slide-128** Now we are almost done, I think.

**slide-129** We just need to add a clause ensuring that the goal in question is maximally desirable; this is an attempt to reduce overgeneration of goals.

**slide-130** OK, I think this is reasonably true to the quote. So we've understood the claim. But is it true?

**slide-131** How good is the agent at optimising the selection of means to her goals? And how good is the observer at identifying the optimality of means in relation to outcomes? **For optimally correct goal ascription, we want there to be a match between (i) how well the agent can optimise her choice of means and (i) how well the observer can detect such optimality.** Failing such a match, the inference will not result in correct goal ascription.

But I don't think this is an objection to the Teleological Stance as a computational theory of pure goal ascription. It is rather a detail which concerns the next level, the level of representations and algorithms. The computational theory imposes demands at the next level.

slide-133 It will work if we can match observer and agent: both must be 'equally optimal'. But how can we ensure this?

slide-134 How good is the agent at optimising the rationality, or the efficiency, of her actions? And how good is the observer at identifying the optimality of actions in relation to outcomes? **If there are too many discrepancies between how well the agent can optimise her actions and how well the observer can detect optimality, then these principles will fail to be sufficiently reliable.**

motor\_theory\_goal\_ascription

## The Motor Theory of Goal Ascription

slide-136 Two forms of goal ascription, representational and functional (Gallese and Sinigaglia 2011). In *representational goal ascription*, three things must be represented: an action, an outcome and the relation between this outcome and the action in virtue of which the outcome is a goal of the action. In *functional goal ascription*, the relation between action and outcome is captured without being represented. To say that this relation is *captured* is to say that there is a process which ensures that the outcome represented is a goal of the action.

slide-137 Motor representations cannot suffice for representational goal ascription. It is true that, in someone observing an action there can be motor representations of outcomes which, non accidentally, are the goals of the observed action. But this is not enough. There would have to be, in addition, a motor representation of an intention, or of a motor representation or of some other goal-state, or of a function. But there are no such motor representations.

slide-139 How does it ever come about that an outcome represented motorically in observing an action is an outcome to which that action is directed? First consider a parallel question about performing rather than observing actions. Suppose you are alone and not observing or imagining any other actions. When performing actions in this situation, outcomes represented motorically in you will normally be among the goals of your actions; that is, they will be outcomes to which your actions are directed. What ensures this correspondence between outcomes represented and goals?

slide-140 It is the role of the representation in controlling how the action unfolds. Representations of outcomes trigger planning-like motor processes whose function is to cause actions that will bring about the outcomes represented (Miall and Wolpert 1996; Arbib 1985; Rosenbaum et al. 2012).

**slide-141** Now return to observing rather than perform actions. What ensures the correspondence between outcomes represented motorically and goals when you are merely observing another act?

**slide-142** The answer, we suggest, is roughly that planning-like processes can be used not only to control actions but also to predict them. Let us explain.

**slide-143** There is evidence that a motor representation of an outcome can cause a determination of which movements are likely to be performed to achieve that outcome (see, for instance, Kilner et al. 2004; Urgesi et al. 2010). Further, the processes involved in determining how observed actions are likely to unfold given their outcomes are closely related, or identical, to processes involved in performing actions. This is known in part thanks to studies of how observing actions can facilitate performing actions congruent with those observed, and can interfere with performing incongruent actions (Brass et al. 2000; Craighero et al. 2002; Kilner et al. 2003; Costantini et al. 2012a). Planning-like processes in action observation have also been demonstrated by measuring observers' predictive gaze. If you were to observe just the early phases of a grasping movement, your eyes might jump to its likely target, ignoring nearby objects (Ambrosini et al. 2011). These proactive eye movements resemble those you would typically make if you were acting yourself (Flanagan and Johansson 2003). Importantly, the occurrence of such proactive eye movements in action observation depends on your representing the outcome of an action motorically; even temporary interference in the observer's motor abilities will interfere with the eye movements (Costantini et al. 2014). These proactive eye movements also depend on planning-like processes; requiring the observer to perform actions incongruent with those she is observing can eliminate proactive eye movements (Costantini et al. 2012b). This, then, is further evidence for planning-like motor processes in action observation.

So observers represent outcomes motorically and these representations trigger planning-like processes which generate expectations about how the observed actions will unfold and their sensory consequences. Now the mere occurrence of these processes is not sufficient to explain why, in action observation, an outcome represented motorically is likely to be an outcome to which the observed action is directed.

To take a tiny step further, we conjecture that, in action observation, **motor representations of outcomes are weakened to the extent that the expectations they generate are unmet** (compare Fogassi et al. 2005). A motor representation of an outcome to which an observed action is not directed is likely to generate incorrect expectations about how this action will unfold, and failures of these expectations to be met will weaken the representation.

This is what ensures that there is a correspondence between outcomes represented motorically in observing actions and the goals of those actions.

`marrs_levels`

## Marr's Threefold Distinction

Marr helpfully distinguishes computational description (What is the thing for and how does it achieve this?) from representations and algorithms (How are the inputs and outputs represented, and how is the transformation accomplished?) and from the hardware implementation (How are the representations and algorithms physically realised?)

**slide-145** One possibility is to appeal to David Marr's famous three-fold distinction between levels of description of a system: the computational theory, the representations and algorithm, and the hardware implementation.

This is easy to understand in simple cases. To illustrate, consider a GPS locator. It receives information from four satellites and tells you where on Earth the device is.

There are three ways in which we can characterise this device.

**slide-146** First, we can explain how in theory it is possible to infer the device's location from it receives from satellites. This involves a bit of maths: given time signals from four different satellites, you can work out what time it is and how far you are away from each of the satellites. Then, if you know where the satellites are and what shape the Earth is, you can work out where on Earth you are.

**slide-147** The computational description tells us what the GPS locator does and what it is for. It also establishes the theoretical possibility of a GPS locator.

But merely having the computational description does not enable you to build a GPS locator, nor to understand how a particular GPS locator works. For that you also need to identify representations and algorithms ...

**slide-148** At the level of representations and algorithms we specify how the GPS receiver represents the information it receives from the satellites (for example, it might in principle be a number, a vector or a time). We also specify the algorithm the device uses to compute the time and its location. The algorithm will be different from the computational theory: it is a procedure for discovering time and location. The algorithm may involve all kinds of shortcuts and approximations. And, unlike the computational theory, constraints on time, memory and other limited resources will be evident.

slide-149 So an account of the representations and algorithms tells us ...

– How are the inputs and outputs represented, and how is the transformation accomplished?

slide-150 The final thing we need to understand the GPS locator is a description of the hardware in which the algorithm is implemented. It's only here that we discover whether the device is narrowly mechanical device, using cogs, say, or an electronic device, or some new kind of biological entity.

slide-151 The hardware implementation tells us how the representations and algorithms are represented physically.

slide-152 How is this relevant to the teleological stance? It provides a computational description of goal ascription. Whereas the Motor Theory provides an account of the representations and algorithms

slide-153 I suggest that an account of radical interpretation is supposed to provide a computational description of social cognition; it tells us what social cognition is for and how, in the most abstract sense, it is possible.

slide-155 Now we've solved this: the Motor Theory of Goal Ascription is the solution.

slide-156 Also recall the question

How could the objects of categorical perception of speech be articulatory gestures?

The puzzle here is simple. Categorical perception of speech happens rapidly, and goal-directed actions are complex. How can something so complex be computed so quickly?

Put it another way: how is it that the articulatory gesture is identified?

'Humans [can] understand speech delivered at a rate of 20 to 30 ... phonemes per second' (Devlin 2006)

We've also solved this with the Motor Theory of Goal Ascription. (a) Motor processes are fast enough (production speed = comprehension speed); (b) The Motor Theory explains how motor processes can implement functional goal ascription, which is to say, how they can get from observed bodily configurations and joint displacements to the identification of goals to which the action is directed.

puzzle\_thought\_experience\_motoric



## A Puzzle about Thought, Experience and the Motoric

Motor representations occur when merely observing others act and sometimes influence thoughts about the goals of observed actions. Further, these influences are content-respecting: what you think about an action sometimes depends in part on how that action is represented motorically in you. The existence of such content-respecting influences is puzzling. After all, motor representations do not feature alongside beliefs or intentions in reasoning about action; indeed, thoughts are inferentially isolated from motor representations. So how could motor representations have content-respecting influences on thoughts?

**slide-161** This conclusion entails that motor representations have content-respecting influences on thoughts. It is the fact that one outcome rather than another is represented motorically which explains, at least in part, why the observer takes this outcome (or a matching one) to be an outcome to which the observed action is directed.

**slide-162** But how could motor representations have content-respecting influences on thoughts? One familiar way to explain content-respecting influences is to appeal to inferential relations. To illustrate, it is no mystery that your beliefs have content-respecting influences on your intentions, for the two are connected by processes of practical reasoning. But motor representation, unlike belief and intention, does not feature in practical reasoning. Indeed, thought is inferentially isolated from it. How then could any motor representations have content-respecting influences on thoughts?

**slide-164** In something like the way experience may tie thoughts about seen objects to the representations involved in visual processes, so also it is experience that connects what is represented motorically to the objects of thought.

[significance] This may matter for understanding thought about action. On the face of it, the inferential isolation of thought from motor representation makes it reasonable to assume that an account of how humans think about actions would not depend on facts about motor representation at all. But the discovery that motor representations sometimes shape experiences revelatory of action justifies reconsidering this assumption. It is plausible that people sometimes have reasons for thoughts about actions, their own or others', that they would not have if it were not for their abilities to represent these actions motorically. To go beyond what we have considered here, it may even turn out that an ability to think about certain types of actions depends on an ability to represent them motorically.

[consequence] One consequence of our proposal concerns how experiences of one's own actions relate to experiences of others' actions. For almost any

action, performing it would typically involve receiving perceptual information quite different to that involved in observing it. This may suggest that experiences involved in performing a particular action need have nothing in common with experiences involved in observing that action. However, two facts about motor representation, its double life and the way it shapes experience, jointly indicate otherwise. For actions directed to those goals that can be revealed by experiences shaped by motor representations, there are plausibly aspects of phenomenal character common to experiences revelatory of one's own and of others' actions. In some respects, what you experience when others act is what you experience when you yourself act.

**slide-166** The claim that there is experience of action is based on an earlier argument. I now want to review and then object to that argument. (The conclusion may be correct, but the argument does not establish it.)

**slide-168** The difference in differences ...

Here is the argument. Consider two sequences of sensory encounters: (a) a sequence of sensory encounters with two phonetic events that do not differ with respect to category (both are realisations of /d/, say), and (b) a sequence of sensory encounters with two phonetic events that do so differ (one is a realisation of /d/ the other of /g/, say).<sup>11</sup> Let the events encountered in the first sequence differ from each other acoustically in the same way and by the same amount as the events encountered in the second sequence differ from each other. (That it is possible to find two such pairs of events follows from the fact that we enjoy categorical perception of speech.) The two sequences are depicted in Fig. 3. Now:

**slide-174** This is less obvious than it might seem. We infer it from the facts about relevant acoustic features of the stimuli, and the way acoustic processes work. But we know that motor processes are also at work. And we can't assume that motor processes do not influence acoustic processes. In fact we know that motor processes affect acoustic experiences.

**slide-176** Repp and Knoblich (2007) asked expert and non-expert pianists to press two keys in sequence, where the first key was sometimes to the left, and sometimes to the right, of the second key. The key presses produced an *ambiguous tone pair*, that is, a pair with the property that the first tone is sometimes perceived as lower in pitch than the second whereas at other times it is perceived as higher in pitch (Deutsch et al. 1987). The tones always occurred in the same order regardless of which key was pressed first. By asking subjects to report how they perceived the relative pitches of the tones, Repp and Knoblich found that, for the expert pianists, the direction of the key presses influenced the perceived direction of the change in pitch. Could what influences experience in this case be not a motor representation but

merely the occurrence of a movement, or perhaps even the perception of a movement of the subject's own fingers? Against these possibilities, note that the effect was not observed in non-expert pianists: for them the direction of movement did not measurably influence the perceived pitches. Since the direction of movement was the same for both groups, if the influence were due to movement only we would expect it to occur irrespective of piano-playing expertise. Instead it seems likely that differences in expertise between the two groups of subjects affected how the movements they performed were represented motorically, and that these differences in motor representation are in turn what explains their perceptions of relative pitches.

It is not only in performing action that motor representation can influence experience: the same can occur in observing action. Thus in another experiment, Repp and Knoblich (2009) compared observing someone else perform a sequence of key presses with performing the same sequence oneself. They found the same effect on experiences of an ambiguous tone pair in expert pianists regardless of whether they were observing or performing the action. Sometimes, which judgement about pitch an experience provides a reason for depends on what is represented motorically; and this dependence is systematic, of course, for it reflects how pianos work. These studies, and others like them,<sup>2</sup> provide relatively direct evidence that motor representation can shape experience.

## References

- Ambrosini, E., Costantini, M., and Sinigaglia, C. (2011). Grasping with the eyes. *Journal of Neurophysiology*, 106(3):1437–1442.
- Arbib, M. A. (1985). Coordinated control programs for movements of the hand. *Hand Function and the Neocortex. Experimental Brain Research*, 10:111–129.
- Bonini, L., Maranesi, M., Livi, A., Fogassi, L., and Rizzolatti, G. (2014). Ventral Premotor Neurons Encoding Representations of Action during Self and Others' Inaction. *Current Biology*, 24(14):1611–1614.
- Brass, M., Bekkering, H., Wohlschläger, A., and Prinz, W. (2000). Compatibility between observed and executed finger movements: Comparing symbolic, spatial, and imitative cues. *Brain and Cognition*, 44(2):124–143.
- Browman, C. P. and Goldstein, L. M. (1986). Towards an articulatory phonology. *Phonology*, 3(01):219–252.

---

<sup>2</sup> See also Zwicker et al. (2010) who investigate effects of action on visual experience of motion, and Schütz-Bosbach and Prinz (2007) for a review.

- Calvo-Merino, B., Grèzes, J., Glaser, D. E., Passingham, R. E., and Haggard, P. (2006). Seeing or doing? influence of visual and motor familiarity in action observation. *Current Biology*, 16(19):1905–1910.
- Costantini, M., Ambrosini, E., Cardellicchio, P., and Sinigaglia, C. (2014). How your hand drives my eyes. *Social Cognitive and Affective Neuroscience*, 9(5):705–711.
- Costantini, M., Ambrosini, E., and Sinigaglia, C. (2012a). Does how I look at what you’re doing depend on what I’m doing? *Acta Psychologica*, 141(2):199–204.
- Costantini, M., Ambrosini, E., and Sinigaglia, C. (2012b). Does how I look at what you’re doing depend on what I’m doing? *Acta Psychologica*, forthcoming.
- Craighero, L., Bello, A., Fadiga, L., and Rizzolatti, G. (2002). Hand action preparation influences the responses to hand pictures. *Neuropsychologia*, 40(5):492–502.
- Csibra, G. and Gergely, G. (1998). The teleological origins of mentalistic action explanations: A developmental hypothesis. *Developmental Science*, 1(2):255–259.
- D’Ausilio, A., Pulvermüller, F., Salmas, P., Bufalari, I., Begliomini, C., and Fadiga, L. (2009). The Motor Somatotopy of Speech Perception. *Current Biology*, 19(5):381–385.
- Dejean, A., Solano, P. J., Ayroles, J., Corbara, B., and Orivel, J. (2005). Insect behaviour: Arboreal ants build traps to capture prey. *Nature*, 434:973.
- Deutsch, D., Kuyper, W. L., and Fisher, Y. (1987). The tritone paradox: Its presence and form of distribution in a general population. *Music Perception*, pages 79–92.
- Devlin, J. T. (2006). Are we dancing apes? *Science*, 314(5801):926–927.
- di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., and Rizzolatti, G. (1992). Understanding motor events: a neurophysiological study. *Experimental Brain Research*, 91(1):176–180.
- Fadiga, L., Craighero, L., Buccino, G., and Rizzolatti, G. (2002). Speech listening specifically modulates the excitability of tongue muscles: a tms study. *European Journal of Neuroscience*, 15(2):399–402.
- Flanagan, J. R. and Johansson, R. S. (2003). Action plans used in action observation. *Nature*, 424(6950):769–771.

- Fogassi, L., Ferrari, P. F., Gesierich, B., Rozzi, S., Chersi, F., and Rizzolatti, G. (2005). Parietal lobe: From action organization to intention understanding. *Science*, 308(5722):662–667.
- Gallese, V., Fadiga, L., Fogassi, L., and Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain*, 119(2):593–609.
- Gallese, V. and Sinigaglia, C. (2011). What is so special about embodied simulation? *Trends in Cognitive Sciences*, 15(11):512–519.
- Holt, L. L. and Lotto, A. J. (2014). The alluring but misleading analogy between mirror neurons and the motor theory of speech. *Behavioral and Brain Sciences*, 37(2):204–205.
- Jeannerod, M. (1998). *The Neural and Behavioural Organization of Goal-Directed Movements*. Oxford University Press, Oxford.
- Kilner, J., Vargas, C., Duval, S., Blakemore, S., and Sirigu, A. (2004). Motor activation prior to observation of a predicted movement. *Nature neuroscience*, 7(12):1299–1301.
- Kilner, J. M., Paulignan, Y., and Blakemore, S. (2003). An interference effect of observed biological movement on action. *Current Biology*, 13(6):522–525.
- Liberman, A. M., Isenberg, D., and Rakerd, B. (1981). Duplex perception of cues for stop consonants: evidence for a phonetic mode. *Perception & psychophysics*, 30(2):133–143.
- Miall, R. C. and Wolpert, D. M. (1996). Forward models for physiological motor control. *Neural Networks*, 9(8):1265–1279.
- Repp, B. H. and Knoblich, G. (2007). Action can affect auditory perception. *Psychological Science*, 18(1):6–7.
- Repp, B. H. and Knoblich, G. (2009). Performed or observed keyboard actions affect pianists’ judgements of relative pitch. *The Quarterly Journal of Experimental Psychology*, 62(11):2156–2170.
- Rosenbaum, D. A., Chapman, K. M., Weigelt, M., Weiss, D. J., and van der Wel, R. P. R. D. (2012). Cognition, action, and object manipulation. *Psychological Bulletin*, 138(5):924–946.
- Schütz-Bosbach, S. and Prinz, W. (2007). Perceptual resonance: action-induced modulation of perception. *Trends in Cognitive Sciences*, 11(8):349–355.

- Urgesi, C., Maieron, M., Avenanti, A., Tidoni, E., Fabbro, F., and Aglioti, S. (2010). Simulating the future of actions in the human corticospinal system. *Cerebral Cortex*, 20(11):2511–2521.
- Wolpert, D. M., Ghahramani, Z., and Jordan, M. (1995). An internal model for sensorimotor integration. *Science*, 269(5232):1880 –1882.
- Zhang, W. and Rosenbaum, D. A. (2007). Planning for manual positioning: the end-state comfort effect for manual abduction–adduction. *Experimental Brain Research*, 184(3):383–389.
- Zwicker, J., Grosjean, M., and Prinz, W. (2010). On interference effects in concurrent perception and action. *Psychological Research*, 74(2):152–171.