Editorial Manager(tm) for The Review of Philosophy and Psychology
Manuscript Draft

Manuscript Number: ROPP47

Title: From joint action to action control: Functional equivalence between acting together and acting alone

Article Type: Joint Action: What is Shared?

Keywords: Joint action; Forward models; Perception-action; Interpersonal coordination; Group flow

Corresponding Author: Mr Lincoln John Colling,

Corresponding Author's Institution: Macquarie Centre for Cognitive Science, Macquarie University

First Author: Lincoln John Colling

Order of Authors: Lincoln John Colling; John Sutton; William Forde Thompson

Abstract: Many forms of joint action occur in situations where time constraints make it difficult to coordinate actions by using language. In these situations, individuals need to be able to anticipate the actions of their co-actors. We provide an account of joint action coordination built on a foundation of individual action control. We draw on findings from action prediction tasks to show that functionally equivalent predictive mechanisms underlie both individual action control and joint action coordination. These predictive mechanisms may underlie the ability to both coordinate the actions of multiple effectors and to coordinate the actions of multiple people. These predictive mechanisms have also been implicated in the phenomenology of individual action. The functional equivalence between individual action and joint action, therefore, hints at the possibility that the same predictive mechanisms can also be used to explain the phenomenology of joint action and feelings of group flow.

Suggested Reviewers:

From joint action to action control:

Functional equivalence between acting together and acting alone

Lincoln J. Colling*, John Sutton
Macquarie Centre for Cognitive Science
Macquarie University
Sydney, NSW 2109
Australia

William F. Thompson
Department of Psychology
Macquarie University
Sydney, NSW 2109
Australia

***Corresponding author:**
Lincoln J. Colling

Macquarie Centre for Cognitive Science
Macquarie University
Sydney, NSW 2109
Australia
Email: lincoln.colling@mq.edu.au
Telephone: +(612) 9850 6729
Fax: +(612) 9850 6059

Keywords: Joint action; Forward models; Perception–action; Interpersonal coordination; Group flow

Review of Philosophy and Psychology manuscript No.
(will be inserted by the editor)

From action control to joint action

Functional equivalence between acting together and acting alone

the date of receipt and acceptance should be inserted later

Abstract Many forms of joint action occur in situations where time constraints make it difficult to coordinate actions by using language. In these situations, individuals need to be able to anticipate the actions of their co-actors. We provide an account of joint action coordination built on a foundation of individual action control. We draw on findings from action prediction tasks to show that functionally equivalent predictive mechanisms underlie both individual action control and joint action coordination. These predictive mechanisms may underlie the ability to both coordinate the actions of multiple effectors and to coordinate the actions of multiple people. These predictive mechanisms have also been implicated in the phenomenology of individual action. The functional equivalence between individual action and joint action, therefore, hints at the possibility that the same predictive mechanisms can also be used to explain the phenomenology of joint action and feelings of group flow.

Keywords Joint action · Forward models · Perception–action · Interpersonal coordination · Group flow

Address(es) of author(s) should be given

1 Introduction

Traditional philosophical analyses of joint action have focused on such problems as differentiating truly joint actions from seemingly similar actions that are not joint actions. Consider the following example from Searle (1990). A group of people converge on a common point in a park. This behaviour could either be a collective behaviour or it could be a collection of individual behaviours. What differentiates the two are the intentions that cause the behaviour. If the people have converged at the common point because each has the intention “it is raining and I am running to shelter”, then it is merely a collection of individual behaviours. What is required for it to be a collective behaviour or joint action is a different sort of intention. This might take the form of a collective intention such as “we intend to get ready to perform a ballet by means of me running to the gazebo” (Searle, 1990), or the form of a shared intention consisting of individual intentions to get ready to perform a ballet together with appropriate interrelations between those intentions (Bratman, 1993).

The purpose of these shared or collective intentions is to allow the individual agents to coordinate their actions in such a way so that they successfully carry out the joint action. For example, our shared intention to bake a cake together will result in behaviour that allows each of us to negotiate our part in the joint action. I can plan to mix the ingredients while I can tell you to grease the cake tin. Here, language can play a role in coordinating our actions so that we can realise our joint goal. However, in other cases of joint action such as playing music together, dancing together or precisely coordinating actions on a sports field, a shared intention, or joint goal, is not sufficient for successfully performing the joint action. Furthermore, there are time constraints inherent in these types of joint actions that make appeals to the coordinating power of language unworkable. Language can be used, for example, to decide on the time and place of our piano duet and to negotiate who will play the primo and who will play the secondo, but for the performance to be successful, what is needed is a means of coordinating, and synchronising, action execution across individuals. Furthermore, individuals cannot simply execute their actions in response to the actions of their co-actors, because this will lead to a breakdown in synchronisation. Rather, individuals require the ability to predict or anticipate the actions of their co-actors, and the ability to coordinate their action execution on the basis of these predictions, and this needs to be possible without the use of language.

This paper focuses specifically on joint action planning in the absence of language. In particular, we provide a framework for understanding how it is that individuals participating in joint actions are able to predict the actions of their co-actors. This framework is firmly based on an account of mechanisms that underlie individual action control. Models of individual action control have relied on the notion of prediction to explain how people are able to plan and execute their actions. For instance, predictive models play an important role in understanding how people coordinate actions between multiple effectors (e.g., Witney, 2004) or coordinate movement between the hand and the eyes (e.g., Miall and Reckess, 2002). The model outlined here suggests that predictive mechanisms similar to those that underlie action control and coordination within individuals also underlie action prediction and coordination between individuals involved in joint actions.

The outline of the paper is as follows. First, we provide a basic introduction to control theory and introduce predictive models and emulators. Next, we outline how predictive models have been implicated in theories of action control. We suggest that the action control system is used offline during joint action thereby allowing individuals to emulate and predict the actions of their co-actors. We present evidence that shows that during joint action individuals represent the goals of their co-actors in a manner that is functionally equivalent to how they represent their own goals. We argue that these goals can then be fed into the action control system, taken offline, in order to, first, generate the control commands that would be issued in order to realise that goal and, second, to generate predictions about how the action will unfold and the sensory consequences that will occur. We also present evidence from studies on action synchronisation that suggest how it is possible to reconstruct these control commands. Finally, we show that by co-representation of co-actors' goals, and by emulating their actions to predict how these goals will be realised, individuals are able to form joint action models that are functionally equivalent to individual action models. This allows for the coordination of group action in a manner that is functionally equivalent to individual action coordination.

The joint actions that are the target of this paper—actions such as playing music together or dancing together—differ from joint actions such as baking a cake together not only in the requirement that actions are coordinated by means of predictive mechanisms and not language, but also in their phenomenology. There is a distinct phenomenology of playing music together that is simply not present when baking a cake together. Ultimately, we suggest that our framework might provide clues to the basis of the phenomenology of acting together. Predictive mechanisms for action control have been implicated in the phenomenology of individual action. We suggest that similar mechanisms underlie action control in individuals and between individuals engaged in joint action; therefore, this raises the possibility that these predictive mechanisms might give rise to the phenomenology of joint action and experiences such as group flow, which are often reported by, for example, professional musicians and sportspeople.

2 Predictive models in control theory

In engineered systems goal-directed behaviour is ordinarily implemented by means of closed-loop control. A simple example is a room heater consisting of a heating element, a temperature control, and a temperature sensor (thermostat). Once a temperature has been selected, the system *controller* transforms the target room temperature, or goal state, into commands that are then sent to the target system (the heating element). These commands might consist of, for example, specifications for the amount of current to be passed through the heating element. The controller is said to perform an *inverse mapping* because it transforms the goal state (target room temperature) into control commands to be sent to the target system. The control commands produce the behaviour in the target that is required to bring about the desired goal state. The target system is said to perform a *forward mapping* because it transforms control commands into the goal state. In closed-loop, or feedback control, the temperature sensor can be used to measure the current state of the system (current room temperature) and compare it with the desired goal state. Any discrepancy between the current system state and the goal state can be fed back to the system controller so that adjustments can be made to the control signals. This allows the controller to pass more current to the heating element if the room is too cold or less current if the room is too hot. Feedback control has a distinct advantage because it allows the system to adjust its behaviour to changes in the environment or to compensate for inaccuracies in the initial control commands. Without feedback (what is known as open-loop control) the control signals need to be precisely defined at the onset, because, without feedback, there is no way to compensate for inaccuracies.

A major obstacle in implementing feedback control is that the measurements received from the sensors can be inaccurate either due to inherent uncertainties in the measurements or because of temporal delays in receiving feedback from the sensors. It is possible to compensate for these inaccuracies by using a model, or emulator, of the target system's behaviour. Grush (1997) has illustrated the usefulness of emulators by using the example of a heater on a space station located a significant distance away from the controller placed on earth. Because there is a limit on how quickly radio signals can travel between the controller and the space station, feedback received from sensors will actually provide a measure of the temperature on the space station as it was at some time in the past. The commands sent on the basis of these measures may, therefore, be inappropriate for the system's current state. For instance, by the time the controller issues a command to turn off the heaters the temperature on the space station will be well in excess of the goal temperature. The temperature will then begin to fall until it falls well below the goal temperature and it will continue to fall until the controller eventually sends a signal to restart the heaters. Rather than maintaining the goal temperature, feedback delays will result in the temperature oscillating around the goal without ever locking on to it.

Grush (1997) suggests that this problem can be overcome my means of an *emulator*. A emulator is a *forward model*, so called because it is a model of the forward mapping performed by the target system. Before the space station leaves, an emulator can be trained to exactly mimics the forward mapping behaviour of the space station—that is, it maintains the relationship between particular control signals and particular changes in temperature. The space station can then be sent into space while the emulator is kept on earth. When the controller issues control signals one copy of the control signal is sent to the space station while another copy is sent to the emulator. The emulator responds to the control signals just as the real space station would, and the feedback from the space station can then be substituted with predicted feedback from the emulator. As the emulator is also located on earth the feedback will not be delayed.

A second problem that can arise during feedback control is that the measurements from the sensors can be inaccurate. For example, temperature sensors in the room may contain random noise so that they never truly reflect the actual temperature of the room. If these sensor readings are used for feedback control then the behaviour of the system would also be susceptible to sensor noise. This problem can also be overcome by means of an emulator. However, unlike the space station example, the actual sensor readings are not completely substituted by the predicted temperature from the emulator. Rather, the predicted temperature and actual sensor readings can be combined in order to produce a more accurate estimate of the real room temperature. The contribution of each value (predicted and actual) to the final feedback value can also be varied according to the expected accuracy of each measurement (Golnaraghi, 2010).

The power of emulators is two-fold. First, they provide a means for controlling a system in real-time under conditions of delayed feedback and imperfect knowledge of the systems state. In the next section, we show how these functions maybe important for action control. Second, the whole system can be taken off-line and the emulator can be used as a model of the target system's behaviour without the target system actually producing any output. This can be done by only sending control commands to the emulator and not the target system. This may be particularly useful in, for example, motor imagery. We will argue that this also occurs during joint action where an individual's action control system can be taken offline and used as an emulator of another agent's action control system. This allows individuals engaged in joint action to predict and anticipate the actions of their co-actors.

3 Prediction and action control

Emulators such as those described in the preceding section have been implicated in many different areas of cognition (see Grush, 2004). However, the strongest evidence of a role for emulators in cognition comes from action control. The model of action control developed by Wolpert (1997) relies heavily on notions borrowed from control theory, including emulators. As with the heater example, an *inverse model* is used for computing what motor commands would be needed in order to bring about a particular goal state given such factors as the current state of the limb (its position, joint angles and so forth). The motor command is then sent to the limbs in order to produce the actual movements. In addition, a copy of the motor command is sent to an emulator. The emulator can be divided into two parts: the *forward output model* used to predict the sensory consequences that will result from performing the action, and the *forward dynamic model* used to predict the change in limb position.

The predictions generated by these forward models can be used for several purposes. For instance, forward modelling can be used to anticipate and compensate for the sensory effects created by movements of, for example, the eye. This can be used to selectively filter the sensory consequences of self-movement thereby making it possible to distinguish between the sensory effects caused by self-movement from movement in the environment. Forward models also allow the action control system to bypass feedback delays which allows for fine-tuning of movements without having to wait for actual sensory feedback from the periphery. Furthermore, because sensory signals might not always provide accurate information about the current state of the system, sensory signals can be combined with forward models in order to derive better state estimates (Miall and Wolpert, 1996; Wolpert, 1997).

The ability to predict the sensory consequences of an action may also play a role in action planning. For instance, ideo-motor theories of action planning suggest that actions are planned on the basis of predicted effects. For example, Hommel (2003; see also Elsner and Hommel, 2001) has suggested a two-stage model to explain the development of voluntary action control. In the first stage, actions are generated in a random fashion. These actions lead to perceivable changes in the environment and the body. These action effects, or perceptions, then become associated with the action used to produce them. This process is analogous to the training of the inverse and forward models that, for example, Wolpert (1997; see also Wolpert and Kawato, 1998) has suggested to underlie action control. Once action and action effects have become strongly associated, particular actions can be selected by activating the perceptual codes that correspond to the desired goal. Due to the strong association between perceptual codes and action codes, activation of a perceptual code partially activates its associated motor codes. This, in effect, allows action selection on the basis of predicted perceptual effects. Furthermore, this suggests that action both creates perception and is the consequence of perception, and perception is both the stimulus for action and the result of action. It has been argued that this implies that perception and action must share a common representational code (Hommel, 2009). Common coding of perception and action also has important implications for joint action. With no clear division between perception and action, this suggests that the same representations are involved in both producing actions and perceiving actions (e.g., see Knoblich, 2008).

The principle of common coding has been extended by Hommel et al (2001) into the theory of event coding. This framework combines the principle of common coding with the assumption that stimuli are represented in terms of collections of stimulus codes (such as colour, shape, or spatial location) and action codes (such as leftward movement or hand of responding). Stimulus codes and action codes are then represented together in terms of abstract feature codes in a common coding system. This framework has been used to explain a number of perception–action interactions including, for example, the Simon task which we describe in the next section.

4 Common mechanisms for individual and joint action

Emulators may also play a role in joint action control. In particular, emulators and internal models similar to those that underlie our ability to engage in goal-directed actions might also underlie our ability to predict the actions of other agents. For example, it may be possible to use inverse models in order to predict what actions a conspecific might use in order to realise a particular goal (Csibra, 2008). Forward models, on the other hand, might be used to generate predictions about the unfolding motor acts produced by observed agents or to predict the sensory consequences of actions performed by observed agents (Csibra, 2008; Wilson and Knoblich, 2005).

4.1 Co-representation of goals in joint tasks

In order to engage in joint action it is useful to know what actions one's co-actor will select in order to realise a particular goal. To do this, it is necessary to represent their goal in some manner. Studies by Sebanz and colleagues suggest that when co-acting, individuals automatically take into account the task or goals of their co-actors and represent them in a manner that is equivalent to how they represent their own goals. This is evident in the observation that a co-actor's task can influence individual action planning. For example, Sebanz et al (2003) tested participants on joint and individual versions of a spatial response compatibility task (Simon task). In the individual version, the typical finding is that when participants are asked to respond to a relevant stimulus feature while ignoring an irrelevant spatial feature of a stimulus, the spatial feature facilitates the response when it is spatially compatible with the response. So, for example, a response to a tone stimulus with the right hand might be facilitated when the tone is presented to the participant's right ear (e.g., Simon, 1969). The standard finding can be explained from an event coding perspective. On this view, the spatial nature of the response primes the spatial processing of the stimuli because they share a common representational format. When a stimulus appears, if the spatial features of the stimulus overlap with the spatial features of the response then that response is facilitated relative to the case where the features do not overlap (Hommel, 2009). Typically, this effect is only seen when participants have responses with a spatial dimension—for example, responding with the left hand to one stimulus and the right hand to another—and are not seen in, for example, go/no-go tasks where participants make a single response to only one stimulus and not to the other.

The setup for Sebanz et al's (2003) version of the Simon task included both individual and joint versions of the task. In the individual conditions, participants performed either the standard two-choice task or a go/no-go task. In the standard two-choice condition participants sat directly in front of a monitor displaying the stimuli (a hand) and made a response to the colour of a ring on the hand with either a left or a right button press. Participants were asked to ignore the direction that the hand was pointing. In the individual and joint go/no-go conditions participants sat to one side of the monitor and made a response to only one colour ring and no response to the other. The results of the experiment showed a typical spatial compatibility effect for the individual two-choice condition with responses facilitated when the hand pointed in the direction that was compatible with the button press (i.e., left button when the hand pointed left). No effect was observed for the individual go/no-go condition. The key condition was the joint go/no-go condition. Rather than mirroring the results of the individual go/no-go condition, the joint go/no-go condition mirrored the individual two-choice condition. That is, participants performed the joint go/no-go task as if they had two action possibilities available. Importantly, this was only the case when the pair (participant and co-actor) actually had two choices available, and the effect did not occur when the participant was paired with a non-co-acting individual. This was the case even when the individuals were not provided with feedback about their co-actors' actions, but were merely aware of their task (or action goals).

Sebanz et al's (2003) findings have been extended by examining whether the joint Simon effect is dependent on being able to conceive of one's co-actor as an intentional agent. Tsai et al (2008) had participants perform an individual go/no-go Simon task under the belief that they were either co-acting with a hidden human agent or a computer. The results showed that a joint Simon effect only occurred when participants believed they were co-acting with a human agent. When participants believed they were co-acting with a computer no Simon effect was observed (the standard result for individual go/no-go Simon tasks).

Taken together, the results of the joint Simon task suggest that when co-acting, individuals automatically represent the action goals of their co-actors when planning their own actions (Knoblich and Sebanz, 2008). When co-actors are not intentional agents, and therefore incapable of forming action goals, this goal representation does not occur. Furthermore, when co-representation of goals does occur, co-actors' goals are represented in a manner that is functionally equivalent to how they represent their own action goals thus creating a shared task representation.

4.2 Action models in prediction of observed action

Once a co-actor's goals have been represented, how are their actions predicted? These goals are fed into the inverse model that is used to plan one's own actions, the inverse model then *predicts* the motor commands that would be used to realise this goal. When limited information is known—for example, only the goal—the predicted commands from the inverse model can be rather inaccurate. Many actions can be employed to realise a particular goal, and without detailed knowledge about the constraints and capabilities of the co-actor's action system it is difficult to constrain the possibilities. One way to constrain these possibilities is to provide additional information to the inverse model. Such additional information might be information about the current state of the co-actor's limbs or information about any actions they might currently be performing. Csibra (2008) suggests that visual analysis of the movements and bodies of other agents may be a crucial step in reconstructing the motor codes driving their actions. If estimates of the motor codes can be generated based on visual analysis of the current movement, these estimates could be used to drive the predictions of the forward model. The discrepancy between the predicted actions/sensory consequences and the actual actions/sensory consequences could then be used to refine the estimates of the motor codes.

Once the motor commands have been reconstructed they can then be fed into forward models in order to generate predictions about the unfolding action trajectory or to generate predictions about the sensory consequences of the observed actions. Wilson and Knoblich (2005) have suggested that the output of these forward models can serve as the basis for anticipatory action planning or in aiding perception of conspecifics. For instance, when tracking the movements of conspecifics in suboptimal conditions vision may be occluded by objects or other agents. As the movements of the observer and the observed agent are subject to similar (but not identical) constraints, the observer can use their forward model to infer the trajectory of partially occluded actions.

Several lines of evidence have been used to support the idea that the same mechanisms recruited by action production are also recruited for action prediction. For example, neurophysiological evidence suggests that overlapping neural populations in motor regions such as premotor cortex (for a review see Rizzolatti and Craighero, 2004) and the cerebellum (e.g., Calvo-Merino et al, 2006) are involved in both execution and the observation of actions. These premotor regions also exhibit a topographic organisation (Buccino et al, 2001) similar to that observed in the so called motor homunculus (see Penfield, 1954). There are, however, competing interpretations for the neurophysiological data (e.g., see Gallese, 2009, and Jacob, 2008, for competing views) and, therefore, direct behavioural evidence may be more compelling. Behavioural evidence that individuals employ their forward action models not only when producing actions but also when predicting actions performed by others comes from an experimental paradigm examining differences in action prediction accuracy when observers view recordings of actions that were either self-produced or produced by another person. The logic of this paradigm is that if action prediction involves emulating the observed action with the observer's action model then idiosyncrasies of the observer's action model should interfere with the predictions they generate. This interference should be minimised, and prediction accuracy enhanced, when observers generate predictions about self-produced actions because the idiosyncrasies of the predicting system are matched by the idiosyncrasies of the predicted system. This self-prediction advantage has been termed an authorship effect.

Several experiments conducted by Knoblich and colleagues have employed this paradigm. For example, Knoblich and Flach (2001) examined the ability of participants to predict the landing position of a thrown dart from a recording of the action. Participants based their predictions on videos that depicted the throwing action up until the point of release, thus excluding any information about the dart's flight trajectory. Consistent with the forward model hypothesis, the results showed that participants were significantly more accurate at predicting the landing position of darts they had thrown themselves. Further experiments by Knoblich et al (2002) found similar effects when participants were asked to predict whether an observed pen stroke was produced in isolation or as part of a character.

If action prediction is to be useful for joint action coordination then the predictions generated by the emulator system must be available in advance of incoming sensory information. One method to test how rapidly these predictions are available is to employ a synchronisation task, because this places a time constraint on prediction. As with the prediction experiments discussed above, it is possible to examine whether an observer uses their own action model during action synchronisation by measuring synchronisation accuracy with self-produced actions compared with other-produced actions.

Flach et al (2003) examined synchronisation accuracy using dynamic pen traces. In this study, participants first drew zig-zag patterns on a drawing tablet. At a later stage, participants were presented with dynamic displays of the zig-zag traces. They were asked to synchronise a button press with the peaks of the trace and synchronisation accuracy was measured as the time difference between the occurrence of the peak and the button press. The results showed that timing error was reduced when participants were synchronising with self-generated patterns, compared with other-generated patterns. This suggests that information from the observers action model is available rapidly enough to be useful for action planning.

Interestingly, the authorship effect observed in this study only emerged in the later experimental trials. One explanation for this is that in order for the forward output model to accurately predict the sensory consequences of an action, the model needs accurate information about the current state of the effectors producing the action and accurate information about the motor codes corresponding to the action. The stimuli used in this study did not contain any information about the current state of the limbs and, therefore, this information would need to be reconstructed from the sensory consequences (pen traces) of the action.

This could be done if estimates are generated about the effectors' state together with estimates about the current motor codes. These estimates could be fed into the forward model, and the discrepancy between the output of the forward model and the actual observed sensory consequences could be used to update these estimates. Once the correct motor codes and state variables have been learned (or rather links have been generated between the motor codes and the observed traces) these can be used with the forward model to generate predictions. This process would not alter the internal dynamics of the forward model and, therefore, observers would still use their own action model to generate the predictions. The use of their own action model explains the presence of the authorship effect in later trials.

One way to circumvent the problem of having to learn the links between sensory effects and actions is by using participants with particular motor expertise, such as musicians, who would already have learned the appropriate mappings between actions and sensory effects. Keller et al (2007) asked expert musicians to synchronise a piano performance with recordings of either self-produced or other-produced music. The results showed that pianists were more accurate when playing a piano duet with a recording of themselves than they were when playing the duet with another person. In contrast to Flach et al (2003), the authorship effect was present in the first block and did not change in magnitude as the experiment progressed. These results are to be expected because trained pianists have more direct access to the links between perceived sensory consequences and the actions required to produce them.

Colling et al (2010) took an alternative approach to circumvent the problem of reconstructing the motor codes and state variables by modifying the paradigm of Flach et al (2003) so that participants were shown depictions of moving bodies. During the movement phase, participants were asked to produce arm movements by pretending to draw zig-zag patterns on an imaginary blackboard. These actions were recorded using motion capture. The motion capture data was then reconstructed as an animated character. During the test phase, participants were asked to synchronise a button press with the arm movements of the character, and timing error was measured. As in the Flach et al (2003) study, enhanced synchronisation accuracy was observed for self-generated actions. However, unlike Flach et al (2003), this authorship effect was present from the first block. This is to be expected if visual analysis of the limbs and limb movement is used to reconstruct the motor codes used to produce the action. With more immediate access to the motor codes, this information did not need to be learned during the course of the experiment.

A follow up study by Colling et al (2009) provided further support for the benefit of providing bodily information in the stimulus. Unlike previous studies that examined the difference between self-synchronisation and other-synchronisation, this study required all participants to synchronise with other produced actions and instead varied the type of information that was available in the stimulus. The authors found that participants were significantly more accurate at synchronising with an animated character than they were at synchronising with a single moving point-light. This study also compared synchronisation accuracy for individuals who had motor experience with the observed actions with those that did not. The synchronisation advantage that resulted from providing bodily information was restricted to participants with motor experience. This suggests that experienced and inexperienced observers may employ different mechanisms for action prediction. Inexperienced observers may employ mechanisms that are not dependent on mapping the exact action onto their motor system (see Schubotz, 2007). Indeed, neurophysiological evidence suggests that the degree to which an observer can map observed actions onto their motor system is dependent on their motor experience (Calvo-Merino et al, 2005, 2006). Importantly, the difference between experienced and inexperienced observers suggests that the effect of limb information cannot be explained by, for example, the added visual complexity or size of the animated character. This supports the idea that the synchronisation differences observed in the motor experience group have a motoric source.

Taken together, these studies show that when generating predictions about merely observed actions, observers' motor dynamics influence the accuracy of the predictions that they generate. This can be explained by common mechanisms for action production and action prediction. As sensorimotor prediction during internally generated (performed) action is performed by means of forward models it seems likely that functionally equivalent forward models are involved during sensorimotor prediction for externally generated (observed) actions.

4.3 Coordinating actions in individuals and groups

From the evidence cited above it is clear that common mechanisms underlie both action control and action prediction. Furthermore, the mechanisms that underlie action prediction can be directly recruited to assist action control in time-critical situations by allowing individuals to predict or anticipate the actions of their co-actors. This has obvious applications in joint action. For many types of joint action, understanding the goals of one's co-actors or representing a joint goal is not sufficient for successfully engaging in joint action. For example, in ensemble music and dance performance actions need to be coordinated with a high degree of temporal precision; this cannot be achieved through joint goal representation alone. For joint actions such as these, it is necessary to temporally coordinate action production between multiple co-actors. This can be made possible by jointly representing and predicting self-generated and other-generated actions, in addition to self-goals and other-goals, within a single system.

Experiments have shown that the rules that govern, for example, the coordination of limbs work equally well regardless of whether both limbs belong to the same person or to different people (Spivey, 2007). For example, Kelso (1984) asked participants to engage in a bimanual coordination task which involved holding their hands out in front of themselves with one hand facing palm downwards and the other hand facing palm upwards. Participants were then asked to engage in simultaneous wrist flexion and extension so that the orientation of each hand (in the horizontal plane) repeatedly alternated. Kelso (1984) found that as cycling frequency increased there was an abrupt change from the original anti-phase movements to in-phase movements (with both hands oriented in the same way). Consistent with ideo-motor theories of action planning, it has been suggested that this phase transition occurs because of a bias towards perceptual symmetry in action planning (Mechsner et al, 2001). What is particularly interesting for our present purposes, however, is that identical effects are observed when the limbs involved are spread across two individuals. For example, Schmidt et al (1990, experiment 2) had two participants sit on chairs next to each other in a manner that allowed them to freely swing their legs. Participants were asked to swing their outside leg (the leg furthest from the other participant) in either an in-phase or anti-phase manner. As with earlier experiments involving a single individual, the pair of participants found in-phase leg swinging stable at all speeds. However, anti-phase leg swinging was only stable at lower speeds and as speed increased there was a rapid and abrupt change from anti-phase to in-phase movements. Taken together, the results of the single case and joint case suggest that functionally equivalent mechanisms are involved in intra-personal and inter-personal action coordination.

Experimental results from Knoblich and Jordan (2003) have similarly suggested functional equivalence between acting alone and acting together. Participants were given the task of controlling a circular ring (tracker) on a computer monitor so that it tracked the movement of a smaller dot (target). The target moved horizontally across the screen and changed direction at the edges. Two response buttons were used to control the speed of the tracker. The left response button increased the tracker's speed when the tracker was moving towards the left of the screen and decreased the tracker's speed when it was moving towards the right of the screen. The right response button had the opposite effect. In order to change the direction of movement the tracker first had to come to a complete stop.

Participants performed the tracking task either individually or in pairs with each participant controlling one button. The results showed that groups were able to learn to perform the task as well as individuals and in the same manner as individuals. However, this only occurred when groups were given feedback about the actions of their partner. It was suggested that without feedback, participants were unable to distinguish the contribution of each participant to the movement of the tracker. But when feedback was presented, this allowed participants to emulate the actions and action effects of their partner along with representing their own actions and action effects (Knoblich and Sebanz, 2008). This parallel emulation allowed them to construct a joint action representation equivalent to an individual action representation thus permitting joint action coordination that was functionally equivalent to individual action coordination.

When engaged in individual actions such as juggling, individuals are required to coordinate the actions of two distinct effectors. Being able to predict how one's own actions will influence the juggling balls has obvious benefits when trying to plan and coordinate the actions of the two hands (Wolfensteller, 2009). Juggling is by no means easy, but it has the advantage that the actions and action goals are represented in a single system and predictions can be generated by a single motor system designed for that task. The case of coordinating the actions of two individuals in time critical situations seems far more difficult. However, the model we have presented here suggests that the gulf between the nature of the mechanisms involved in the individual case and the joint case is not so large. When engaged in joint actions individuals automatically represent the tasks of their co-actors in a manner that is functionally equivalent to how they represent their own tasks. Furthermore, they predict the actions and action effects of their co-actors using the same mechanisms they use to predict their own actions and action effects. And the system of individuals engaged in a joint task coordinates action execution in the same way an individual coordinates action execution in an individual task.

5 Phenomenology of joint action

The similarities between action coordination in the joint case and in the individual case may have further implications. The predictive mechanisms involved in action control have been implicated in the phenomenology of action. This raises the possibility that the use of similar predictive mechanisms during joint action may help explain the phenomenology of joint action.

Frith et al (2000) have used the model of action control developed by Wolpert (1997) to explain the sense of agency or the sense of oneself being the cause of an action. In its simple form, this model, known as the comparator model, states that a sense of agency is derived by comparing the predictions generated by the forward model with, for example, actual sensory feedback. If the predicted sensory feedback matches actual sensory feedback, then the action is perceived as self-produced—that is, a sense of agency is felt over the action. This effect can be illustrated by the case in which movement of the limb is caused by an external force (for example, as a result of electrical muscle stimulation). As the movement was not intended, the initial motor command, a copy of which ordinarily drives the functioning of the forward models, is missing. As a result, no sensory feedback is predicted by the forward model. When sensory feedback is perceived, as a result of the limb's movement, a discrepancy between the predicted feedback and the actual feedback occurs, which prevents a sense of agency being felt over the action.

Although it may seem that a veridical sense of agency is fairly robust in healthy people, this is in actual fact not the case. The sense that we are both the cause of our actions and not the cause of merely observed actions is very fragile (Wegner, 2002). A classic example of the malleability of agency comes from experiments on vicarious agency. These experiments, which employ the helping hands paradigm, demonstrate that people may feel a sense of agency over actions they are not in fact performing. In order to create this illusion, Wegner et al (2004) had participants stand in front of a mirror while another person (the hand helper) stood behind them hidden from view. The hand helper extended their arms forward so that one arm was positioned on either side of the participant as if they were the participant's own arms. A series of instructions to perform simple movements (such as "wave hello with your right hand") were then given either to the hand helper alone or to both the participant and the hand helper. The instructions given to the participant could either be the same as the instructions given to the hand helper (Experiment 1 and 3) or could vary between being consistent or inconsistent (Experiment 2). Once the movements had been completed, participants were asked a series of questions including questions designed to probe their feelings of agency over the hand helpers actions (for example, "How much control did you feel that you had over the arms' movements?").

The results of the experiments showed that participants reported a sense of agency over the movement of the hand helpers' hands when they heard movement instructions that were consistent with the movement of the hands. (It should be noted that a sense of agency is not all-or-none but, as these experiments suggest, is graded). This effect persisted even when participants made simultaneous subtle movements of their own hands in a manner inconsistent with the movement of the hand helpers hands' (Experiment 3). Wegner et al (2004) suggest that this feeling arises because when participants are able to hear the movement instructions they are able to anticipate the movement of the hands.

The results of the helping hands paradigm have ordinarily not been interpreted from a forward-modelling perspective. Both proponents (e.g., Carruthers, 2010) and critics (e.g., Synofzik et al, 2009) of the comparator model have claimed that the helping hands paradigm poses a problem for the comparator model, because it is assumed that there is no motor prediction with which to compare actual sensory feedback. However, considered as a joint task, the model we present here suggests that motor prediction does occur during the helping hands experiments. The experiments by Sebanz and colleagues (Sebanz et al, 2003, 2005, 2006) have demonstrated that when co-acting with another individual, people automatically represent the action goals of their partners. Furthermore, experiments on action prediction and action synchronisation demonstrate that forward models are employed to predict merely observed actions. If predictions resulting from forward models for action control can explain a sense of agency over our own actions (Frith et al, 2000; Sato and Yasuda, 2005), then common mechanisms for action production and action perception, along with co-representation of co-actors tasks, might help to explain the sense of agency over merely observed, but predictable, actions. While merely being able to anticipate observed actions or action effects may not lead to a full blown sense of agency (for example, Sato, 2009 reports significantly higher levels of agency for self-produced actions compared to observed, but predicted, actions), it may be enough to cause subtle disruptions in the sense of agency during some instances of joint action.

Sebanz et al (2006) and Hove (2008) have also suggested that joint action can lead to disruptions in the feeling of agency. For instance, Sebanz et al (2006) suggests that when co-actors act together by producing actions that have similar effects, confusions may occur over who is the agent responsible for the action effects. In turn, Hove (2008) has proposed that during synchronised action, difficulty in distinguishing self-produced from other-produced actions can lead to a blurring in the distinction between self and other.

Hove's (2008) argument builds on the claim that when performing actions we are able to attribute our own actions to ourselves when we accurately predict the sensory consequences of the action. However, when observing actions we use functionally equivalent action models to predict the actions of our co-actors. This means that action attribution based on prediction accuracy might be ineffective. Action attribution may also occur because we do not actually produce the actions we observe and, therefore, it is possible to distinguish self and other produced action on the basis of motor output. However, when engaged in synchronous action where individuals mimic each other's action in time and form, lack of motor output is also an ineffective cue. In these situations, Hove (2008) suggests that the distinction between self and other can dissolve, in a kind of blurring that might lead to enhanced social bonding. Examining this idea further, Hove and Risen (2009) asked participants to tap either synchronously or asynchronously with an experimenter. An affiliation measure was then recorded by asking participants how likable they found the experimenter. Affiliation scores were reliably modulated by the degree of synchronisation between the participant and the experimenter, suggesting that engaging in synchronised behaviour facilitates social bonds.

Hove and Risen (2009) suggest that their results may help to explain the social bonding effect of joint music performance. Freeman (2000) has argued that music, in preliterate cultures, likely took the form of synchronised social actions comprised of rhythmic chanting, clapping, and drumming. Furthermore, highly constrained, rhythmically synchronised actions (such as group chanting) are common in religious ceremonies even today. Freeman (2000) has suggested that music evolved as a technology for group formation, and Hove and

Risen (2009) suggest that blurring of the self–other distinction may be the means by which this occurs.

Agency confusion might not only occur during imitative actions, but might also occur during ensemble music performance or joint sport performance. Sebanz et al (2006) suggests that during ensemble music performance agency confusion might actually be preferred and musicians might strive to experience a sense of agency that transcends the self. This type of agency confusion may also lie at the heart of what Sawyer (2006) has termed *group flow*, and the model presented here suggests how group flow might arise.

The term group flow is used to describe the state of a group when they perform as a single unit. When experiencing group flow “each of the group members can even feel as if they are able to anticipate what their fellow performers will do before they do it” (Sawyer, 2006, p. 158). Berliner (1994) provides many firsthand accounts of ensemble jazz musicians that illustrate the experience of group flow, which jazz musicians describe as “striking a groove”. Berliner (1994) states, “the highest points of improvisation occur when the group members strike a groove together” (p. 388). Striking a groove allows “everything to come together in complete accord” (Harold Ousley, quoted in Berliner, 1994 p. 349).

Jazz musicians often practice extensively in order to develop precise timing, and work hard to ensure that their playing is appropriately synchronised (Berliner, 1994). Keller (2008; see also Keller et al, 2007) has suggested that the ability to engage in rhythmically coordinated music performance may be partially mediated by predictive action models. Extensive practise may allow fine-tuning of these predictive models so that musicians are able to accurately predict the actions of their co-performers (see Cross et al, 2009, for an example of training effects on action emulation for dance). Indeed, the musicians interviewed by Berliner (1994) often report the ability to accurately predict the actions of the other performers. For example, George Duvivier (quoted in Berliner, 1994, p. 390) reports that, “I’ve experienced times in which it was almost like I've been able to read a soloist's mind. I'll play a phrase, like a descending passage, at the same time he does." Some performers even report feelings of disembodiment: "I remember the first time I experienced that floating, out-of-the-body feeling," (Leroy Williams quoted in Berliner, 1994, p. 413). We suggest that these feelings of disembodiment and group flow result in part from agency confusion that occurs because the action systems of performers in these ensembles are finely tuned to those of their co-actors. In attempting to synchronise their musical performance they predict how their fellow musicians will perform, and the accuracy of these predictions leads to agency confusion. Furthermore, Hove and Risen's (2009) experiment suggests that "striking a groove" or the experience of group flow can also be a profound bonding experience. For example, Chuck Israels reports "[y]ou play every beat in complete rhythmic unison... if it's working, it brings you very close... The relationship is very intimate" (quoted in Berliner, 1994, pp. 349–350).

We have outlined a model of joint action control that is built on a foundation of individual action control. Individual action control relies on inverse and forward models in order to transform goals first into control commands and then into predictions about how the action will unfold and to predict the sensory consequences of the action. We suggest that this system can be taken offline in order to transform a co-actor's goals into action and sensory predictions. Ultimately, the similarities between individual action and joint action, suggested by our framework, provides hints for understanding the phenomenology of joint action. While explanations of the phenomenology of group performance are still speculative, the model proposed here suggests that the mechanisms used to explain the phenomenology of acting alone may also be adept at explaining the phenomenology of acting together. However, further empirical investigations are needed in order to test the validity of this claim. Future work will also need to address in more detail how control commands are reconstructed during joint action, and how prediction mechanisms are fine-tuned for joint action and group flow. An understanding of the neural mechanisms involved in biological motion perception and body perception may be particularly suited to answering these questions. However, we believe that there is now a strong case to be made that a future understanding of joint action will be built on a foundation of individual action control.

References

Berliner PF (1994) Thinking in jazz: The infinite art of improvisation. University of Chicago Press, Chicago

Bratman ME (1993) Shared intention. Ethics 104(1):97–113

Buccino G, Binkofski F, Fink GR, Fadiga L, Fogassi L, Gallese V, Seitz RJ, Zilles K, Rizzolatti G, Freund HJ (2001) Action observation activates premotor and parietal areas in a somatotopic manner: an fMRI study. European Journal of Neuroscience 13(2):400–404

Calvo-Merino B, Glaser DE, Grézes J, Passingham RE, Haggard P (2005) Action observation and acquired motor skills: an fMRI study with expert dancers. Cerebral Cortex 15(8):1243–1249

Calvo-Merino B, Grézes J, Glaser DE, Passingham RE, Haggard P (2006) Seeing or doing? influence of visual and motor familiarity in action observation. Current Biology 16(19):1905–1910

Carruthers G (2010) A problem for wegner and colleagues' model of the sense of agency. Phenomenology and the Cognitive Sciences 9(3):341–357

Colling LJ, Thompson WF, Sutton J (2009) The influence of limb and joint information on action synchronization [Abstract]. In: Stevens C, Schubert E, Kruithof B, Buckley K, Fazio S (eds) Proceedings of the 2nd International Conference on Music Communication

Science (ICoMCS2), HCSNet, University of Western Sydney, Sydney

Colling LJ, Thompson WF, Sutton J (2010) Action synchronisation with biological motion. In: Christensen W, Schier E, Sutton J (eds) ASCS09: Proceedings of the 9th Conference of the Australasian Society for Cognitive Science, Macquarie Centre for Cognitive Science, Sydney, pp 49–56

Cross ES, Kraemer DJ, de C Hamilton AF, Kelley WM, Grafton ST (2009) Sensitivity of the action observation network to physical and observational learning. Cerebral Cortex 19(2):315–326

Csibra G (2008) Action mirroring and action understanding: an alternative account. In: Haggard P, Rossetti Y, Kawato M (eds) Sensorimotor foundations of higher cognition, Oxford University Press, New York, pp 435–459

Elsner B, Hommel B (2001) Effect anticipation and action control. Journal of Experimental Psychology: Human Perception and Performance 27(1):229–240

Flach R, Knoblich G, Prinz W (2003) Off-line authorship effects in action perception. Brain and Cognition 53(3):503–513

Freeman W (2000) A neurobiological role of music in social bonding. In: Wallin NL, Merker B, Brown S (eds) The origins of music, MIT Press, Cambridge MA, pp 411–424

Frith CD, Blakemore SJ, Wolpert DM (2000) Abnormalities in the awareness and control of action. Philosophical Transactions of the Royal Society B: Biological Sciences 355:1771–1788

Gallese V (2009) Motor abstraction: a neuroscientific account of how action goals and intentions are mapped and understood. Psychological Research 73(4):486–498

Golnaraghi M (2010) Automatic control systems, 9th edn. Wiley, Hoboken NJ

Grush R (1997) The architecture of representation. Philosophical Psychology 10(1):5–23

Grush R (2004) The emulation theory of representation: motor control, imagery, and perception. Behavioral and Brain Sciences 27(3):377–396

Hommel B (2003) Planning and representing intentional action. TheScientificWorldJOURNAL 3:593—608

Hommel B (2009) Action control according to TEC (theory of event coding). Psychological Research 73(4):512–526

Hommel B, Müsseler J, Aschersleben G, Prinz W (2001) The theory of event coding: A framework for perception and action planning. Behavioral and Brain Sciences 24(5):849–878

Hove MJ (2008) Shared circuits, shared time, and interpersonal synchrony. Behavioral and Brain Sciences 31(1):29–30

Hove MJ, Risen JL (2009) It's all in the timing: Interpersonal synchrony increases affiliation. Social Cognition 27(6):949–960

Jacob P (2008) What do mirror neurons contribute to human social cognition? Mind & Language 23(2):190–223

Keller PE (2008) Joint action in music performance. In: Morganti F, Carassa A, Riva G (eds) Enacting intersubjectivity: a cognitive and social perspective on the study of interactions, IOS Press, Amsterdam, pp 205–221

Keller PE, Knoblich G, Repp BH (2007) Pianists duet better when they play with themselves: On the possible role of action simulation in synchronization. Consciousness and Cognition 16(1):102–111

Kelso JA (1984) Phase transitions and critical behavior in human bimanual coordination. American Journal of Physiology: Regulatory, Integrative and Comparative Physiology 246(6):R1000–1004

Knoblich G (2008) Bodily and motor contributions to action perception. In: Klatzky R, MacWhinney B, Behrmann M (eds) Embodiment, ego-space, and action, Psychology Press, New York, pp 45—78

Knoblich G, Flach R (2001) Predicting the effects of actions: interactions of perception and action. Psychological Science 12(6):467–472

Knoblich G, Jordan JS (2003) Action coordination in groups and individuals: Learning anticipatory control. Journal of Experimental Psychology: Learning, Memory, and Cognition 29(5):1006–1016

Knoblich G, Sebanz N (2008) Evolving intentions for social interaction: from entrainment to joint action. Philosophical Transactions of the Royal Society B: Biological Sciences 363(1499):2021–2031

Knoblich G, Seigerschmidt E, Flach R, Prinz W (2002) Authorship effects in the prediction of handwriting strokes: Evidence for action simulation during action perception. The Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology 55(3):1027–1046

Mechsner F, Kerzel D, Knoblich G, Prinz W (2001) Perceptual basis of bimanual coordination. Nature 414(6859):69–73

Miall RC, Reckess GZ (2002) The cerebellum and the timing of coordinated eye and hand tracking. Brain and Cognition 48(1):212–226

Miall RC, Wolpert DM (1996) Forward models for physiological motor control. Neural Networks 9(8):1265–1279

Penfield W (1954) Mechanisms of voluntary movement. Brain 77(1):1–17

Rizzolatti G, Craighero L (2004) The mirror-neuron system. Annual Review of Neuroscience 27(1):169–192

Sato A (2009) Both motor prediction and conceptual congruency between preview and action-effect contribute to explicit judgment of agency. Cognition 110(1):74–83

Sato A, Yasuda A (2005) Illusion of sense of self-agency: discrepancy between the predicted and actual sensory consequences of actions modulates the sense of self-agency, but not the sense of self-ownership. Cognition 94(3):241–255

Sawyer RK (2006) Group creativity: musical performance and collaboration. Psychology of Music 34(2):148–165

Schmidt RC, Carello C, Turvey MT (1990) Phase transitions and critical fluctuations in the visual coordination of rhythmic movements between people. Journal of Experimental Psychology: Human Perception and Performance 16(2):227–247

Schubotz RI (2007) Prediction of external events with our motor system: towards a new framework. Trends in Cognitive Sciences 11(5):211–218

Searle J (1990) Collective intentions and actions. In: Cohen PR, Morgan J, Pollack ME (eds) Intentions in communication, MIT Press, Cambridge MA, pp 401–415

Sebanz N, Knoblich G, Prinz W (2003) Representing others' actions: just like one's own? Cognition 88(3):B11–B21

Sebanz N, Knoblich G, Prinz W (2005) How two share a task: corepresenting stimulus-response mappings. Journal of Experimental Psychology: Human Perception and Performance 31(6):1234–1246

Sebanz N, Bekkering H, Knoblich G (2006) Joint action: bodies and minds moving together. Trends in Cognitive Sciences 10(2):70–76

Simon JR (1969) Reactions toward the source of stimulation. Journal of Experimental Psychology 81(1):174–176

Spivey MJ (2007) The continuity of mind. Oxford University Press, New York

Synofzik M, Vosgerau G, Newen A (2009) Reply to carruthers. Consciousness and Cognition 18(2):521–523

Tsai C, Kuo W, Hung DL, Tzeng OJL (2008) Action co-representation is tuned to other humans. Journal of Cognitive Neuroscience 20(11):2015–2024

Wegner DM (2002) The illusion of conscious will. MIT Press, Cambridge MA

Wegner DM, Sparrow B, Winerman L (2004) Vicarious agency: experiencing control over the movements of others. Journal of Personality 86(6):838–848

Wilson M, Knoblich G (2005) The case for motor involvement in perceiving conspecifics. Psychological Bulletin 131(3):460–473

Witney AG (2004) Internal models for bi-manual tasks. Human Movement Science 23(5):747–770

Wolfensteller U (2009) Juggling with the brain – thought and action in the human motor system. In: Raab M, Johnson JG, Heekeren HR (eds) Mind and motion: the bidirectional link between thought and action, Elsevier, Amsterdam, pp 289–301

Wolpert DM (1997) Computational approaches to motor control. Trends in Cognitive Sciences 1(6):209–216

Wolpert DM, Kawato M (1998) Multiple paired forward and inverse models for motor control. Neural Networks 11(7-8):1317–1329