

Mechanisms of copying, social learning, and imitation in animals

Author links open overlay panel

Thomas R. Zentall

[Show more](#)

Add to Mendeley

Share

Cite

<https://doi.org/10.1016/j.imot.2022.101844>

[Get rights and content](#)

Abstract

When an animal matches the behavior of another animal of the same species it likely has evolutionary value because that behavior often avoids the negative outcomes that may accompany trial and error learning. Psychologists have been interested in the mechanisms that might be responsible for the copying behavior. Flocking behavior, an example of behavioral contagion, for example, is likely to be under genetic control. Alternatively, if an animal needs to acquire a new behavior to obtain a reinforcer, the mere presence of another animal may influence motivation, and perhaps also performance. Also, when observing the performance of a response, attention to the manipulated object may be sufficient to facilitate learning. Furthermore, a distinction can be made between imitating the movements of the demonstrator and learning the relations between the manipulated object and the reinforcer (learned affordance). Research using the two-action procedure controls for all of these alternative mechanisms and shows what appears to be evidence for a more cognitive form of imitation. Although human children show a remarkable propensity to imitate there is evidence that many other species (not only other apes, other mammals, and bird species) show evidence that they too will imitate the behavior of others, under appropriate conditions. These results suggest that imitation, thought to be a conceptually cognitive behavior, is widely distributed among many animal species.

Introduction

When one organism matches or copies the behavior of another, because of the widespread finding of matching behavior across species it has attracted the attention of ethologists, behavior ecologists, anthropologists, and psychologists (Heyes, 1994; Heyes, & Galef Jr., 1996; Huber et al., 2009; Hoppitt & Laland, 2008; Galef & Giraldeau, 2001; Leadbeater & Chittka, 2009; Piaget, 1962; Schuster, Wohl, Griebisch, & Klostermeier, 2006; Shettleworth, 1998; Whiten & Ham, 1992; Whiten, Horner, & de Waal, 2005; Wilkinson, Kuenstner, Mueller, & Huber, 2014). However, the interests of researchers trained with a biological orientation and those who come from a more psychological orientation have approached their interest in behavioral similarity by animals somewhat differently. Researchers with a biological orientation tend to be interested in the potential value to the organism of behaving like others. Researchers with a more psychological orientation tend to be interested in the implications that such behavior matching has for the intelligence of the organism.

Section snippets

Biological function versus psychological mechanism

Biologists focus primarily on behavior that is genetically predisposed. Such behavior is often referred to as

species-typical behavior. Its advantage is that such behavior is relatively certain. For birds, courting, mating, nest building, and caring for young do not have to be learned. They may be triggered by maturation or by the environment. There is a potential cost, however, to such species-typical behaviors. They are relatively inflexible. Should the environment change in a way that would

Social influence

Some copying behaviors may not actually involve learning. They may be influenced by the presence of others, but they do not have to be learned. Birds that live in flocks typically leave a foraging patch together. They are copying the behavior of the other birds but that would qualify as species-typical defensive behavior – a genetically predisposed behavior - because leaving together would likely be safer than staying. Similarly, some bird species act in concert when they attack or “mob” a

Motivational effects

Imagine that one is interested in asking if an animal can learn to perform a behavior from watching another animal engaged in the behavior. If that behavior is typically followed by a reward, the observing animal might be able to learn the behavior on its own (by trial and error). Thus, one would need to ask if learning by watching another animal perform the behavior is faster than learning by trial and error. But if one is interested in understanding the underlying mechanism, it may be that

Perceptual factors

The implication of learning by imitation is that the observer understands what the demonstrator is doing and is motivated to copy that behavior. It is possible, however, that the way in which the demonstrator interacts with the apparatus affects the observer's attention, and the change in attention may facilitate acquisition of the target response. For example, an observer sees a demonstrator pressing a lever. The movement of the lever attracts the attention of the observer to the location of

Bird song

A special case of matching behavior by animals is the acquisition of bird song (Hinde, 1969; Marler, 1970; Nottebohm, 1970; Thorpe, 1961; see also vocal mimicry, e.g., Pepperberg, 1986; Thorpe, 1967). Although for many species of songbird the development of species-typical song is regulated to a large extent by maturation and the seasonally fluctuating release of hormones, for some species, regional variations in the song appear to depend on the bird's early experience with conspecifics (

Imitation

Piaget (1962) suggested that imitation must involve what is now referred to as having a "theory of mind" (the ability to attribute mental states to ourselves and others, Premack & Woodruff, 1978). Following Piaget's approach, Tomasello (1990) proposed that in true imitation, an imitating observer understands that the demonstrator used particular actions with a particular goal in mind. This may be a useful definition when testing adult humans for their ability to imitate but it would be

Transparent versus opaque imitation

Considering the different two-action procedures mentioned above, a distinction can be made between transparent and opaque imitation. If one imitates someone who presses a button with their finger, the button pushing might be considered transparent because the image of the demonstrator's finger pressing the button and the image of one's own finger pressing the button could be seen as similar. Thus, stimulus matching could be proposed to account for such imitation (stimulus matching is a concept

Over-imitation: are chimpanzees smarter than children?

A striking feature of children's imitation is the extent to which they are prepared to imitate actions that

appear to be unrelated to the apparent outcome of the task. This phenomenon, known as over-imitation, was first studied by Horner and Whiten (2005) and later labeled by Lyons, Young, and Keil (2007). In the Horner and Whiten experiment, 4-year-old children were shown a puzzle-box with a hole at the top covered by a bolt that could be pushed to the side to allow a stick to be inserted into

Imitation as a measure of intelligence

Imitation is sometimes taken as a measure of intelligence. This may be because Piaget (1962) interpreted certain kinds of imitation (often referred to as opaque imitation) as requiring the ability to take the perspective of another (theory of mind) and perspective taking is often viewed as a measure of intelligence (Harwood & Farrar, 2006). Relative to other species, humans appear to have a greater tendency to imitate (see Meltzoff, 1988).

The finding of imitation of the kind proposed by Piaget

The question of enculturation

Researchers have long been interested in the study of imitation in apes because they are our closest living relatives, and potentially they provide an important means of separating the effects of our culture from genetically predisposed behavior. Among primate researchers, a distinction has been made between so called enculturated (non-human) apes that have had extensive human interactions and non-enculturated apes that have been raised by their mothers (or peers). On the one hand, when

Conclusions

Our interest in animal imitation started with the question, can animals learn by copying the behavior of others? The simple answer is yes. This would appear to be a straight-forward question that should be relatively easy to answer but understanding the mechanisms involved in copying behavior is not so easy to determine. If we try to order the ways in which behavior copying can take place, we might start with predisposed behavior that an animal frequently often does on its own, such as eating

Highlights

- Understanding how the human brain translates visual information into skilled motor performance has been assisted and constrained by the discovery of mirror neurons.
Emerging evidence highlights how observational motor learning involves a far more diffuse network of brain regions and cognitive processes, which are shaped by the context and complexity of the motor task to be learned.
A greater emphasis on combining functional decomposition and functional integration

approaches should facilitate paradigms and discoveries that move us closer toward understanding how we learn from watching others in complex, real-world scenarios.

The mirror neuron system has dominated understanding of observational learning from a cognitive neuroscience perspective. Our review highlights the value of observational learning frameworks that integrate a more diverse and distributed set of cognitive and brain systems, including those implicated in sensorimotor transformations, as well as in more general processes such as executive control, reward, and social cognition. We argue that understanding how observational learning occurs in the real world will require neuroscientific frameworks that consider how visuomotor processes interface with more general aspects of cognition, as well as how learning context and action complexity shape mechanisms supporting learning from watching others.

Keywords

observational learning
human mirror neuron system
motor learning
motor system
social cognition
reward
real-world neuroscience

Learning from Watching Others Is a Fundamental Human Skill

Whether learning to dance Gangnam Style, open a bottle of champagne, or tie shoelaces, humans learn a great deal by simply watching others [1.,2.]. Learning by

observation can have many benefits over physical practice without observation. This is especially true in dangerous or novel environments where poor initial performance can be costly, such as learning to drop in on a skateboard ramp without first watching someone else do it successfully. Equally, learning by watching others is beneficial in social situations where active participation is not possible, such as watching your supervisor give critical feedback to a colleague. As such, observational learning has been of considerable interest to experts from a variety of disciplines, including social and developmental psychology, sport and exercise science, comparative biology and robotics, where researchers are examining how we learn from watching others across a range of behaviors, from simple motor movements to complex social interactions [1.,3., 4., 5.].

Contemporary empirical study into how we learn from watching others originated, to a large extent, in the **social learning** (see [Glossary](#)) research of Albert Bandura and colleagues starting in the 1960s [3.,6., 7., 8.]. Much of this early work focused on the role of an observer's motivation to learn and social factors relevant to the model. Over the past several decades, however, empirical interest in examining observational learning has come increasingly from the action domain, with considerable attention from cognitive neuroscientists who have begun to explore and characterize the systems within the human brain that translate visual signals into motor output [2.,9., 10., 11., 12.]. While tremendous progress has been made in identifying the neural correlates

supporting observational learning of others' actions, our understanding of how we learn from watching the actions of others remains in its infancy. The foundational structure of relevant cognitive and brain systems remains largely unknown and many key questions still need to be addressed. By synthesizing initial findings from several relevant subdisciplines of human neuroscience and placing them within a wider context of observational learning across species and agents, we aim to provide an overview of the progress that has been made towards understanding brain-based mechanisms that support observational learning. We also outline important next steps that, if taken, could generate a more complete understanding of this ubiquitous and vital capacity.

The Value of a Holistic Approach to Studying Observational Learning

Much like human neuroscience research in general, observational learning research within cognitive neuroscience has tended to focus on understanding the role of a restricted set of brain areas, such as the human **mirror neuron system**. This focus has undoubtedly contributed to our knowledge about how we learn from watching others, but it has also come at the expense of understanding and appreciating the roles played by a wider set of systems, as well as how such systems interact. These additional systems include the motor system more generally, as well as regions associated with semantic processing, attention, and memory. What is currently lacking, therefore, are frameworks describing how the brain enables

information to be incorporated across distributed neural networks, some of which may be particularly relevant for action learning and others that may support a host of more general cognitive processes. Human social life is simply too complex to understand by focusing exclusively on the operation of subsystems in isolation. For example, if a clear understanding of motor development in children or expertise in dancers and athletes is to be developed, it is vital to study more than a restricted set of motor skills suitable for laboratory study (such as key pressing or moving a manipulandum), where the focus is on a subset of component systems. Likewise, if socially harmonious robotic agents that learn from and adapt to human actions are to be developed, this will require a fuller understanding not only of how skills are acquired by observation but also how they fit with other states and processes such as motivation, goals, and intentions. Without incorporating these perspectives, one is likely to fall short of a holistic understanding of the processes and systems that support observational learning in the human brain, how these develop and change across the lifespan, and how these might be modeled in artificial agents.

Given the varieties of observational learning that have been studied to date ([Box 1](#)), it is important to establish the scope of this review. We focus on a specific type of observational learning – namely observational motor learning, where action observation leads to an enduring refinement to motor performance. We have organized the review into three main parts. First, owing to its dominance, we review research into the role of the

human mirror neuron system in observational learning. Next, we review evidence implicating motor processes extending beyond the mirror system in observational learning. We outline how research on sequence learning and **motor adaptation** in particular informs our understanding of the neural substrates underlying observational learning. Finally, we review evidence highlighting the roles played by non-motor systems in observational learning, such as those associated with cognitive control and reward. We also detail how systems or frameworks that might seem less relevant to understanding observational learning, such as semantic and memory systems, as well as 'real-life' neuroscience approaches, also hold considerable potential for making progress in this endeavor.

Box 1

We conclude by articulating what is known and what remains to be discovered regarding the cognitive neuroscience of learning from watching others. If the ultimate goal is to understand observational learning as it occurs in daily life, there is a need to consider a distributed and diverse set of interacting neurocognitive systems both within and beyond the human mirror neuron system. It is also imperative to combine evidence gathered from controlled laboratory conditions with work that puts observational learning to the test in complex and rich real-world contexts to better understand how to support and bolster this vital skill.

The Dominance of the Human Mirror Neuron System in Cognitive Neuroscience Investigations of Observational Learning

Cognitive neuroscientific investigations into

observational learning have focused extensively on the role played by the human mirror neuron system (Box 2). Since the discovery of **mirror neurons** within premotor and, later, inferior parietal cortices of the non-human primate brain [13., 14., 15.], researchers have sought to determine whether similar neural processes exist within the human brain. To do so, researchers have used a variety of noninvasive neuroimaging and neurostimulation measures to probe links between action and perception [16., 17., 18., 19., 20., 21., 22.]. This work revealed compelling evidence that the human mirror neuron system is engaged not only during action perception and performance but is also sensitive to learning and experience, with more familiar actions leading to more robust engagement [23., 24., 25., 26.]. Furthermore, research examining learning guitar chords [27., 28., 29.], dance movements [9.,30.], and assembling or using novel objects [10.,31., 32., 33.] reported converging evidence that frontoparietal brain regions are modulated when observation of others' actions resulted in subsequent learning. This evidence has led to a general understanding within cognitive neuroscience that action learning by observation relies, at least in part, on functions performed by the mirror neuron system.

Box 2

A role for the human mirror neuron system in observational learning makes intuitive sense, given that the neurons found in these cortical regions have well-documented sensory and motor properties. At the same time, we argue that an outsized focus on the human mirror neuron system has produced an oversimplified

account of the neural underpinnings of observational learning. Indeed, as already foreshadowed in an early meta-analysis [16.], as well as more recent empirical work (e.g., [28., 34., 35., 36.]), several lines of evidence document brain regions and systems extending beyond the mirror neuron system that contribute to this fundamental capacity. In the following sections we examine the role(s) played by some of these systems. Our starting point is that the mirror neuron system is unlikely to be the 'center' of any complex cognitive process. Instead, we argue that the mirror neuron system is likely to perform a range of different subprocesses that contribute to many broader cognitive functions [37.]. Consequently, we offer a deflationary account of the mirror neuron system in observational learning, which assigns a more limited role to this system than is typically proposed. Instead of the mirror system being the star (or even solo) performer in observational learning, we emphasize how this type of learning is far more likely to rely on coordinated interplay between the mirror neuron system and other brain networks.

The Role of an Extended Motor Network in Observational Learning

In this section we review recent work on observational motor learning highlighting roles for cognitive and brain systems extending beyond the core brain regions traditionally associated with the mirror neuron system. We focus on two broad types of task: sequencing learning and motor adaptation [38.] (Box 3). Studying sequence learning holds value because, in real life,

much of skilled human behavior follows a sequential structure (e.g., learning to speak, dance, play the piano, ride a bike, and drive a car). Moreover, like many other kinds of human behavior, sequence learning can be achieved via physical practice or by observing others [2.]. Observational motor sequence learning is often studied via the **serial reaction time (SRT)** task (Box 3), and shares two important behavioral characteristics with motor learning achieved primarily through physical practice. First, increasing the number of observation trials improves learning in much the same manner as increasing the number of physical practice trials does [39.]. Second, contextual interference effects occur in observational settings [40.] exactly as they do in physical practice settings [41., 42., 43.]. Therefore, learning via physical and observational practice appears to rely on partially shared cognitive foundations [1., 2., 44., 45., 46.].

Box 3

To date, a small but growing number of functional magnetic resonance imaging (fMRI) studies have investigated the neural substrates underpinning observational motor sequence learning. Some researchers have used tasks involving learning the correct sequence of movements to assemble an object [31.]. When watching the sequence of actions with the intention to learn compared with a perceptual control task, widespread engagement emerged across inferior and superior frontal and parietal cortices, presupplementary motor area (pre-SMA), the cerebellum, basal ganglia, and the hippocampus. Likewise, others have used dance and complex knot-

tying paradigms to study observational sequence learning [9.,10.]. These studies show that when people watch dance sequences they learned via observation, or view pictures of knots that they observed another person learn to tie, dorsal premotor cortex shows stronger engagement compared with observing untrained stimuli. In addition, several fMRI studies have used the SRT task to probe the neural correlates of observational learning [11.,28.,36.,47.]. Across these studies, researchers consistently report sensitivity within dorsal premotor cortex and superior parietal lobule, as well as ventral premotor cortex, when learning SRT tasks via observation. These results show that observational motor sequence learning is not restricted to engagement of inferior frontal and parietal brain regions but extends to a broader motor network that includes dorsal premotor, superior parietal, and cerebellar regions (Figure 1).



Figure 1 Human Brain Regions Implicated in Observational Motor Learning.

[Show full caption](#)

[View Large Image](#)[Download Hi-res image](#)[Download \(PPT\)](#)

Brain imaging research on the motor system also highlights how responses can increase or decrease following learning, depending on a range of factors, which can complicate the interpretation of learning-related neural activity changes. Activation decreases following both physical and observational sequence learning have been reported in inferior and superior portions of the anterior parietal lobe, as well as in ventral and dorsal premotor cortex [11.,28.,36.,48.]. This is consistent with a general pattern reported in the literature of activity decreases across a number of areas, including primary motor cortex (M1), pre-SMA,

and dorsolateral prefrontal cortex (DLPFC), during the initial fast stages of learning a sequential motor task (reviewed in [49.]). Importantly, later stages of learning are associated with activation increases in M1, SMA, ventral premotor cortex (PMv), and primary somatosensory cortex (S1) ([50.,51.], reviewed in [49.]; *cf* [52.] for a contrasting view). Whether observational motor sequence learning occurring over even longer time-periods (months and years rather than weeks and days) results in similar increases in activation in areas that exhibited short-term learning-related decreases remains unexplored.

Whereas motor sequence learning supports the acquisition of complex motor skills, motor adaptation supports the maintenance of consistent performance in response to changes in the body or external environment (Box 3). According to the dominant theoretical framework, this type of learning involves the recalibration of internal models (i.e., representations of body–environment interactions) used to support feedforward and feedback motor control [53., 54., 55., 56.]). Observational motor adaptation describes how individuals can learn to respond to novel sensorimotor perturbations by observing someone else adapt their movements to those same perturbations. Observation has been shown to facilitate visuomotor adaptation [46., 47., 48.], force-field adaptation [46.,57.,58.], and force estimation [59.,60.]. For example, in experiments probing the effects of observation on subsequent visuomotor adaptation [61., 62., 63.], participants observe someone else learning to reach using rotated visual feedback. When

observers are subsequently exposed to the same visuomotor rotation, their reaches become more accurate during initial exposure compared with those who had no previous observational experience. Despite their similarities, some important differences have also been noted between observational and first-hand, physical visuomotor learning. First, although passive observers exhibit improved learning when they encounter the same visuomotor perturbation they previously observed, they do not show after-effects (reaching errors in the normal unperturbed environment following exposure to perturbed conditions) – a hallmark of implicit motor adaptation [61.,62.]. Second, observers do not suffer interference costs when consecutively learning two opposing visuomotor perturbations, another defining characteristic of implicit motor adaptation [64.,65.]. These results imply that the benefits of observation may reflect learning explicit aiming strategies [66.] as opposed to using motor errors on a trial-by-trial basis to incrementally update an internal model [67., 68., 69.]. More work is needed in this area to identify the underlying mechanisms.

Examining the neural mechanisms underlying observational learning using motor adaptation tasks is a growing area of research, and novel insights are coming from studies employing neurostimulation and neuroimaging methods. In one study, **repetitive transcranial magnetic stimulation (rTMS)** was applied over M1 after participants observed someone learning to reach in a force field [70.]. Even though observation improved subsequent learning in a control group, participants in the rTMS group showed degraded

adaptation to the same force field they observed at levels comparable with another group of controls who received no observational experience at all. Similarly, single-pulse TMS over M1 has been shown to induce larger increases in corticospinal excitability when observing motor learning compared with observing similar movements that involved no learning [71.]. In addition to M1, emerging evidence supports a tight interplay between sensory and motor areas during motor learning by observing [72.,73.]. For example, a role for S1 in observational learning in motor adaptation tasks has been demonstrated using median nerve stimulation to disrupt the function of S1 [35.]. Stimulation delivered to the same arm used in the observed action impaired subsequent learning, whereas stimulation of the opposite arm did not. Together, results from neurostimulation studies indicate a crucial role for M1 and S1 in learning by observation.

Neural reorganization following observational learning has also been studied using fMRI combined with motor adaptation tasks. For example, in a study where participants observed movement errors performed when adapting to novel force fields [57.], the authors found activation of portions of the posterior parietal cortex, dorsal premotor cortex, and cerebellum. In another fMRI study [58.] the authors investigated changes in resting-state functional connectivity after participants observed others complete a force-field learning task and identified a network consisting of the middle temporal visual area (V5/MT), S1, M1, and cerebellum. Importantly, connectivity changes correlated with the amount of learning gained through observation. Subsequent work

showed that individual differences in pretraining resting-state functional connectivity in sensorimotor brain areas could predict the amount of learning by observation [74.]. The studies reviewed in this section demonstrate that, even if one only considers standard motor learning paradigms, a widespread and distributed neural network is involved when learning from watching other people's actions. Accordingly, even simple, pared-down tasks involving relatively simple actions (e.g., planar reaching, button pressing, etc.) recruit a complex network of brain areas that extends beyond the mirror neuron system (Figure 1). These considerations underscore the value to be gained from broadening the focus of observational learning research.

The Role of Non-motor Systems and Frameworks

Brain circuits beyond the motor system have received less attention in the context of observational motor learning. This makes sense, given that the motor system is the obvious place to start when attempting to understand the brain mechanisms that underpin motor skill learning (observational or otherwise). However, as outlined earlier, to scientifically understand complex processes or systems, one must consider how the component subprocesses or subsystems operate both in isolation and in combination [75.,76.]. In addition, one of the earliest and arguably most influential psychological theories of observational learning to date emphasized that a wide range of non-motor factors, such as an individual's motivation to learn, also make important contributions to observational learning [3.]. In this section we review research implicating a role for

non-motor systems in observational learning, including those associated with cognitive control and reward. We also outline how considerable value should be gained from exploring observational learning from different perspectives, such as those associated with semantic or memory systems and so-called 'real-life' neuroscience approaches. Finally, we suggest that observational learning is a prime case where 'neuroscience needs behavior' [77.]. In other words, our understanding of how brain systems support observational learning will be fundamentally limited until neuroscientific methods can be connected to real-world behavior in meaningful ways. The first two perspectives that we consider have already been studied in observational learning contexts, albeit to a relatively small degree. The first concerns general cognitive control processes, which are reliably associated with bilateral DLPFC [78.,79.]. Cognitive control processes regulate other mental processes and guide attention onto relevant features of the environment through processes such as alerting, orienting, filtering, and inhibition [80.]. The versatility of this frontoparietal brain network to operate across a range of tasks and contexts has led to it being labeled the 'multiple demand network' [79.]. In the context of observational learning, depending on the type of learning task or aims of the individual, cognitive control is necessary to direct attention to features that are most relevant in a given context. The efficiency and effectiveness of such control processes also likely contributes to the observational learning rate. Indeed, evidence has shown that DLPFC correlates with performance gains following with observational practice

of a guitar chord learning task [28.]. Furthermore, transcranial magnetic stimulation to DLPFC makes behavior more error-prone when performing motor sequences previously learned via observation, but not through physical practice [81.]. As such, visual and motor processes do not appear to operate in isolation, but are instead modulated by other task demands, which are likely to reflect a range of other supervisory and control processes.

The second system concerns reward and value-based learning [34., 134., 135.]. For example, some authors have made the case that basic principles of learning, such as those grounded in Pavlovian and instrumental learning, must be considered together with advances in social cognition to better understand observational learning [34.]. Across a range of non-human species, they show that social learning is partly indexed by similar reward centers in the brain that coordinate learning by direct experience. Similarly, in humans, single-neuron recordings have identified neurons in the amygdala that track the expected value of any given trial in a gambling task based on one's own experience, as well as knowledge gained from watching others [82.]. Accordingly, it has been suggested that learning from others is partly mediated by a general-purpose value system in combination with more socially specific processes such as mental state reasoning [34.]. Future research building on these value processes and extending them more deeply into the domain of human neuroscience would be particularly valuable, especially given that a strong foundation exists in the non-human literature to guide predictions.

In addition to cognitive control and reward processes, it is also important to consider perspectives that have not yet been widely considered in cognitive neuroscience models of observational learning. Consider semantic cognition research, for instance, which investigates how meaning is extracted from interactions with the environment (e.g., [83,84.]). This perspective has recently been applied to social cognition more generally, under the view that observing other people is nothing special and is merely one more way to learn about the world and people in it [85.]. Another example is research on memory systems. Recent theoretical work makes the case for how research on impression formation, which is a form of learning about people and their trait characteristics, could benefit from a much greater consideration of developments in the neuroscience of memory [86.]. The proposal outlines how distinct types of memory, such as episodic, associative, and instrumental, are associated with distinct but interacting brain circuits, and such knowledge is likely to inform how social aspects of knowledge are stored and retrieved. We would argue that the same lessons apply to research on observational learning. Specifically, important questions to explore include which types of memory systems are involved when we learn from watching others, how the learning or social context shapes their engagement, and how they interface with social information processing systems more generally. One major advantage of considering semantic and memory frameworks in more depth is that they are based on decades of research across multiple species, thus providing a rich foundation to build upon.

Finally, we offer two further proposals. First, to accelerate progress toward understanding how we learn from others in daily life, it will be important to consider the fast-changing developments that are emerging in 'real-life' neuroscience [87., 88., 89., 136.]. For instance, how do individuals learn from other people when interacting in groups? How do complex contextual signals, such as person identity, type of action, and setting (e.g., exam conditions, combat, surgery, or leisure time) interfere with or facilitate observational motor learning? Although some laboratory work has taken steps toward understanding action learning among dyads present in a laboratory (as opposed to watching video-based representations of to-be-learned actions; e.g., [10.]), current understanding of how learning unfolds in complex real-world contexts where multiple models and environmental pressures and opportunities are present remains limited at the behavioral (let alone neural) level. This progress will require the development of approaches that work both within and beyond the confines of the research laboratory. Portable neuroscience technologies, such as **functional near-infrared spectroscopy (fNIRS)**, may offer a suitable methodological approach to bridge this gap between the research laboratory and the real world [90., 91., 92.]. Embracing more 'real-life' neuroscience approaches will likely further reinforce the need to take an expanded view of the cognitive and brain systems involved in observational learning. A fuller understanding of how we learn from others in daily life stands to further advance several related areas of inquiry, including motor development and learning

across the lifespan, how expertise is established, and how best to program robots to accomplish joint actions with human users.

Second, much like neuroscience research more generally, the role of behavior in revealing principles of observational learning should not be undervalued [77.]. For instance, in a laboratory sequence-learning task, behavioral research has shown that individual differences in dimensions of personality, working memory, and intelligence play a minimal role in observational learning [93.]. Such findings may help to constrain expectations about the underlying brain mechanisms, especially those that may be more or less variable across individuals. That said, while behavioral experiments are certainly crucial for understanding the phenomenon of observational learning, we advocate a pluralistic approach in which investigations across multiple levels of description can play complementary roles in elucidating the links between brain and behavior [77.,94., 95., 96.]. It is becoming increasingly clear that both behavioral and neural data can help to adjudicate between competing mechanistic models and place useful constraints on mechanism discovery in the human brain [97.,98.].

Overall, our central argument in this section is that, for more substantial progress to be made, we encourage the field to fully embrace research perspectives from separate but related domains. Progress will be facilitated by a greater focus on generalized processes and other general frameworks, which are not necessarily or strictly tied to visual or motor processing *per se* [99., 100., 101., 102.].

Concluding Remarks

Our review of the current state of knowledge on the cognitive neuroscience of observational learning raises several considerations for future research (see Outstanding Questions). Studying component processes in isolation and under controlled laboratory conditions has made valuable contributions to understanding cognitive and brain mechanisms of observational learning and will continue to do so. At the same time, an overemphasis on one system or process can give the illusion that complex and multifaceted processes can be reduced to the workings of a few solitary cognitive and brain subsystems [103.,137.].

Future observational learning research may benefit from taking a different starting point, which is predicated on an updated and richer set of assumptions (Figure 2, Key Figure). We suggest that it will be valuable to start from a position that expects a large number of 'many-to-many' mappings between cognitive processes and brain circuits. To better understand how we learn from others in the real world, the contributions of a multitude of factors (social, semantic, motivational, contextual) that may up- or downregulate basic observational learning processes must be considered. A consequence is that, to understand basic motor processes, one must understand how they operate in tandem with a diverse set of social goals that those actions sometimes serve. This is not to say that all research in this domain should change tack to focus exclusively on interactions between and integration across networks. Instead, going forward, a combination of functional segregation and

integration perspectives will be necessary. One particular promising approach for accelerating progress toward a deeper understanding of observational motor learning is the type of neural network modeling that is gaining momentum in human motor learning research (e.g., [104., 105., 106.]). A combination of methodological and theoretical expertise from a wider range of cognitive neuroscience subdisciplines should bring this field closer to characterizing how the human brain can translate information about other people's actions from observation into sound tooth-brushing technique, the intricate hand and foot moves of Bharatanatyam dance, or the ability to successfully perform minimally invasive surgery.

Action observation versus motor imagery in learning a complex motor task: A short review of literature and a kinematics study

Author links open overlay panel

R. Gatti ^a, A. Tettamanti ^a, P.M. Gough ^b, E. Riboldi ^a, L. Marinoni ^a, G. Buccino ^{c d}

[Show more](#)

Add to Mendeley

Share

Cite

<https://doi.org/10.1016/j.neulet.2012.11.039>

[Get rights and content](#)

Abstract

Both motor imagery and action observation have been shown to play a role in learning or re-learning complex motor tasks. According to a well accepted view they share a common neurophysiological basis in the mirror neuron system. Neurons within this system discharge when individuals perform a specific action and when they look at another individual performing the same or a motorically related action. In the present paper, after a short review of literature on the role of action observation and motor imagery in motor learning, we report the results of a kinematics study where we directly compared motor imagery and action observation in learning a novel complex motor task. This involved movement of the right hand and foot in the same angular direction (in-phase movement), while at the same time moving the left hand and foot in an opposite angular direction (anti-phase movement), all at a frequency of 1 Hz. Motor learning was assessed through kinematics recording of wrists and ankles. The results showed that action

observation is better than motor imagery as a strategy for learning a novel complex motor task, at least in the fast early phase of motor learning. We forward that these results may have important implications in educational activities, sport training and neurorehabilitation.

Highlights

► The mirror neuron system sub-serves both motor imagery and action observation. ► Both motor imagery and action observation play a role in motor learning. ► In this study we compared these strategies in learning a novel complex motor task. ► Action observation revealed better than motor imagery as a motor learning strategy. ► This is relevant in educational activities, sport training and neurorehabilitation.

Introduction

This introduction includes two main parts: the first one is intended as a short review of literature on motor learning. Within the theoretical framework depicted in the first part, the second one outlines the background and aim of a kinematics study where the efficacy of two common strategies used to learn a novel motor task, motor imagery and action observation, has been directly compared.

Motor learning is the process through which we come to perform actions effortlessly after practice and interactions with the environment. By means of motor learning we continuously extend our motor repertoire.

Motor learning involves a number of interacting components [69]: processing and collecting sensory information relevant to action in an effective and efficient manner, applying a series of decision making strategies aimed at defining which movements to perform and which order to follow while performing them, activating control processes during motor performance including a feed-forward control, a reactive control and a biomechanical control.

Two experimental paradigms are usually used to study the mechanisms leading to learning new motor skills: motor sequence learning, assessing the incremental acquisition of movements in a specific behavior, and adaptation model, through which the capacity to compensate for changes in the environment is assessed [16].

For both paradigms distinct learning phases can be distinguished: a fast phase, in which improvement occurs within the very first training session; a consolidation phase, in which an improvement of performance occurs after at least 6 h from the first practice session; a slow phase in which further gains can be achieved across several training sessions; an automatic stage in which the motor task is performed automatically with poor cognitive demand; a retention state in which the motor performance can be executed in the absence of any practice after a long delay [16], [31].

Two frameworks have been proposed so far describing the neural mechanisms underlying motor learning [13], [16], [35]. In one model [13] two parallel circuits operate in learning spatial and motor features of motor sequences. Learning spatial coordinates would be supported by a frontoparietal associative striatum-cerebellar circuit, while learning motor coordinates would be supported by a primary sensorimotor striatum-cerebellar circuit. In another model [16] it was proposed that during fast learning a cortico-striato-thalamo-cortical loop and a cortico-cerebello-thalamo-cortical loop are both recruited operating in parallel. Interactions between the two systems are believed to be crucial for establishing the motor routines necessary for learning new motor skills. When consolidation has occurred the motor representation of the learned motor skill is believed to be distributed both in the cortico-striatal loop and in the cortico-cerebellar loop, with a preeminent role of the cortico-striatal loop in the case of motor sequence learning and a preeminent role of cortico-cerebellar loop in the case of adaption learning. Little attention has been paid so far by current literature on motor learning to the best way to start a motor learning process and/or to the best way to gain a solid performance in a number of situations.

Two well assessed strategies for motor learning are action observation and motor imagery. Both action observation and motor imagery are believed to share the same neural mechanisms in the mirror neurons [38]. The discovery of mirror neurons [25], [54], first identified in monkey premotor area F5, has revealed

a mechanism in the brain which allows one to match an observed action with its motor counterpart in the observer's brain. In fact these neurons discharge when an animal performs an object-directed action and when the same, or a motorically related action, is performed by another individual.

Neurons endowed with mirror properties were found in both frontal and parietal areas, thus constituting the mirror neuron system (MNS) [19]. A system similar to the one described in the monkey has also been found in humans. A number of neurophysiological studies have shown that the observation of object-directed, as well as non object-directed actions, modulates the activity of those motor areas, normally involved in the actual execution of the observed actions, while brain imaging experiments have shown that during the observation of both object-directed and non object-directed actions different sectors of the premotor and parietal cortex are recruited [2], [10], [15], [21], [24], [26], [30], [33], [43], [62]. In normal adults it has been shown that the activation of the MNS during action observation is related to the experience the observer has of the observed actions [3], [8]. This seems to suggest that the development of the MNS runs in parallel with the motor experience of the observer. On the other hand classical studies do suggest that human newborns, only a few days old, are able to resonate with other adult individuals' actions [45], and infants less than two years old can predict other people's action goals [46] thus suggesting that the brain is endowed with a mechanism subserving a close coupling between action observation and execution from early development.

MNS has been involved in a number of motor as well as cognitive functions [19], [34], [55]. Several studies have consistently shown that action observation is an effective way to learn or enhance the performance of that specific motor skill (observational learning) [for review see 66]. In a recent study participants who were required to perform a reaching task in a novel environment performed better after observing a video depicting a person learning to reach in the same novel environment as compared to participants who observed the same movements in a different environment [44]. In a pivotal study aimed at investigating the cortical mechanisms of human imitation, it has been found that areas involved in the actual execution of simple finger movements are crucially recruited also during the observation of identical movements made by another individual as compared to the presentation of spatial or symbolic cues [36]. Indeed during observational learning of complex actions, where mastering motor sequences not already in one's own motor repertoire is required, areas within the MNS have been found to be active from the observation of the model until its actual execution, with a specific contribution of prefrontal area 46 in reorganizing motor competences which are already part of our own motor repertoire in novel patterns fitting an observed model [4], [65]. Moreover, using transcranial magnetic stimulation (TMS) it has been shown that the observation of another individual performing simple repetitive thumb movements leads to a kinematically specific memory trace of the observed movements in the motor cortex [60].

As for motor imagery, it has been defined as the ability to “mentally rehearse simple or complex motor acts that are not accompanied by overt body movements” [39]. That is, an individual imagines himself executing a particular action, almost perceiving the kinesthetic experience of the movement. This may be referred to as kinetic imagery. Imaging studies have shown activation of regions including supplementary motor area, superior and inferior parietal lobules, dorsal and ventral premotor cortices, pre-frontal areas, inferior frontal gyrus, superior temporal gyrus, primary motor cortex (M1), primary sensory cortex, secondary sensory area, insular cortex, anterior cingulate cortex, superior temporal gyrus, basal ganglia and cerebellum [14], [20], [28], [32], [43], [53], [57], [59], [61]. That is, motor imagery is associated with activity in a variety of regions, some of which have specifically been shown to be involved in action execution and action observation. It has been forwarded that the MNS is recruited whenever motor representations are recalled, and therefore not only during action observation, but also during motor imagery, dreams with a motor content and so on even in the absence of overt action [38]. Evidence for an improvement in motor performance after motor imagery has been collected in several studies. In an early study participants were requested to learn a five-finger exercise on a piano key-board over five days. Participants could either actually perform the motor task or practice it mentally (motor imagery). Using TMS it has been shown that the reorganization of the motor cortex following actual practice or motor imagery was similar [50]. In a kinematics study in which participants had to point with their right arm to different targets placed in the frontal plane, both physical and motor imagery training led to an improvement of performance as revealed by a decrease of movement duration and an increase of peak acceleration, respectively [27]. Similarly, in a finger tapping task both participants who received motor training and those receiving motor imagery improved significantly in motor performance [49].

Motor imagery has been suggested by some to be as effective as physical practice in improving motor performance, whereas others suggest this improvement should be more modest [37]. Improvement may also depend on the degree to which an action falls within one's repertoire [48].

While both action observation and motor imagery appear to result in improved performance, it is still an open question as to which of these is more effective as the two processes have never been directly compared in a complex task.

We addressed this issue with a kinematics study aimed at directly comparing the efficacy of motor imagery and action observation (observational learning) in learning a novel complex motor task. As a control condition we used a motorically neutral task. All participants had to learn a complex and unusual motor task which involved movement of the right hand and foot in the same angular direction (in-phase movement), while at the same time moving the left hand and foot in an opposite angular direction (anti-phase movement) all at a frequency of 1 Hz. Kinematics of wrists and ankles were recorded to assess motor learning. The study included only one practice session and therefore was aimed at assessing the role of both motor imagery and action observation only in the fast phase of a typical motor learning process.

Section snippets

Participants

Forty-five healthy participants (29 women and 16 men), recruited at the Vita-Salute San Raffaele University of Milan took part into the study. They all had neither history of orthopedic or neurological diseases affecting upper and/or lower limbs nor special motor competence (e.g., athletes, dancers, musicians). Participants gave written consent before taking part into the experimental procedure. The study was approved by the Ethical committee of the Scientific Institute San Raffaele, Milan.

Discussion

The findings of our kinematics study clearly show that action observation is more effective than motor imagery (and, of course, than the control condition) in learning a novel, complex motor task. The results however were collected after one training session and therefore apply only to the fast phase of motor learning process. Both action observation and motor imagery most likely target the mirror neuron system, thus recruiting neural structures involved in the actual execution of the observed

Searching for roots of entrainment and joint action in early musical interactions

Jessica Phillips-Silver^{1*} Peter E. Keller^{2,3*}

¹ International Laboratory for Brain, Music and Sound, Montreal, Quebec, Canada

² Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

³ MARCS Institute, University of Western Sydney, Australia

When people play music and dance together, they engage in forms of musical joint action that are often characterized by a shared sense of rhythmic timing and affective state (i.e., temporal and affective entrainment). In order to understand the origins of musical joint action, we propose a model in which entrainment is linked to dual mechanisms (motor resonance and action simulation), which in turn support musical behavior (imitation and complementary joint action). To illustrate this

model, we consider two generic forms of joint musical behavior: chorusing and turn-taking. We explore how these common behaviors can be founded on entrainment capacities established early in human development, specifically during musical interactions between infants and their caregivers. If the roots of entrainment are found in early musical interactions which are practiced from childhood into adulthood, then we propose that the rehearsal of advanced musical ensemble skills can be considered to be a refined, mimetic form of temporal and affective entrainment whose evolution begins in infancy.

Introduction

What joint human behavior reveals greater coordination in time and affect – or entrainment – than music? When individuals join together in almost any musical behavior, ranging from a simple tune shared between a young child and caregiver, to a rhythmically complex performance of a Cuban Jazz band (Figure 1), their joint action is characterized by entrainment. The term entrainment (from the Middle French *entrainer*, v., to draw, drag) is used across diverse domains to refer generally to spatiotemporal coordination between two or more individuals, often in response to a rhythmic signal (e.g., [Phillips-Silver et al., 2010](#)). Used in musical contexts (including dance), entrainment typically refers to the rhythmic synchronization as a product of individuals' interaction ([Clayton et al., 2005](#)) – that is, rhythmic coordination with simple (e.g., 1:1 in-phase or anti-phase) or more complex (e.g., 2:3 or 3:4 polyrhythmic) phase relations.

figure 1



FIGURE 1. ILLUSTRATIONS OF TEMPORAL AND AFFECTIVE ENTRAINMENT. The skills of musical ensembles such as a Cuban Jazz band embody temporal entrainment and musical joint action. Affective entrainment can be seen in early forms of joint action between an infant and caretaker, and might provide a foundation for the development of musical skills.

One component of entrainment is thus temporal. Temporal entrainment to music can be observed at hierarchical levels of metrical periodicity in the body ([Leman and Naveda, 2010](#); [Naveda and Leman, 2010](#); [Toiviainen et al., 2010](#)) and brain ([Large and Jones, 1999](#); [Large and Palmer, 2002](#); [London, 2004](#); [Jones, 2009](#); [Nozaradan et al., 2011](#)). People often experience temporal musical entrainment in the automatic, even uncontrollable head-bobbing or toe-tapping that occurs when we listen to musicians play (or the covert activation of motor areas of the brain in the absence of overt movement), while the height of temporal entrainment might be seen in the complex rhythmic timing and exchange between partners or ensemble members in music or dance ([Keller et al., 2007](#); [Keller, 2008](#); [Goebel and Palmer, 2009](#)) as

portrayed by the band in Figure 1.

A second component of entrainment derives from the mutual sharing of an affective state between individuals. Authors have even spoken of a kind of affective “synchrony,” although we prefer the term entrainment in order to distinguish the temporal and affective components while encapsulating the essence of both. Therefore, for the purpose of this paper we will refer to both temporal and affective entrainment. Affective entrainment involves the formation of interpersonal bonds (see [Feldman, 2007](#)) and is related to the pleasure in moving the body to music and being in time with others (known in the vernacular as “groove”; [Pressing, 2002](#); [Madison, 2006](#); [Janata et al., 2012](#)). [Janata et al. \(2012\)](#) have put forth that temporal and affective components of entrainment in music might be inextricably linked, as for example, the quality of sensorimotor synchronization (temporal) can predict the degree of experienced groove (affective), and affiliation ([Hove and Risen, 2009](#)). In searching for the biological roots of temporal and affective entrainment, early musical interactions between infants and their caregivers might provide the first experiences in entrainment (e.g., [Malloch and Trevarthen, 2009](#)) and thus provide a foundation for the development of skills in musical joint action (Figure 1).

In joint action in general, multiple individuals coordinate their movements in space and time, with a goal to communicate, or to effect a change in the environment (see [Sebanz et al., 2006](#)). Here we take on the term “coordination” commonly used in this literature because movements between co-actors might be coupled in either a synchronized or complementary fashion, in time, or in space. Studies of the spatiotemporal aspects of interpersonal joint action reveal that coordination in joint action can be emergent or planned ([Knoblich et al., 2011](#)). Emergent coordination occurs spontaneously due to automatic processes that are grounded in perception and action couplings in the brain ([Knoblich et al., 2011](#)). In a non-musical example, when two conversational partners are engaged, they show a coordination of body sway that is unintentional and does not require planned action ([Shockley et al., 2003](#)). Unintentional coordination of body sway can also occur between duetting musicians ([Keller and Appel, 2010](#)), an example that will be discussed in more detail later. The emergent process can also be described in terms of concepts from dynamical systems theory (e.g., coupled oscillators; see [Schmidt and Richardson, 2008](#); [Oullier and Kelso, 2009](#); [Riley et al., 2011](#)). In contrast, planned coordination requires shared representations of the intended outcome of the joint action and a plan for each individual’s own role in the joint action. This can be seen in a variety of activities in which multiple individuals engage, including spoken conversation, or carrying a heavy object together (e.g., [Clark, 1996, 2005](#); [Knoblich and Sebanz, 2008](#); [Bekkering et al., 2009](#); [Sebanz and Knoblich, 2009](#)). In music, both emergent (automatic head-bobbing) and planned (playing and dancing along) coordination reflect entrainment and can exert simultaneous influences on the individuals’ actions.

How do we experience emergent and planned coordination in musical interactions where entrainment involves demands on temporal synchronization and complex rhythmic timing? To understand how these forms of musical joint action arise requires an excursion into several different research fields. The same terminology is occasionally used to different ends in these fields, and part of our task consists of an attempt to reconcile this with the aim of merging these diverse fields on common

ground (see **Box 1** for a glossary of our terminology and definitions). We begin in the next section by identifying two primary generic modes of musical joint action, chorusing and turn-taking, and their underlying entrainment mechanisms. Then in the section entitled “Ensemble Skills in Musical Joint Action” we consider how these mechanisms are crystallized in specific cognitive and motor skills that support musical ensemble performance. We then discuss in the section entitled “The Beginnings of Musical Joint Action” some of the potential ontogenetic roots for the cognitive and motor systems required in musical joint action. In the final section, we speculate on a possible mimetic relationship between developmental trajectories in social coordination skills and musical ensemble skills.

Box 1. A glossary of terms as they pertain to the concepts of musical joint action in this paper. Terms are listed in the order they appear in text.

Entrainment: The spatiotemporal coordination between two or more individuals, often in response to a rhythmic signal (e.g., [Phillips-Silver et al., 2010](#)). Examples include playing music or dancing together, in which temporal entrainment is particularly important, and infant–caregiver interactions, in which affective entrainment is often the primary goal (see [Figure 1](#)). Temporal and affective components of entrainment often coincide in musical joint action, and might be inextricably linked ([Janata et al., 2012](#)).

Joint action: Social interaction wherein multiple individuals coordinate their behaviors in space and time to communicate or to effect a change in the environment (see [Sebanz et al., 2006](#)). Two types of joint action, emergent coordination and planned coordination, are described below.

Emergent coordination: A type of joint action that occurs by spontaneous, automatic processes that are grounded in links between perception and action.

Planned coordination: A type of joint action that, in addition to basic entrainment, requires shared representations of the intended outcome of the joint action, such as in musical ensemble playing.

Chorusing: The production by separate individuals of simultaneous sounds or movements that signal joint action, such as in a chorus of voices, the beat of a drum circle, or a dance chorus line.

Turn-taking: The production by separate individuals of alternating or dovetailing sounds or movements that signal joint action, such as in “call and response” or musical fugue.

Imitation: An intentional or unintentional spatiotemporal mirroring of the actions (or their effects) of one individual by another. Examples occur in a choral unison or a group’s entrainment to a musical pulse.

Complementary joint action: An action achieved by partners with a common goal, and a systematic but non-identical spatiotemporal relation between the partners’ actions. An example occurs in hierarchical turn-taking as heard in the

trading of licks in a Jazz band.

Motor resonance: The automatic, bottom-up, exogenously driven activation of movement-related brain areas (e.g., premotor and sensorimotor cortex). Examples include an infant's non-conscious mimicry of the posture or affective expression of an adult or another infant.

Action simulation: The controlled, top-down activation of sensory and movement-related brain areas, that does not necessarily rely upon exogenous stimulation. Examples include the mental imagery of a sequence of musical sounds or dance movements.

Mimesis: An aesthetic representation or recreation of a physical reality. We propose an example of mimesis in the process of musical practice and expertise, which recalls and builds upon the stage of learning affective and temporal entrainment during infancy (see Figure 1).

Coordination Modes in Musical Joint Action: Chorusing and Turn-Taking

Chorusing and turn-taking are canonical modes of joint action that exemplify emergent and planned coordination. Chorusing occurs when communicative signals (sounds or movements) produced by separate individuals make simultaneous and roughly equal contributions to the joint action as a whole (as in synchronous chorusing observed in several species; [Merker, 2000](#)). For example, in multi-part music, monophonic (unison), and homophonic (chordal) textures can be considered to be instances of chorusing. In dance, the “chorus line” and the “corps de ballet” provide definitive examples of chorusing. Turn-taking involves the ordering of communicative signals produced by separate individuals in such a way that there is little temporal overlap (i.e., serial organization) or, when there is overlap, information in the signal produced by only one individual has priority at any given time (hierarchical organization). Musical turn-taking occurs in antiphonal “call and response” textures (as in Afro-American blues and gospel music; [Williams-Jones, 1975](#); [Waterman, 1999](#)) and, more generally, when instruments or dancers exchange roles, taking turns at leading and accompanying, such as happens in a Baroque fugue.

The distinction between chorusing and turn-taking is based on the temporal relations between actions of interacting individuals. Musical joint actions, however, have behaviorally relevant spatial, as well as temporal, components. Spatiotemporal movements shape dance forms and define the embodiment of metrical hierarchies ([Leman, 2007](#)). In a musical melody, the pitch intervals and contours define trajectories through mental representations of pitch space that can correspond cross-modally to physical space ([Rusconi et al., 2006](#); [Lidji et al., 2007](#); [Keller and Koch, 2008](#); [Eitan and Timmers, 2010](#)). To account for such spatiotemporal relations, chorusing and turn-taking can be classified with respect to two further general

classes of action: imitation and complementary joint action (Figure 2).

figure 2

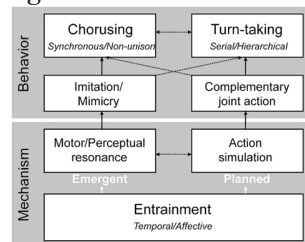


FIGURE 2. DIAGRAM ILLUSTRATING DUAL ROUTES BY WHICH ENTRAINMENT SUPPORTS MECHANISMS AND BEHAVIORS IN MUSICAL JOINT ACTION. One route leads to emergent coordination (e.g., automatic imitation and chorusing) via motor and/or perceptual resonance. The other route leads to planned coordination (e.g., complementary joint actions involving turn-taking) through action simulation. The vertical arrows in the figure represent relations between mechanisms and behaviors within each route, while horizontal arrows represent potential areas of overlap between the routes (e.g., imitative behavior may arise through resonance alone or through a combination of resonance and simulation). The dotted arrows imply that the behaviors and mechanisms at each level are not mutually exclusive, and may exist in hybrid forms (e.g., non-unison chorusing and hierarchical turn-taking).

Imitation entails one individual mirroring the spatiotemporal features of another's movements or their effects (see [Brass and Heyes, 2005](#)), often in a goal-directed manner ([Sebanz et al., 2006](#)). Imitation may occur intentionally, such as when a musician or dancer incorporates a pre-existing sequence in a variation on a theme, or unintentionally, as when two duetting players inadvertently copy each other's body sway or the effects of that sway on expressive parameters of the sound that they are producing ([Keller and Appel, 2010](#)). Imitation may occur at varying time lags, ranging from short intervals that give the impression of simultaneity (as in unison chorusing, or moving bodies or instruments to a common musical pulse) to longer intervals at which the model behavior and the copy do not overlap (serial turn-taking). Complementary joint action, which requires more sophisticated partnership skills, occurs when spatiotemporal features of one individual's behavior are different from, but systematically related to, those of another individual (see [Bekkering et al., 2009](#)). This is exemplified in non-unison chorusing and hierarchical turn-taking.

We assume that entrainment supports imitation and complementary joint action through two brain mechanisms that capitalize on links between perception and action. Specifically, imitation is mediated by motor resonance while complementary joint action requires internal simulation processes that may involve mental imagery. Motor resonance occurs when brain regions that control movement – including motor, premotor, and sensorimotor cortices – are activated automatically by the observation of another's movements ([Rizzolatti and Craighero, 2004](#)). This process may lead to non-conscious mimicry, where individuals unwittingly adopt the facial expressions, mannerisms, or postures of interaction partners ([Chartrand and Bargh, 1999](#)). Motor resonance may contribute to the perception of musical pulse ([van Noorden and Moelants, 1999](#); [Grahn and Brett, 2007](#); [Chen et al., 2008](#); [Large,](#)

2008) and, in both music and dance, may promote readiness for temporally coordinated, planned action.

Motor resonance evoked during the perception of music and dance does not only function in the service of action, but may also modulate aesthetic responses (Calvo-Merino et al., 2008; Kornysheva et al., 2010; Cross and Ticini, 2011). Motor representations of practiced actions result in stronger motor resonance in adults as well as in infants (Calvo-Merino et al., 2006; van Elk et al., 2008), which suggests that motor resonance is tuned by experience and, when combined with perceptual resonance (Schütz-Bosbach and Prinz, 2007a), can contribute to action simulation.

Action simulation, as conceived here, is richer than motor resonance to the extent that it involves the activation of sensory, in addition to movement-related, brain regions (cf. Schütz-Bosbach and Prinz, 2007a). Specifically, action simulation occurs when sensorimotor neural processes that resemble those associated with executing an action are engaged in the absence of overt movement (see Gallese et al., 2004; Wilson and Knoblich, 2005; Decety and Grèzes, 2006). Such covert activity may be triggered by observing – or by merely imagining – an action or its effects, for example, body movements in the case of dance (Cross et al., 2006), or tones in the case of music (Keller, 2008). This highlights what we believe to be a fundamental distinction between motor/perceptual resonance and action simulation: While resonance is driven exogenously and automatically (i.e., preattentively) by the perception of external events, simulation may be generated endogenously in the absence of external stimuli (cf. Grush, 2004) and may be modulated by attention.

Action simulation is a mark of expertise, as it is mediated by experience-based associations between sensory and motor processes (Haueisen and Knösche, 2001; Baumann et al., 2005; Bangert et al., 2006; Zatorre et al., 2007). Auditory-motor simulation is, therefore, especially strong in musicians (Lahav et al., 2007) and visuo-motor simulation is particularly potent in dancers (Calvo-Merino et al., 2005, 2006; Cross et al., 2006). It has been proposed that action simulation facilitates the understanding of others' intentions and affective states, as well as playing a role in predicting an observed action's immediate outcome and future course (e.g., Wilson and Knoblich, 2005; Decety and Grèzes, 2006; Schütz-Bosbach and Prinz, 2007b; Sebanz and Knoblich, 2009). In musical joint action, predictions based on simulation can facilitate interpersonal coordination in the context of musical textures characterized by chorusing and turn-taking (e.g., Keller et al., 2007; Keller and Appel, 2010).

Instances of unison chorusing – for example, singing a tune such as Auld Lang Syne, or moving one's body in matched fashion with other individuals as in a group folk dance – can be considered to be forms of imitation to the extent that co-actors adopt leader or follower roles. Spatiotemporal interpersonal coordination at this level can especially reveal the contribution of motor resonance triggered exogenously by perceptual input, in addition to the simulation, and action-planning required to execute the musical unison. A complex example that illustrates the contribution of multiple cognitive-motor processes is the Afro-Brazilian Congado – a musical ritual of religious significance that entails the simultaneous performance of separate but proximal communal groups of singers, percussionists, and dancers. A rigorous analysis of Congado performances (Lucas et al., 2011) has shown that inter-group

entrainment is influenced by proximity and visual contact, similarity of tempo, and intention (i.e., whether proximal groups resist entrainment), factors which reveal contributions of emergent entrainment and imitation, as well as simulation and action-planning in planned coordination. Finally, during turn-taking, the serial or hierarchical ordering of signals produced by different individuals (e.g., when “trading licks” in jazz or exchanging melody and accompaniment roles in chamber music) constitutes complementary joint action. Effective coordination during turn-taking therefore requires the participating individuals to predict the spatiotemporal trajectories of each other’s actions via simulation.

The cognitive and motor demands of chorusing and turn-taking vary as a function of the degree to which these modes of musical joint action require imitation or complementary joint action. The scheme in Figure 2 captures this through the use of arrows to delineate potential links between the two levels of behavior (chorusing/turn-taking and imitative/complementary) and entrainment-based mechanisms (motor/perceptual resonance and action simulation). Note that musical joint action may adopt hybrid forms that house elements of both imitative and complementary joint action (e.g., when one improvising musician imitates another’s rhythm while altering the pitches). In the next section we consider in more detail the cognitive and motor processes that serve advanced forms of musical joint action, such as skilled ensemble performance, through links to basic mechanisms of entrainment, motor resonance, and action simulation.

Ensemble Skills in Musical Joint Action

Musicians coordinate their actions with precision and flexibility. To engage in such planned coordination, whether it entails turn-taking or chorusing, ensemble musicians in many musical traditions invest considerable time into rehearsing together in order to form shared representations of the ideal sound (see [Davidson and King, 2004](#); [Ginsborg et al., 2006](#); [Davidson, 2009](#)). According to [Vesper et al. \(2010\)](#), the “minimal architecture” supporting generic joint action includes shared task representations and processes related to action monitoring, prediction, and behavioral modifications that simplify coordination. The latter “coordination smoothers” involve modulations of an individual’s own behavior that render the action more predictable and make it easier for others to coordinate with it ([Vesper et al., 2010](#)). In music performance, exaggerated movements associated with breathing, body sway, and ancillary performance gestures such as head nods ([Goebel and Palmer, 2009](#); [Keller and Appel, 2010](#); Keller, forthcoming) may serve as smoothers. Once shared goals are established, these musicians rely upon three core cognitive–motor ensemble skills that are based on entrainment, and allow each individual to coordinate with the actions of co-performers in real time (see [Keller, 2008](#)).

The most fundamental ensemble skill is adaptive timing, or adjusting the timing of one’s movements in order to maintain synchrony in the face of unintended temporal perturbations (e.g., by other players) and intended tempo changes in the music. Adaptive timing is mediated by temporal error correction mechanisms that enable internal timekeepers (i.e., neural oscillations at timescales relevant to musical meter) to remain coupled, or entrained in stable phase relations with external signals ([Repp,](#)

2005; Large, 2008; Repp and Keller, 2008). Temporal error correction processes that operate automatically support emergent coordination, while others that require attention may be invoked in contexts requiring planned coordination, such as when an ensemble musician intentionally adapts to the expressively motivated tempo changes of a co-performer (Keller, 2008). Research on adaptive timing has yielded evidence for assimilation in interpersonal action timing during dyadic sensorimotor synchronization. This has been observed in joint finger-tapping tasks that require basic in-phase coordination akin to chorusing (Konvalinka et al., 2010), as well as in tasks that require individuals to produce movements in alternation, which involves turn-taking. In these studies, cross-correlation analyses revealed interpersonal dependencies in the timing of co-actors' finger taps that were suggestive of mutual assimilation. Temporal assimilation may be a form of imitation that facilitates ensemble cohesion by making multiple individuals sound collectively as one.

A more cognitively advanced skill is prioritized integrative attending, that is, simultaneous attention to one's own actions (high priority) and those of others (lower priority) while monitoring the integrated ensemble sound (Keller, 1999, 2001). Prioritized integrative attending thus relies both on an individual's ability to divide attention between different sound sources, and on the group's joint attention skills (to the extent that multiple individuals must attend to the aggregate structure that results from their coordinated actions). This variety of joint attention is particularly important in music characterized by complex turn-taking. For example, to perform the interlocking rhythms of Central African music, according to Nketia (1962), each player must have a general awareness of the resulting aggregate pattern, as well as the "knack" for coming in at the right moment. It has been proposed that the dynamic allocation of attentional resources that is required for prioritized integrative attending relies on entrainment (Keller, 2008). Specifically, the coupling of internal timekeepers to periodicities associated with the music's metric structure may provide a temporal schema that allows attention to be modulated in a manner that is optimal for monitoring multiple levels of the musical texture concurrently (Keller, 1999; London, 2004; Jones, 2009). Evidence that such metric schemas facilitate prioritized integrative attending has been found in studies in which musicians are required to memorize one instrumental part, as well as the aggregate structure, of multi-part rhythm patterns (Keller and Burnham, 2005). For example, in one experiment listeners were required simultaneously to memorize a target (high priority) part and the overall aggregate structure (resulting from the combination of parts) of short percussion duets. The key finding was that recognition memory for both aspects of each duet was influenced by how well the target part and the aggregate structure could be accommodated within the same metric framework.

The third ensemble skill is anticipatory imagery, which involves the use of mental imagery in planning the production of one's own sounds and predicting upcoming sounds of co-performers. Auditory or motor imagery may thus assist in anticipating others' actions with a view to optimizing ensemble cohesion (Keller, 2008). In support of this hypothesis, it has been shown that individual differences in synchrony between pianists in duos are positively correlated with performance on a task designed to measure the vividness of anticipatory auditory imagery (Keller and Appel, 2010). With regard to the mechanisms underling this relationship, it has been claimed that anticipatory imagery relies on entrainment-based sensorimotor coupling to drive internal models that trigger mental images during action

simulation (Keller, 2008). These internal models, which allow sensorimotor transformations between bodily states and events in the immediate environment to be represented in the brain, are harnessed during action simulation to generate predictions about an action's future course (Wolpert et al., 1998, 2003). The accuracy and stability of sensorimotor synchronization in the context of real and virtual instances of interpersonal coordination have been shown to depend on temporal prediction abilities (Wöllner and Cañal-Bruland, 2010; Pecenka and Keller, 2011), which are, in turn, positively correlated with temporal imagery abilities (Pecenka and Keller, 2009a,b).

Thus a suite of cognitive and motor skills, operating in concert, supports successful joint action in musical contexts via entrainment-based mechanisms that enable diverse forms of chorusing and turn-taking behavior. This is particularly evident in the performances of expert ensembles, such as the Cuban Jazz band depicted in Figure 1. Some of these ensemble skills, in addition to coordination smoothers and shared task representations, can be founded on principles of joint action learned very early in life (Figure 1). We now turn to the early foundations of musical joint action.

The Beginnings of Musical Joint Action

Joint action as a general form of coordinated social action begins with the earliest relationship. Affective entrainment in social contexts seems to be a predisposition of the infant and mother (Trehub, 2003), as in the “intuitive parenting” that helps to regulate the infant's emotional state and guide learning in his pre-verbal environment (Papoušek, 1996b). One of the highlighted aspects of the mother–infant interaction is its musical nature. Musical qualities of typical infant-directed speech include slow and regular tempo, repetition, exaggerated prosody and the accompaniment of movement and gesture. These features may serve as coordination smoothers, as well as functioning to engage the infant's attention (Fernald, 1991), to enable emotional communication (Trainor et al., 2000; Trainor, 2008), and to promote language acquisition (Papoušek, 1996a; Kuhl, 2004).

The infant is responsive to such expressive musical characteristics in the mother's (or other caretaker's) displays, and also participates in “communicative musicality” (Malloch and Trevarthen, 2009; the infant's behavior has also been referred to as “protomusical,” e.g., Cross, 2003), for example, in predicting the timing of the mother's expressions and producing responses that are to some degree temporally coordinated (e.g., Crown et al., 2002). Jaffe et al. (2001) define temporal coordination as a form of interpersonal contingency, and they describe interactional (non-periodic) rhythms which allow an infant and parent to each predict the timing pattern of the other's behavior. This ability to predict timing is necessary in order to eliminate the time lag of a sensorimotor reaction (Merker et al., 2009), and is considered to be essential to bonding within the dyad (Jaffe et al., 2001). In a common game played by adults to engage infants, violation of expectation in timing is used to make the infants laugh, also supporting infants' ability for predictive timing (Stern, 2007). Infants' production of musical actions, as in sung tones or rhythmic movements, begins to emerge during the first year (perhaps especially with encouragement) and can be considered in the context of the communication of the

pre-verbal infant (Trehub, in press). Yet before full musical actions emerge, the infants' responsiveness to musicality and the timing of their responses may build upon a repertoire of non-verbal vocal and gestural expressions (Eckerdal and Merker, 2009).

Infants thus engage in a kind of "sympathetic conversation" with their mothers, the timing of which enables them to anticipate and relish in the mothers' expressive displays, and causes them distress if timing is mismatched (see Trevarthen and Aitken, 2001). Temporally coordinated actions can be simultaneous, dovetailing, or alternating (Feldman, 2007), though these interactions are not typically strictly periodic, and are stochastic and bidirectional in organization (Cohn and Tronick, 1988). When mother–infant interactions take the form of chorusing, they have been referred to as "coaction" (Dissanayake, 2000). In this context, the behaviors that infants display with their caregivers include social gaze, facial expressions, and vocal behaviors (by 3 months of age), gesture, and shared attention to objects (after 6 months of age; Feldman, 2007). The roughly unison or matched forms of interaction between infants and their caregivers might be attributed to motor resonance (e.g., Meltzoff and Decety, 2003), and reflect emergent coordination behavior. The function of this emergent coordination is to establish affective entrainment, cooperation, and bonding between the infant and her caregiver (Feldman, 2007; Feldman et al., 2007).

Turn-taking in infancy has been studied primarily in the context of observing spoken conversation, and indicates the importance of the development of social cognition and attention. Infants shift gaze or attention as they observe videos of adults engaged in conversational turn-taking (von Hofsten et al., 2005), and gaze shifts become increasingly predictive with age (Bakker et al., 2011). By 4–6 months infants are sensitive to cues of social cognition (such as selective attention to face-to-face interactions) in turn-taking (Augusti et al., 2010), which coincides with the development of their ability to use infant-directed speech cues to choose their preferred social partners (Schachner and Hannon, 2011). Before semantics and syntax are shared between conversational partners, cues to turn-taking are manifest in culturally dependent vocal prosody, eye gaze, and body movements (Wilson and Wilson, 2005) but with a universal target of timing in turn-taking (Stivers et al., 2009). A further cognitive skill in interpretation of conversational turn-taking is anticipating transitions in conversation (i.e., when the next speaker will begin to speak), which does not develop until around 3 years of age and might be influenced by language development (von Hofsten et al., 2009). This anticipatory timing skill might also correspond to improvements in planned coordination of actions (Figure 2).

Spatiotemporal imitation appears in the repertoire of young infants – for example in facial (see Meltzoff and Moore, 1997) and manual gestures (Bekkering et al., 2000), as well as affective mirroring, beginning with the first social smiles at just 6 weeks of age (Rochat, 2007) – which are all examples of emergent coordination (Figure 2). The timing of imitation is limited by cognitive–motor maturity but is taught in part by caregivers to young children, often via imitation games aimed at encouraging joint action (Gergely et al., 2002; Sebanz et al., 2006; Papoušek, 2007). In early stages imitation is automatic and relies primarily on motor resonance (Paulus et al., 2011), even more strongly in practiced behaviors such as crawling than in novel behaviors

(i.e., walking; [van Elk et al., 2008](#)). While many actions of young infants may rely on resonance, according to [von Hofsten \(2004\)](#), even from birth such actions are not mere reflexes but can be motivated, informed, and goal-directed. Evidence for this claim includes infants' interest in tracking and imitating the purpose and the outcomes of observed actions (e.g., [von Hofsten and Siddiqui, 1993](#); [Gergely et al., 2002](#); [Gergely and Csibra, 2003](#)). The goal-directed nature of infants' actions, revealing planning, prediction, and motor representations, could enable the progression (with muscular and especially cognitive development) from resonant imitation to more complex forms of joint and complementary joint action.

In the earliest joint actions that are performed by infants with adults, the adult typically helps the infant to achieve her goal, in which case a precise motor representation is not required ([Vesper et al., 2010](#)). By around 1 year of age, several cognitive changes facilitate joint action. First, joint attention has emerged (see [Rochat, 2007](#)), in which individuals knowingly attend to the same object or event. This skill coincides with monitoring of relative shared attention between the infant and his social partner ([Rochat, 2007](#)). The 1-year-old understands intentions, which has been argued ([Tomasello et al., 2005](#)) to be the basis for understanding beliefs (i.e., theory of mind) – emerging around 15 months ([Onishi and Baillargeon, 2005](#)) and continuing to develop with experience with language and shifting of perspective ([Tomasello et al., 2005](#)). At 1 year action-planning is evident, as infants show goal-directed eye movements that reflect motor representations ([Falck-Ytter et al., 2006](#)). These motor representations are thought to be supported by the mirror neuron system: a brain network that is recruited similarly during action perception and action execution ([Falck-Ytter et al., 2006](#)), and is thought to be involved in the interpretation of music and dance in adults ([Stevens et al., 2001](#); [Zatorre et al., 2007](#); [Overy and Molnar-Szakacs, 2009](#)).

Beyond the first year emerge the abilities to interpret goal-directed actions in a rational manner (e.g., [Gergely et al., 2002](#)), suppress imitative motor representations, and eventually perform complementary joint actions ([Sebanz et al., 2006](#)). Motivation for these changes stems in part from “shared intentionality,” or the sharing of psychological states in order to reach mutual collaboration ([Tomasello and Carpenter, 2007](#)). Planned coordination calls upon the more advanced cognitive–motor skills of precise (and shared) sensory and motor representations, action monitoring, and behavioral modification (cf. [Keller, 2008](#)). Presumably, concurrent with the maturation of the above-mentioned cognitive and motor skills (hence less reliance on scaffolding by adults) as well as language development (e.g., [von Hofsten et al., 2009](#)), the practice of musical activities during childhood may continue to engage attention and foster coordination skills. For example, prioritized integrative attending, when musical ensemble performers monitor the aggregate sounds (which children's choirs can do to an extent) presumably builds upon the earlier abilities for joint attention and dividing attention between actors, once a shared goal representation is also established. Between the ages of 2.5 and 3 years, the skills of coordination (timing) in joint action show substantial improvement even if individual performance on a task improves only marginally ([Meyer et al., 2010](#)). Complementary roles in joint action appear to be mastered from the age of 3 years (similar to linguistic turn-taking; see [von Hofsten et al., 2009](#)), as action-planning and control become refined ([Meyer et al., 2010](#)). Improvisation in social contexts may have a role in the building of planned coordination capacities

upon emergent coordination capacities. For example, turn-taking has been observed in children's vocal play, even in improvised and complementary forms (Dissanayake, 2000), and such vocal play is expressed in virtual social contexts, as in imaginary dialogs and play-acting (see Rochat, 2007).

To achieve temporal entrainment in music, it may be necessary to practice the skill of sensorimotor synchronization – that is, the ability to move in time with perceived external events (Repp, 2005). For example, the clapping games and nursery rhymes, songs, and group dances, of school-age children begin to demonstrate the kind of coordination that is required to keep time as a group (e.g., Provasi and Bobin-Begue, 2003), with an external periodic pulse (McAuley et al., 2006), and using multiple levels of metrical hierarchy (Drake et al., 2000). Such music and games, which represent collective social entrainment (Phillips-Silver et al., 2010), may also foster the development of ensemble playing, and play an important role in the improvement of automatic and deliberate adaptive timing skills, attention, and auditory–motor imagery (Keller, 2008).

The ability for precise synchronization seems to mature gradually, probably building on early perceptual abilities for processing the musical beat (Hannon and Trehub, 2005; Phillips-Silver and Trainor, 2005; Winkler et al., 2009). In studies of synchronization of body movement with music or with a musical partner (in children between the ages of 5 months and 5 years), the children's motions – or the sounds produced by them – are not tightly synchronized (phase-locked) to the musical beat (Eerola et al., 2006; Kirschner and Tomasello, 2009; Merker et al., 2009; Zentner and Eerola, 2010). This suggests that sensorimotor synchronization in music is not typically developed until sometime later in childhood or near adolescence (Merker et al., 2009; although cases of exceptional children's musical performances can suggest otherwise, e.g., Merker et al., 2009; Sowinski et al., 2009)¹.

Practice of the affective component of entrainment in joint action is natural, as infants and young children show a predisposition to “groove” to music in social contexts – that is, they are compelled to move to the music and derive pleasure from it (Janata et al., 2012). Infants and toddlers display a variety of dance gestures in response to music (Eerola et al., 2006), and they produce spontaneous dance motions more to music than to speech (Zentner and Eerola, 2010). The social component is clear in that infants' and toddlers' dancing is associated with positive affect (Zentner and Eerola, 2010), and young children's musical drumming is enhanced in a social context (Kirschner and Tomasello, 2009). The development of action simulation in childhood could further facilitate the understanding of others' affective states (Decety and Grèzes, 2006) and complementary joint action as in musical exchange (Kirschner and Tomasello, 2009), especially when the roles are truly complementary.

Does Ensemble Practice Show Mimesis of Early Coordination?

We have proposed that temporal and affective forms of entrainment together support musical joint action by enabling chorusing, turn-taking, and hybrid modes of interpersonal coordination via dual routes. On a low road, emergent coordination

arises through motor and perceptual resonance that underlies imitative behavior, while on a high road, planned coordination is achieved with the assistance of covert internal simulations that facilitate complementary joint action. We then described the relevance of these mechanisms and behaviors to ensemble skills that allow experienced individuals to coordinate their actions during group music making and dance, as well as to the consolidation of social coordination skills in human development. In the present section, we speculate that the development of coordinated social action in infancy and early childhood, and the practice of advanced musical ensemble skills at later stages, are mimetic processes. Specifically, the process of acquiring and refining skills pertaining to temporal adaptation, attention, and anticipation in early human development is repeated in the acquisition and refinement of skills that are required for precise and flexible interpersonal coordination in aesthetic displays of music and dance (Figure 1). This may be viewed as a form of mimesis where advanced interpersonal coordination skills serving artistic purposes (aesthetic communication through music and dance) are grounded in early coordination skills that serve basic functions (mother–infant bonding, cognitive–motor development, and social and cultural learning, cf. [Cross, 2001](#)). Indeed, it has been proposed that both the infant–caregiver relationship and the arts including music and dance reveal a propensity in humans to respond to temporally dynamic social stimuli with emotional affiliation and temporal cooperation ([Dissanayake, 2000](#); [Miall and Dissanayake, 2003](#)).

We have suggested that musical joint action builds upon early manifestations of affective and temporal entrainment. The affective component of entrainment – which is central to interpersonal synchrony in music – is grounded in emotional resonance and affective mirroring in early infancy ([Rochat, 2007](#)). This component is arguably the first and foremost in ontogeny of coordinated action and musical communication. For example, imitative forms of chorusing and turn-taking are mastered more readily than complementary varieties of musical joint action. Imitation may also provide the easiest route to the socio-emotional benefits of joint musical experience. Indeed, synchronous, matched motor activity has been found to foster interpersonal affiliation ([McNeill, 1995](#); [Hove and Risen, 2009](#)), cooperation and pro-social behavior ([Wiltermuth and Heath, 2009](#); [Kirschner and Tomasello, 2010](#)), and even altruism ([Valdesolo and DeSteno, 2011](#)).

The temporal component of entrainment emerges in various forms during childhood: through joint attention activities in infancy, the improvement of adaptive timing, and finally anticipatory processes in school-aged children. Formal music training plays a role in developing such capacities for attention and executive functioning (see [Hannon and Trainor, 2007](#)), as well as refined auditory–motor interactions (see [Zatorre et al., 2007](#)). [Hannon and Trainor \(2007\)](#) have suggested that joint task goals, action monitoring, and synchronization can contribute to the benefits of musical training on cognitive and motor processing abilities in musical joint action. As increasing capacities for imitation and complementary joint action are developed, the cognitive and motor demands of chorusing and turn-taking can be met in musical joint action.

Musicians and dancers often strive to attain the height of temporal and affective entrainment, and so we look to the roots of those behaviors to understand the process by which they are embodied. From the earliest musical exchanges between

the infant and mother, temporal and affective entrainment serve more than the primary bond: they lay the groundwork for the refinement of skills in joint action and musicianship. In infants and children, as well as in musical experts via a mimetic process, maturation and musical experience result in entrainment that allows for precision and flexibility in timing, a sense of participation and emotional communion, and a musical aesthetic that reveals the complexity and richness of human interaction.