Visuomotor correspondence in imitation and self-recognition

Richard Cook

Division of Psychology and Language Sciences

University College London

A thesis submitted to the Faculty of Brain Sciences of University College London for the degree of Doctor of Philosophy

Abstract

In order to imitate the actions of others actors must solve the visuomotor correspondence problem: Visual representations of actions must somehow be ‘matched-up’ with their corresponding motor programs. In addition, hypothesised motor contributions to perception require a solution to the reverse motor-to-visual correspondence problem. This thesis is concerned with the origins of visuomotor correspondence knowledge – how actors match visual representations of actions to the corresponding motor representations.

Chapter 1 describes rival accounts of the acquisition of visuomotor correspondence knowledge and evaluates the evidence for each. Two types of theory are reviewed: Associative solutions - which argue that correlated sensorimotor experience is necessary to link visual and motor representations - and nativist solutions - which posit innate visuomotor links or innate means to achieve such connections.

The first two empirical chapters address previous findings that appear to challenge associative accounts. Experiments 1 and 2 (Chapter 2) sought a better understanding of the mechanisms mediating superior recognition of self-produced movements relative to those produced by friends. Experiments 3 and 4 (Chapter 3) sought to identify the conditions necessary for the refinement of visuomotor correspondences. The results from these experiments indicate that our ability to match observed and executed actions is mediated by links acquired associatively during correlated sensorimotor experience.

Chapters 4 and 5 sought to determine which associative solution best describes the acquisition of visuomotor links. Experiments 5 and 6 (Chapter 4) tested whether the acquisition of visuomotor links is sensitive to contingency, while Experiments 7 and 8 (Chapter 5) sought to determine whether ‘second-learned’ visuomotor links are subject to contextual modulation. The results from both chapters suggest that the acquisition of visuomotor links conforms to the principles of associative learning established through the study of conditioning in humans and animals.

Acknowledgements

I, Richard Cook, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

I am grateful to the Economic and Social Research Council (ESRC), who, through their award of a studentship, made it possible for me to complete this thesis.

I would like thank my parents for making the last seven years possible. Without their patience and encouragement I would not have had the opportunity to study Psychology at UCL.

I would like to thank Inci Ayhan, Geoff Bird, Caroline Catmur, Brad Duchaine, Tony Dickinson, Clare Press, Liz Ray and Natalie Salem for their support and advice. I would also like to thank Chris McManus who was instrumental in the award of my PhD studentship.

Finally, I would like to extend special thanks to my supervisors – Alan Johnston and Cecilia Heyes. Their scholarship, razor-sharp insight, clarity of exposition, experimental rigor and outstanding capacity for original thought, has provided the best academic apprenticeship imaginable.

Contents

Chapter 1: Introduction

1.1 The visuomotor correspondence problem

Imitation refers to the replication by an individual of the topography - the configural spatial properties - of observed actions. Imitation may thus be distinguished from emulation; a related behaviour where actors seek to replicate the effects of observed actions on objects or the environment, but not necessarily by the same means (Tomasello, 1996). The capacity to imitate is important not least because it mediates a fundamental form of social learning - the ability to acquire adaptive behaviours through the observation of conspecifics' movements (Heyes, in press; Heyes & Galef, 1996). Moreover, being imitated seems to have significant interpersonal consequences, promoting pro-social affect and feelings of social fluency (Chartrand & Bargh, 1999; Lakin & Chartrand, 2003).

In order to imitate, actors have to identify the motor programs that will replicate the observed configuration of limbs and effectors when viewed from third person perspectives. However, this poses a visuomotor correspondence problem (Brass & Heyes, 2005). How are we able to 'match up' visual representations of actions to their corresponding motor programs? The visuomotor correspondence problem parallels the visual-haptic correspondence problem posed by Molyneux (Locke, 1694). In a letter to John Locke, Molyneux asked whether a man recovering from congenital blindness who had learned to discriminate a cube from a sphere based on tactile exploration of the two forms, would subsequently be able to identify the two objects by sight? In essence, Molyneux was asking whether the ability to match a haptic representation of an object to the corresponding visual representation is innate; that is, determined by factors present from birth, or whether this ability is learned through experience. This thesis is concerned with a conceptually similar question: How do third-person or allocentric visual representations of actions become 'linked' to their corresponding motor representations? In other words, how do we acquire visuomotor correspondence knowledge?

The need to solve the correspondence problem distinguishes imitation from other types of action. When executing goal-directed actions, actors are typically faced with a degrees of freedom problem. Given the large number of musculoskeletal elements involved, there is a high degree of biomechanical redundancy, giving rise to thousands of ways to achieve a particular goal. Actors are therefore faced with the problem of overwhelming choice due to under specification. However, when we imitate we must identify a motor program, or configuration of motor programs, which will replicate the topography of observed action, out of a repertoire of many hundreds of thousands of potential movements. Actors are therefore faced with a ‘needle in a haystack’ problem imposed by over specification.

The challenge posed by the correspondence problem is compounded for ‘perceptually opaque’ actions (Heyes & Ray, 2000). Some actions, such as hand movements, have a high degree of perceptual transparency in that the consequences of action can be viewed directly by the actor from third-person perspectives. Identifying corresponding visual and motor representations for perceptually transparent actions is easier because we can directly compare what we execute with actions observed from allocentric perspectives. However, other types of actions, such as facial expressions and whole-body movements, have greater perceptual opacity, in that the sensory consequences of action performance do not afford direct comparison with the actions of others. The acquisition of correspondence knowledge for perceptually opaque actions is therefore particularly challenging.

1.1.1 The correspondence problem posed by motor contributions to perception

Solving the visuomotor correspondence problem is a necessary step in a series of processes involved in imitation. However, imitation may not be the only cognitive ability which requires correspondence knowledge. Several authors have argued that the motor system contributes to the perception of action. This idea was first advanced over a century ago by James (1890) and Lotze (1852). More recently, several contemporary motor theories have been proposed which argue that a process of ‘covert simulation’ provides an online contribution to action perception by predicting likely limb trajectories and configurations (Blake & Shiffrar, 2007; Blakemore & Frith, 2005; Schutz-Bosbach & Prinz, 2007; M. Wilson & Knoblich, 2005). In the case of imitation, actors must solve a visual-to-motor correspondence problem; a visual representation of target action must be matched to the corresponding motor program. The hypothesised motor contribution to perception requires a solution to the reverse motor-to-visual correspondence problem; motor programs must be matched to their corresponding visual representation. Covert simulation can only aid perception if excitation of a motor program propagates to the corresponding visual representation – simulating an observed hand movement is unlikely to aid perception if it excites the visual representation of a facial expression. If the motor system were found to contribute to perception by enhancing the topographic descriptions of observed actions, it is possible that this contribution is mediated by the same visuomotor links that mediate imitation.

1.1.2 Mirror neurons and the correspondence problem

In recent years, there has been considerable interest in the correspondence problem prompted by the discovery of mirror neurons (MNs). MNs are single units identified in the ventral premotor (area F5) and inferior parietal cortices (area PF) of the macaque (di Pellegrino, Fadiga, Fogassi, Gallese, & Rizzolatti, 1992; Fogassi et al., 2005; Gallese, Fadiga, Fogassi, & Rizzolatti, 1996; Rizzolatti, Fadiga, Gallese, & Fogassi, 1996) which respond to both the sight and execution of transitive and communicative actions. Since their discovery in monkeys, evidence has accumulated suggesting that humans also have a mirror neuron system (MNS). Several neuroimaging studies have revealed premotor and parietal areas of the human brain which respond to the observation and execution of transitive and intransitive actions (Buccino et al., 2001; Gazzola, Rizzolatti, Wicker, & Keysers, 2007; Iacoboni et al., 1999; Kilner, Neal, Weiskopf, Friston, & Frith, 2009). Single-cell recording in patients with intractable epilepsy has also identified neurons with mirror properties in the medial wall (supplementary motor area, cingulate cortex) and the medial temporal lobe (hippocampus, parahippocampal gyrus, entorhinal cortex and amygdala) of the human brain (Mukamel, Ekstrom, Kaplan, Iacoboni, & Fried, 2010). Strictly congruent MNs, which respond selectively to the observation and execution of the same action, may be the neural substrate of visuomotor correspondence knowledge – congruent MNs may be the links between corresponding visual and motor representations of action (Heyes, 2011; Iacoboni, 2009; Rizzolatti & Craighero, 2004).

1.1.3 A note on terminology

In the context of this thesis, the term “visuomotor correspondence problem” will be used to refer to the challenge of matching *third-person* visual descriptions of action with the corresponding representations within the observer’s motor repertoire. A related correspondence problem is how *first-person* visual descriptions of actions become linked to the corresponding motor representations. Knowledge of this sort is likely to make an integral contribution to the online control of actions (Haruno, Wolpert, & Kawato, 2001; Wolpert, Doya, & Kawato, 2003) and contribute to visual perception via sensory prediction (Haruno et al., 2001; Voss, Ingram, Haggard, & Wolpert, 2006; Voss, Ingram, Wolpert, & Haggard, 2008). However, knowledge of the first-person visual consequences of our actions is insufficient to explain our ability to imitate the actions of others or how motor processes contribute to the topographic description of others' actions. Instead, these abilities require third-person visual representations to be matched to corresponding motor representations and vice-versa. For this reason the term "visuomotor correspondence problem" will be used throughout to refer to the '*third-person'* visuomotor correspondence problem.

The term "visuomotor correspondence knowledge" will be used throughout to refer to the means by which we solve the visuomotor correspondence problem – the knowledge that allows us to identify the motor programs which replicate the topography of observed actions, and potentially allows action performance to excite corresponding visual representations. Where necessary, the individual units of correspondence knowledge will be described as "visuomotor links" in preference to "associations" or "synapses", which in the present context, may convey unintended meaning. I will depart from this convention only when discussing specific theories.

Theories of the origins of correspondence knowledge will be broadly described as "associative" or "nativist". The former term will be used to refer to theories which argue that correlated sensorimotor experience is necessary to acquire correspondence knowledge. The latter term will be used to refer to theories which assert that correlated sensorimotor experience is not required. This dichotomy is appropriate because those theories which do not ascribe a crucial role to correlated sensorimotor experience, all rely on innate knowledge or mechanisms, and none deem independent sensory or motor experience to be critical to the emergence of correspondence knowledge.

1.2 Putative solutions to the correspondence problem

Having described the visuomotor correspondence problem in the previous section, the current section will consider putative solutions. While numerous theories describe functional roles of links between corresponding visual and motor representations in mature cognitive systems, relatively few address explicitly how this correspondence knowledge is acquired. For example, the predictive coding model (Kilner, Friston, & Frith, 2007a, 2007b) specifies how a mature MNS might achieve action understanding but makes no claims regarding the origins of the MNs (Press, Heyes, & Kilner, 2011). Similarly, the MOSAIC architecture has been used to model action understanding and imitation in a mature system, in which the correspondence problem has already been solved (Wolpert et al., 2003). Specifically, this account fails to explain how inverse- and forward-models can flexibly represent 'states' in either first- or third-person perspectives, despite being derived from 'first-person' sensorimotor experience.

The following section will consider those theories which address the origins of visuomotor correspondence knowledge. Associative theories, which assert that correlated sensorimotor experience is necessary to identify visuomotor correspondences, will be considered first. This will be followed by a review of those theories which argue that correlated sensorimotor experience is not required to identify corresponding visual and motor representations. A detailed review of the relevant empirical evidence will be presented in section 1.3.

1.2.1 Associative accounts

Several theories state explicitly, or imply, that correlated sensorimotor experience is necessary to solve the correspondence problem. In one form or another, associative theories all conceptualise visuomotor correspondence knowledge as direct links between the visual and motor representations of the actions, acquired through correlated sensorimotor experience. Some accounts (Associative Sequence Learning, Ideomotor Theory) invoke the psychological construct of 'associations' to describe these links. In contrast, Hebbian MN theory characterises these links in neurophysiological terms, as synapses between sensory and motor neurons. The strength of the associative models lies in their ability to account for the effects of sensorimotor training. In contrast, findings of neonatal imitation and motor contributions to perception in the absence of correlated sensorimotor experience are challenging for associative theories.

1.2.1.1 Associative Sequence Learning

Of the associative theories, the most detailed is the associative sequence learning (ASL) model. ASL argues that the correspondence problem is solved directly by associations between corresponding visual and motor representations, acquired through the same domain-general associative mechanisms that mediate conditioning. Where the visual and motor representations of actions are active, either at the same time, or soon after each other, in a predictive relationship, the two become associated. Thereafter the sight of an action may excite the corresponding motor representation (Figure 1). Insofar as it appeals to the learning principles established through the study of conditioning, ASL predicts that visuomotor link acquisition is sensitive to several factors, including temporal contiguity (the extent to which two events occur at the same time), contingency (the predictive relationship between two events), and context (the internal or physical environment present during learning).

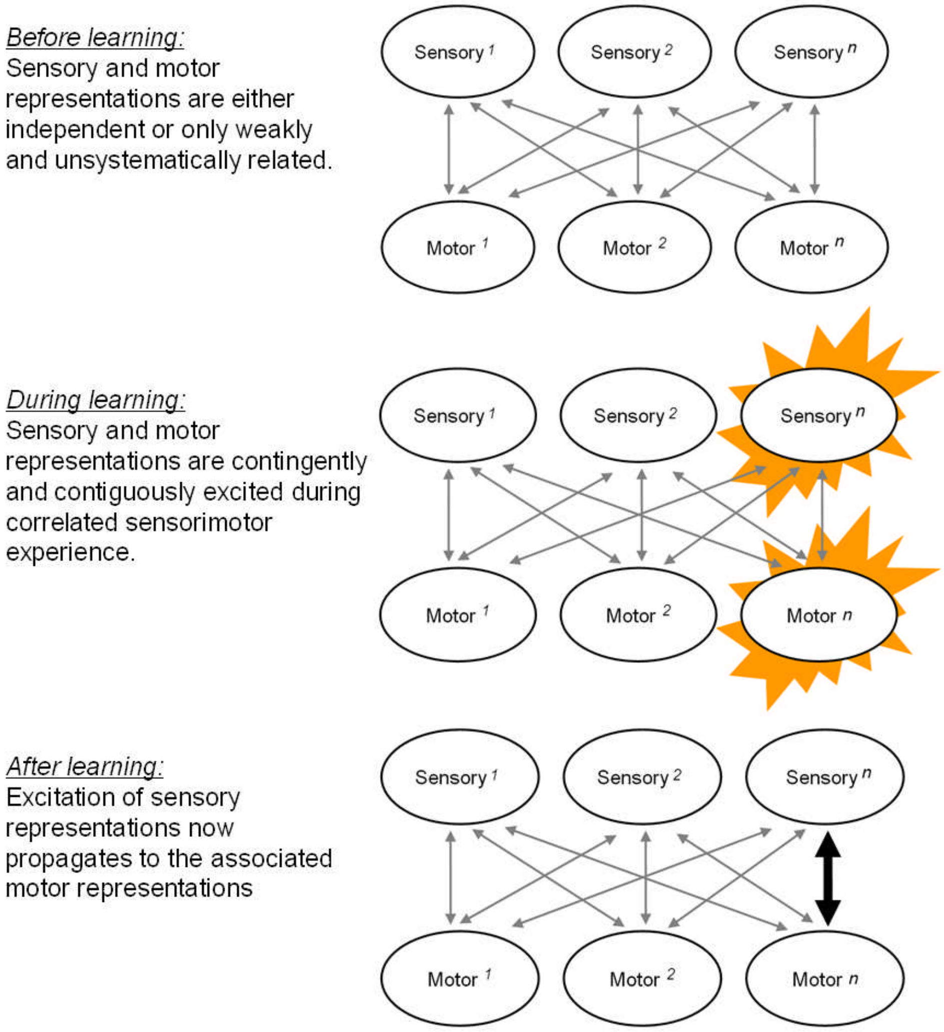

Figure 1: Schematic overview of the ASL model adapted from Heyes (2010).

Several types of situation might give rise to the contingent and contiguous activation of corresponding visual and motor representations (Ray & Heyes, 2011). When we observe our own actions, either directly or in mirrors, the motor commands to act predict the sight of self-produced actions. Thus the motor command to ‘open my hand’ is paired contingently with the sight of my hand opening. Although direct self-observation excites first-person visual representations of actions, this experience may still support the acquisition of the sensorimotor links thought to mediate imitation of actions viewed allocentrically. For actions with a high degree of perceptual transparency (e.g. hand movements) there is, by definition, a degree of overlap between the appearance of actions viewed from first- and third- person perspectives. For perceptually transparent actions, the links acquired between first-person visual representations and the corresponding motor representations may therefore help individuals match third-person visual representations to the corresponding motor representations.

Synchronous activities, where people react in the same way to a common event, may also give rise to sensorimotor experience which is both contiguous and contingent. For example, when a crowd at a sporting event rise in synchrony to cheer victory, or when an aerobics class respond to their instructor, corresponding visual and motor representations may be active at the same time in a predictive relationship. Being imitated by others is a further source of the correlated sensorimotor experience, deemed critical by ASL. There is mounting evidence that we are imitated by others from the moment we are born. Detailed analysis reveals that a striking proportion of parents' responses to infants are imitative (Malatesta & Haviland, 1982; Pawlby, 1977). Imitation also appears to be a pervasive feature of social interactions between adults, with interactants frequently engaging in unconscious, overt mimicry (Chartrand & Bargh, 1999; Cook, Bird, Lunser, Huck, & Heyes, 2011).

The ASL model allows that the visuomotor associations acquired during correlated sensorimotor experience are bidirectional. Once acquired associations may therefore permit both action observation to excite motor representations (as in the case of imitation), and action execution to excite visual representations (as in putative motor contributions to perception). Evidence that associations acquired during conditioning can show bidirectionality is provided by reports of 'backward conditioning'. A stimulus presented after an aversive reinforcer during training, can excite representations of the reinforcer at test, despite having been presented subsequently (Ayres, Haddad, & Albert, 1987). Similarly, having learned that particular responses predicted different tones during a response-effect training procedure, participants were faster to execute those responses to the onset of the tones in a stimulus-response task (Elsner & Hommel, 2001, 2004).

1.2.1.2 Ideomotor Theories

Ideomotor theories, which assert that motor programs come to be represented in terms of their sensory consequences through learning (Elsner & Hommel, 2001; Hommel, Musseler, Aschersleben, & Prinz, 2001; Prinz, 1997), imply a critical role for correlated sensorimotor experience. Many ideomotor theories fail to specify the precise nature of this learning process. However, the first stage of Elsner and Hommel's two-stage model of goal-directed actions (Elsner & Hommel, 2001, 2004) explicitly invokes associative learning theory to explain how actions may become associated with contingent sensory consequences. During self-observation, the sensory consequences of perceptually transparent actions are reliably predicted by the corresponding motor programs. The execution of a precision grip may therefore become associated with the observation of a precision grip, in much the same way as predicted by ASL. However, ideomotor theories are not purposely formulated to address the visuomotor motor correspondence problem. Because they focus on generic sensory consequences, they fail to address the acquisition of visuomotor correspondence knowledge for perceptually opaque actions, such as facial expressions or whole body actions, which yield very different perceptual consequences when observed and executed.

1.2.1.3 Hebbian mirror neuron theory

Hebbian MN theory (Del Giudice, Manera, & Keysers, 2009; Keysers & Gazzola, 2006; Keysers & Perrett, 2004) also suggests that the correspondence problem is solved directly by visuomotor links acquired through an associative process. This solution was formulated explicitly to explain the origins of MNs, and argues that these units are a product of contiguous sensorimotor experience. As the name of this model indicates, it is based on the principles advanced by Hebb (1949), and these principles assume that temporal contiguity is sufficient for associative learning. This is evident in Hebb's famous dictum "Cells that fire together, wire together," and was more formally stated: "any two cells or systems of cells that are repeatedly active at the same time will tend to become 'associated,' so that activity in one facilitates activity in the other" (Hebb, 1949, p. 70). Thus, the Hebbian perspective on MNs implies that contiguity is sufficient for MN development; that the learning process does not also depend on contingency. This represents an important distinction between the ASL and Hebbian models.

In its most recent formulation, the Hebbian account stresses the role of canalisation in the emergence of congruent MNs - units which respond to both the observation and execution of the same action (Del Giudice et al., 2009). The degree to which a trait or ability is 'canalised' describes the extent to which the emergence of a particular phenotype is robust to variations in genetic and environmental factors. The canalised Hebbian account argues that certain innate behaviours ensure the emergence of congruent MNs even if the Hebbian learning that mediates the acquisition of these units takes place in 'noisy' sensorimotor environments. Three canalising factors are suggested: First, infants show an innate visual preference for hands. This maximises the opportunities for Hebbian learning by ensuring visual and motor representations of action are frequently excited simultaneously. Second, infants typically engage in cyclical movements with a frequency of five cycles per minute. This frequency may stabilise long-term potentiation (LTP) at the synaptic level. Third, frontal and parietal regions of infants' brains show evidence of EEG synchronization in the theta frequency. The authors speculate that this may coordinate neuronal firing, further enhancing Hebbian learning. The combination of these factors is thought to protect the emerging MNS from perturbations in the form of non-congruent MNs - units which respond to different actions in observe and execute conditions.

1.2.2 Nativist accounts

Nativist theories argue that the acquisition of visuomotor correspondence knowledge does not require correlated sensorimotor experience. Instead, these theories assert that we are born with innate correspondence knowledge (innate MN theory), or the means to identify corresponding visual and motor representations endogenously – in other words to match observed and executed actions in the absence of correlated sensorimotor experience (active intermodal mapping). In contrast to associative theories, nativist models can readily account for neonatal imitation and evidence of motor contributions to perception in the absence of correlated sensorimotor experience. The weakness of nativist solutions is their inability to account for the effects of correlated sensorimotor experience, when the independent effects of sensory and motor experience, are controlled for.

1.2.2.1 Active intermodal mapping (AIM)

The AIM model proposes that humans are endowed with an innate mechanism which permits the 'seen-but-unfelt' to be matched to the 'felt-but-unseen'. The AIM hypothesis was originally formulated to explain facial imitation by neonates (Meltzoff & Moore, 1997), but has since become a prominent account of both infant and adult imitation (Meltzoff & Decety, 2003). Whereas associative accounts offer 'direct' solutions to the correspondence problem (sensory and motor representation are connected directly by an association or synapse), the solution offered by AIM is ‘indirect’ in two senses. First, the solution is mediated by proprioception; the visuomotor correspondence problem is solved by an innate mechanism which first solves the visual-proprioceptive correspondence problem. Second, motor representations are not immediately and automatically triggered following action observation; rather imitation is thought to be an intentional controlled process, which may occur after a delay (Meltzoff & Moore, 1997).

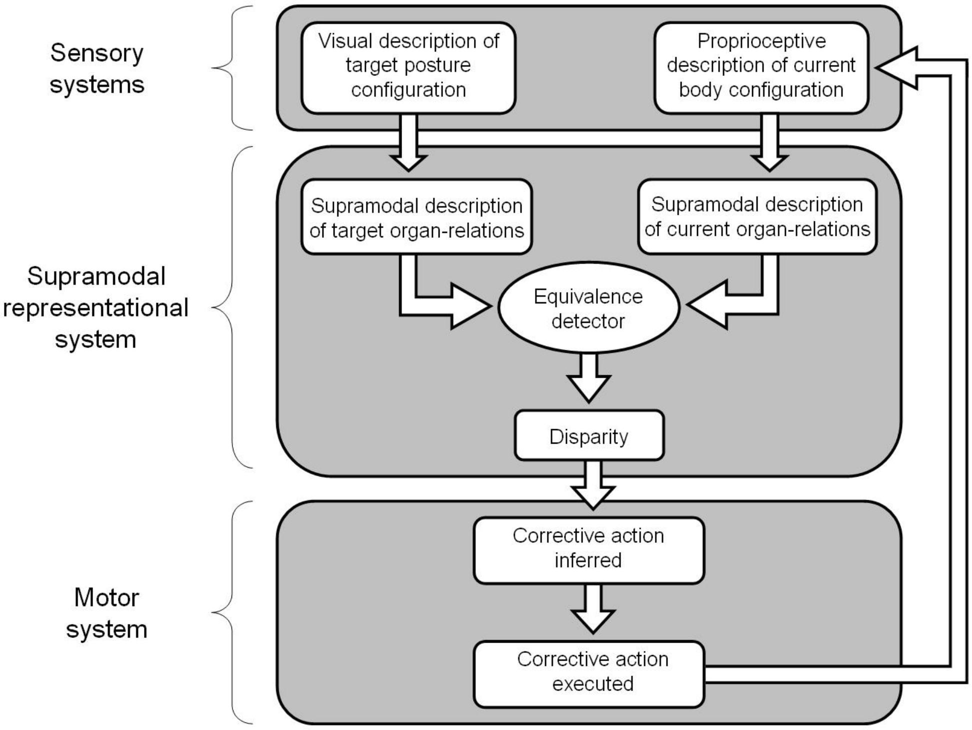

Figure 2: Schematic overview of the AIM model, adapted from Meltzoff & Moore (1997).

The model asserts that human infants are endowed with innate knowledge of the correspondence between their body schema and the physiology of those around them. This innate knowledge mediates ‘organ identification’, the first step in the imitation process, where actors identify what part of their body to move. Organ identification ensures that initial imitative movements are executed with the appropriate effector, even if the observed action topography is not replicated precisely. According to AIM, humans are also endowed with innate imitation machinery, specifically a supramodal representational system which encodes visual and proprioceptive representations of observed and executed actions in terms of ‘organ relations’ using a ‘supramodal code’. Within the AIM framework this supramodal code is the common metric which allows the correspondence between observed and executed acts to be computed (Figure 2). The precise nature of this code is not specified by the authors. However insofar as it is thought to be modality independent, it may be distinguished from a motor, visual or proprioceptive description. It has been suggested that MNs may mediate supramodal representation (Lepage & Theoret, 2007).

Once the actor's current posture and the desired target configuration are described in a common supramodal metric, imitation is achieved through an iterative, match-to-target process, mediated by a proprioceptive feedback loop. Where a mismatch between the organ relations of the observed target and the current bodily state is identified by an 'equivalence detector', a corrective action is inferred and executed. Thereafter, the proprioceptive feedback from the new body configuration is processed, and where necessary, a new iteration commenced. This process continues until the disparity between the visual and proprioceptive representations is minimised or eliminated.

According to AIM, the process of 'organ identification' and the first imitative attempts are thought to be mediated by innate knowledge. Moreover, the mechanism which detects and describes the disparity between observed and executed actions is also thought to be innate. However, Meltzoff and Moore assert that the ability to infer and execute corrective actions is informed by motor experience. The authors speculate that a process of 'body-babbling' where infants explore the range of their motor abilities, allows the infant to map specific muscle movements onto supramodal descriptions of the resulting 'organ-relations'. The accuracy of imitative error correction is therefore constrained by the amount of motor experience - only through acting can the individual develop a 'directory' of muscle movements mapped in terms of their supramodal consequences. Once acquired this directory allows the infant to infer and execute corrective movements where target-attempt disparities are detected.

1.2.2.2 Innate mirror neuron theory

Several authors have claimed that MNs are a product of natural selection. The view that MNs are an evolutionary adaptation 'for' a particular function implies that MNs are present from birth, because innate correspondence knowledge conveyed upon our ancestors an evolutionary advantage (Heyes, 2010). It has been proposed that early selection pressure favoured those monkeys with MNs because they afforded 'action understanding' (Nelissen, Luppino, Vanduffel, Rizzolatti, & Orban, 2005; Rizzolatti &

Arbib, 1998; Rizzolatti & Craighero, 2004; Rizzolatti et al., 1996). Because MNs discharge both when a monkey observes and executes similar actions, some of the neurons responsible for action execution are active during observation. This direct mapping between visual and motor representations of actions may provide the observer with ‘first-person’ insight into the actions they witness.

At subsequent stages in primate evolution, MNs may have started to convey further adaptive benefits including theory of mind (Gallese & Goldman, 1998), imitation learning (Rizzolatti & Craighero, 2004) and language development (Rizzolatti & Arbib, 1998). In each case, covert simulation within the motor system of the observer is thought to excite the consequences, goals, and motivations that are associated with the execution of the observed action. This simulation induced propagation allows the observer to use their own experiences of action execution to interpret the actions of others. The evolutionary timeline (action understanding > imitation, language development, theory of mind) is inferred from reports that while monkeys have MNs, only humans and apes are thought capable of imitation, language and theory of mind (Rizzolatti & Craighero, 2004). However, this inference has not gone unchallenged (Hickok & Hauser, 2010).

1.3 Evidence

Having outlined the rival associative and nativist solutions to the correspondence problem, this section will review the existing empirical evidence. First, evidence will be considered which suggests that correlated sensorimotor experience is necessary to acquire visuomotor correspondence knowledge. This will be followed by discussion of evidence which appears to challenge this view.

1.3.1 Evidence in favour of associative accounts

Evidence consistent with associative accounts has been derived from three sources: i) neuroimaging studies of sensorimotor expertise; ii) sensorimotor training effects; and iii) evidence from single-cell recording in macaques.

1.3.1.1 Neuroimaging studies of sensorimotor expertise

Neuroimaging studies of sensorimotor expertise take advantage of naturally occurring variations in participants’ sensorimotor experience. Studies of expertise suggest that learning of some kind is important in the development of the MNS (Calvo-Merino, Glaser, Grezes, Passingham, & Haggard, 2005; Calvo-Merino, Grezes, Glaser, Passingham, & Haggard, 2006; Cross, Hamilton, & Grafton, 2006; Haslinger et al., 2005; Jackson, Meltzoff, & Decety, 2006; Margulis, Mlsna, Uppunda, Parrish, & Wong, 2009; Vogt et al., 2007) For example, Haslinger et al. (2005) presented expert pianists and nonpianist controls with sequences of matched finger movements either similar or dissimilar to piano playing. They found that observation of the piano-related finger movements elicited stronger MNS responses (left inferior frontal gyrus, bilateral inferior parietal areas) in the pianists, but not in the nonpianist controls.

Similarly, the results of Calvo-Merino et al. (2005) also suggest that MNS activation is greater when observers have related motor expertise. Experts in classical ballet, experts in capoeira, and inexpert controls were presented with sequences of capoeira and ballet actions, matched for kinematic properties. Whereas the inexpert controls showed comparable mirror activation for both types of stimuli, the dancers showed a clear effect of expertise. A stronger blood-oxygen-level-dependent (BOLD) response was observed in the premotor, parietal, and posterior superior temporal sulcus (STS) regions when the ballet experts observed the ballet sequences, than when they viewed the capoeira stimuli. Using a similar design, Calvo-Merino et al. (2006) subsequently reported that male and female ballet dancers show greater MNS activation when viewing movement sequences unique to their gender.

While studies of sensorimotor expertise suggest a role for sensorimotor learning in the development of the MNS, they do not specify the precise nature of this contribution. Several of the effects (Calvo-Merino et al., 2005; Haslinger et al., 2005; Jackson et al., 2006; Margulis et al., 2009; Vogt et al., 2007) could be due to independent sensory or motor experience. For example, the stronger MNS response observed by Calvo-Merino et al. (2005) when dancers viewed movements from their own genre could be a perceptual learning effect resulting from greater visual experience. This account cannot be advanced to explain why ballet dancers show stronger responses to movements unique to their gender (Calvo-Merino et al., 2006), because dancers have extensive visual experience of male and female dance moves. However, this effect could still be a product of greater motor fluency resulting from repeated execution, rather than a product of correlated sensorimotor experience.

1.3.1.2 Sensorimotor training effects

1.3.1.2.1 Enhancement of visuomotor links

Automatic imitation is a widely documented effect in which the topographical features of task-irrelevant action stimuli facilitate similar, and interfere with dissimilar, motor responses (Heyes, 2011). The effect may be quantified using a stimulus-response compatibility (SRC) paradigm; a two-by-two factorial design where responses (e.g. hand-open or hand-close responses) can be compatible or incompatible with a task-irrelevant stimulus (e.g. hand-open or hand-close stimuli). Participants are typically faster to make compatible, imitative responses than non-compatible, non-imitative responses. For example, hand-open responses to the onset of hand-open stimuli are faster than to the onset of hand-close stimuli (Heyes, Bird, Johnson, & Haggard, 2005). Similarly, participants execute finger lift responses faster to the onset of finger lifting stimuli than finger tapping stimuli (Brass, Bekkering, & Prinz, 2001). The finding that participants make faster imitative responses than non-imitative responses is extremely robust (Heyes, 2011), having been reported for several effector systems (Gillmeister, Catmur, Liepelt, Brass, & Heyes, 2008; Leighton & Heyes, 2010), and for both transitive (Craighero, Bello, Fadiga, & Rizzolatti, 2002) and intransitive actions (Press, Bird, Walsh, & Heyes, 2008). Moreover, studies have confirmed that this effect is truly imitative; it depends on the topography of observed action - on how body parts move relative to one another; not merely on spatial compatibility - the position of the action relative to an external frame of reference (Catmur & Heyes, 2010).

The observation of human actions elicits stronger automatic imitation effects than actions that appear to be executed by ‘robotic’ agents (Press, Bird, Flach, & Heyes, 2005). This is consistent with the view that automatic imitation is mediated by direct visuomotor links established through correlated sensorimotor experience; we imitate robotic actions less because we rarely execute actions while observing robotic movements. However, Press, Gillmeister & Heyes (2007) demonstrated that automatic imitation of robotic actions can be increased by periods of compatible sensorimotor training where participants made hand-open responses to the opening of robotic hands, and hand-close responses to the closing of robotic hands. At a post-test, administered after six blocks of compatible training, automatic imitation of robotic actions was significantly stronger than at a pre-test, administered prior to training. In contrast, a second group given incompatible training (making hand-open responses to closing robot stimuli and hand-close responses to opening robot stimuli) showed a significant decline in automatic imitation between pre- and post-test.

The foregoing studies inferred enhancement of visuomotor links through greater automatic imitation. However, other studies indicate that sensorimotor training can enhance the BOLD response of the MNS. Cross, Hamilton, Kraemer, Kelley and Grafton (2009) trained participants to perform dance sequences over a five-day period. Some sequences were trained using compound stimuli depicting a human dancer and abstract ‘symbolic’ cues to the appropriate dance movements, while other sequences were trained using the symbolic cues alone. Following training, the authors observed that right ventral premotor cortex – a key element of the putative MNS – not only showed a main effect of training (trained sequences > untrained sequences), but showed broadly comparable responses to the compound stimuli (human dancer present) and the abstract cues alone (human dancer absent). If the visuomotor links comprising the MNS were an innate endowment, one might expect them to show biological-action specificity. Clearly our ancestors were never exposed to the sorts of arbitrary cues used to train dance movements in this study. These data are therefore consistent with the view that the correspondence knowledge encoded within the MNS is acquired through correlated sensorimotor experience.

Sensorimotor training has also been shown to establish novel visuomotor links between arbitrary shapes and hand actions within the MNS (Press, Catmur, Widman, Heyes, & Bird, under review). During a training phase, participants executed various hand actions contingent on the presentation of different coloured shapes (e.g. index finger point to green triangle). Thereafter, fMRI adaptation was used to demonstrate that the trained shapes elicited a MNS response similar to that typically seen during action observation. The fMRI adaptation paradigm takes advantage of the fact that repeated stimulus presentation or repetitive execution of actions causes a decrease in neural responses due to habituation or adaptation (Grill-Spector, Henson, & Martin, 2006). Following sensorimotor training with the shape-action response mappings, the consecutive presentation of a trained shape and the corresponding action stimulus elicited a suppressed BOLD response to the second stimulus in premotor and inferior parietal regions. Moreover, a cross-modal effect was also observed whereby presentation of a shape followed by execution of the trained mapping, or vice versa, produced a suppressed response to the second event. These data suggest that sensorimotor training causes geometric shapes and the contingent hand actions to recruit common neural representations within the MNS. This finding therefore accords with the view that visuomotor links represented within the MNS develop through correlated sensorimotor experience.

Each of these studies (Cross et al., 2009; Press et al., under review; Press et al., 2007) suggest that sensorimotor training can establish or enhance visuomotor links. In particular, the robotic action training study provides a compelling demonstration of this effect. In this study, both the compatible and incompatible groups were exposed to the same number of robotic hand stimuli and executed the same number of open- and close-hand responses. The effect observed therefore appears to be due to the nature of the sensorimotor contingency experienced during training, and cannot be attributed to independent sensory or motor learning. The fMRI adaptation evidence is harder to interpret. There is considerable debate regarding the underlying cause of fMRI adaptation effects (Grill-Spector et al., 2006; Larsson & Smith, 2011), further complicated by an absence of neurophysiological evidence that MNs adapt. In addition it is not clear whether the learning observed reflects the acquisition of stimulus-response (S-R) associations between the shapes and the motor representations of actions, or stimulus-stimulus (S-S) associations between the shapes and visual representations of actions.

1.3.1.2.2 Counter-mirror training

‘Counter-mirror’ training studies have provided evidence that several markers of human MNS functioning may be modulated by correlated sensorimotor experience. The logic behind these training studies is simple: If MNS responses are mediated by visuomotor links acquired through sensorimotor learning; mirror effects should be reduced or reversed by training in which the execution of one action is contingent upon the observation of another. The first counter-mirror training study demonstrated that automatic imitation can be abolished through sensorimotor training (Heyes et al., 2005). Participants in one group received incompatible training in which they were required to open their hand in response to closing-hand stimuli, and vice versa; whereas participants in a second group received compatible training in which they simply had to imitate opening- and closing-hand stimuli. When subsequently tested, the compatible training group showed the typical automatic imitation effect, responding faster to the compatible than to the incompatible stimuli. However, the counter-mirror training group showed no effect of S-R compatibility, thus failing to demonstrate automatic imitation.

Since this original demonstration, counter-mirror learning has been shown to modify several other markers of MNS function (Catmur et al., 2008; Catmur, Mars, Rushworth, & Heyes, 2010; Catmur, Walsh, & Heyes, 2007). During passive observation of actions, participants typically display a ‘mirror pattern’ of motor evoked potentials (MEPs). When transcranial magnetic stimulation (TMS) is applied to the primary motor cortex during passive action observation, the amplitude of the MEPs recorded from the muscles required to execute that action is greater than the amplitude of the MEPs recorded from task-irrelevant muscles (Borroni & Baldissera, 2008; Enticott, Johnston, Herring, Hoy, & Fitzgerald, 2008; Fadiga, Fogassi, Pavesi, & Rizzolatti, 1995; Maeda, Kleiner-Fisman, & Pascual-Leone, 2002; Patuzzo, Fiaschi, & Manganotti, 2003; Strafella & Paus, 2000). MEPs are widely interpreted as an index of corticospinal excitability. Selectively enhanced MEPs during action observation are therefore thought to reflect modulation of the motor system by a human MNS. However, Catmur et al. (2007) found that counter-mirror training reverses the typical ‘mirror’ pattern of MEPs. Prior to training, the typical pattern was observed: MEPs recorded from the index finger muscle were greater during observation of index finger abduction than during observation of little finger abduction, and vice versa for MEPs recorded from the little finger muscle. Following incompatible training, where participants made index-finger abductions to the onset of little-finger stimuli and little-finger abductions to index-finger stimuli, a counter-mirror pattern of MEPs was observed: Greater MEPs were recorded from the index finger muscle having observed a little finger movement and vice versa. In a subsequent follow-up it was confirmed that the counter-mirror MEP effect originated from the same premotor region, as the original mirror MEP effects (Catmur et al., 2010).

Similarly, Catmur et al. (2008) also found counter-mirror training reversed the baseline BOLD response seen in fMRI paradigms. Participants in a compatible training group were required to make foot-lifting responses to a raised-foot stimulus, and hand-lifting responses to a raised-hand stimulus. Conversely, participants given counter-mirror training had the opposite pattern of incompatible S-R mappings. Comparison of the groups following training revealed that the relative action observation properties of the MNS were reversed. Premotor areas that showed greater fMRI responses to observed hand actions in the compatible training group, responded more strongly to observed foot movements in the incompatibly trained group.

1.3.1.3 Evidence from single-cell recording in macaques

Neurophysiological evidence also suggests that the MNs studied in the macaque, thought to mediate visuomotor correspondence knowledge, acquire their properties ontogenetically. MNs have been reported in the ventral premotor area F5 which respond when the macaque observes actions performed by an experimenter with a tool (pliers or a stick) and when the monkey executes the same action with their hands (Ferrari, Rozzi, & Fogassi, 2005). Testing was conducted after a two month training period during which tools were used to grasp and pass food items to the monkey. According to associative accounts, this sort of experience would be likely to cause motor representations for grasping food items to become associated with the visual representations of actions made with sticks and pliers, because the former was reliably predicted by the latter.

So-called ‘audiovisual’ MNs have also been identified in the F5 region of the macaque premotor cortex (Keysers et al., 2003; Kohler et al., 2002). In addition to responding to the sight and execution of actions, these neurons also respond to the sounds associated with actions. A range of ripping and tearing sounds cause F5 MNs to discharge including the sound of a peanut breaking; paper ripping; plastic crumpling; metal hitting metal and paper shaking. This finding is entirely consistent with an associative view. There is a strong contingency between the sound of a peanut breaking and the sight and execution of that action. Repeated exposure to these sensorimotor contingencies is likely to cause the motor representation for ripping and both the auditory and visual sensory representations of paper tearing to become associated. Consequently, what may have once been a motor neuron may become an audiovisual MN.

An alternative ‘mediated activation’ account may be advanced to explain both of these findings if it is assumed that the sight of the tool actions, or the sounds associated with ripping and tearing actions, were associated not with motor programs directly, but rather with existing visual representations of hand actions. The observation of actions made with tools or the sound of ripping and tearing might not excite motor representations directly, but rather indirectly, via a hardwired visual representation of grasping (Figure 3). According to this account, links between sensory representations are acquired through experience, rather than links between sensory and motor representations. These accounts are not mutually exclusive, indeed associative accounts explicitly predict both types of learning.

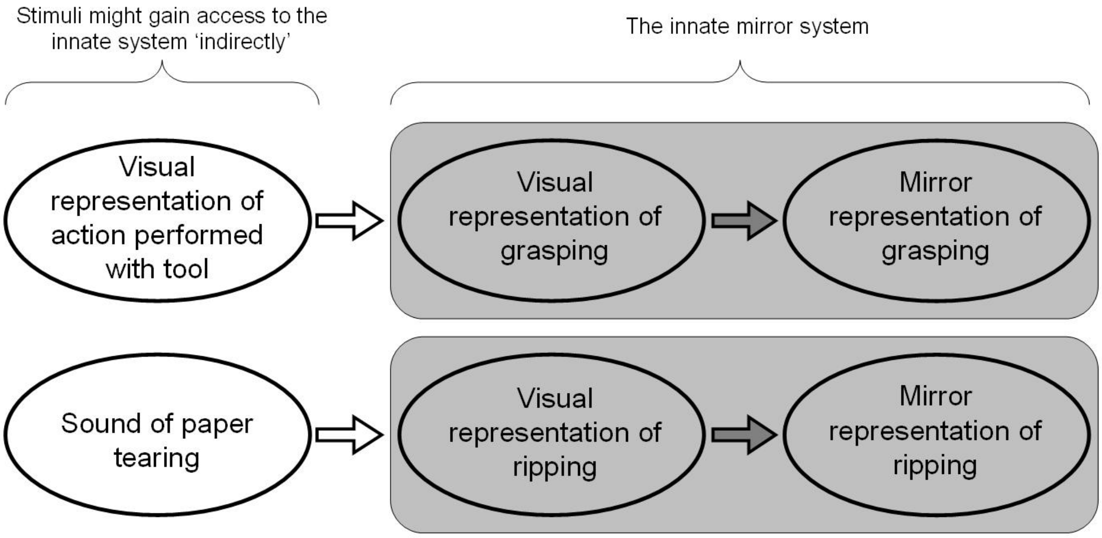

Figure 3: A 'mediated activation' account of MN responses to the sight of tool use and the sounds associated with actions. While this interpretation is logically possible, findings that MNs show stronger responses to the sight of actions with tools and to action sounds than to the sight of grasping or ripping alone, argues against mediated activation. Instead, some mirror neurons appear to have receptive fields tuned to sensory inputs other than the sight of gripping or tearing.

However, while logically possible, mediated activation accounts cannot explain all of the neuronal responses observed. Crucially, tool use MNs discharged significantly less often, if at all, to the sight of actions performed with biological effectors, than to actions performed with tools. Similarly, several of the audiovisual MNs studied also showed no response to the sight of the actions alone (Kohler et al., 2002), or responded more strongly to the sound of actions than to the combined sight and sound of actions (Keysers et al., 2003). These observations are inconsistent with a mediated activation account as they imply that the receptive fields of tool use and audiovisual MNs are tuned to sensory inputs associated with tool use and action sounds rather than the sight of actions executed with biological effectors. A mediated activation account predicts the opposite pattern; that MNs ought to respond maximally to the sight of hand actions executed with biological effectors, indicative of tuning, and weaker responses to other sensory consequences. This provides a clear indication that the sight of actions executed with tools or the sounds of actions, excited motor representations of grasping directly and not indirectly via the visual representation of grasping.

Reports of tool-use and audiovisual MNs do not argue decisively in favour of associative accounts because it is not clear what type of experience is causing these units to emerge. In both cases individuals were exposed to correlated sensorimotor experience, however, in the absence of suitable control conditions, one cannot exclude the possibility that independent perceptual or motor experience was crucial. Nevertheless, these findings indicate that the receptive fields of MNs may be tuned to sensory inputs to which the subjects' ancestors could not possibly have been exposed; e.g. the sight of actions performed with pliers or to the sound of a plastic crumpling. The existence and properties of tool-use and audiovisual MNs therefore challenge the view that MNs are an innate endowment.

1.3.2 Evidence against associative accounts

Evidence that appears to challenge associative solutions to the correspondence problem has been derived from three sources: i) neonatal imitation; ii) MNS responses in the absence of correlated sensorimotor experience; and iii) motor-to-visual transfer in the absence of correlated sensorimotor experience.

1.3.2.1 Neonatal imitation

1.3.2.1.1 The basic tongue-protrusion effect

Reports of imitation by newborn infants (< 6-weeks-old) have been interpreted by some as evidence of innate correspondence knowledge (Casile, Caggiano, & Ferrari, 2011; Ferrari, Bonini, & Fogassi, 2009; Gallese, Gernsbacher, Hickok, Heyes, & Iacoboni, 2011) or innate means to derive such knowledge (Meltzoff & Decety, 2003; Meltzoff & Moore, 1997). Neonatal imitation implies that visual and motor descriptions of actions can be matched by actors who have had little or no opportunity to link them through correlated sensorimotor experience. The most consistent finding is that newborns appear to match observed tongue protrusion (Anisfeld, 1996; Ray & Heyes, 2011). Blind scorers typically report a greater frequency of tongue protrusions when infants observe a model performing the movement, than when infants observe a different modelled action (Abravanel & Deyong, 1991; Abravanel & Sigafoos, 1984; Legerstee, 1991; Meltzoff & Moore, 1977, 1983, 1989).

It remains highly controversial whether tongue protrusion matching by neonates is a product of visuomotor correspondence knowledge. Several studies indicate that tongue protrusion is a generic arousal response elicited by a range of arbitrary stimuli including felt-tip pens moved towards and away from the infant's mouth (Jacobson, 1979) or by flashing lights or bursts of music (Jones, 1996, 2006).

Some authors have also argued that the matching of tongue protrusion by infants reflects an innate releasing mechanism (IRM) (Anisfeld, 1996; Jacobson, 1979). Several innate responses are known to be elicited by seemingly arbitrary stimulus features. For example, the distinctive red spot located on the bright yellow beak of adult herring gulls is thought to be a sign stimulus; the trigger for an innate food begging response in herring gull chicks (Tinbergen, 1953). The sight of a protruding tongue might resemble a similar sign stimulus, possibly the sight of a mother's nipple (Jacobson, 1979). Consistent with this possibility, studies have shown that tongue protrusion matching ceases to occur in infants at around three-months, and reappears around six-months together with a broader range of imitative responses (Abravanel & Sigafoos, 1984; Fontaine, 1984; Jacobson, 1979). Both arousal and IRM accounts also accord with evidence that matching is remarkably robust to stimulus degradation. Comparable tongue-protrusion responses are elicited by a protruding tongue in the context of a whole human face and a robotic tongue protruding from a disembodied robotic mouth (Soussignan, Courtial, Canet, Danon-Apter, & Nadel, 2011).

1.3.2.1.2 Other gestures

Detailed review on a gesture-by-gesture basis reveals reliable evidence of tongue protrusion matching only (Anisfeld, 1996; Ray & Heyes, 2011). Sporadic reports have described neonatal imitation of other gestures, including mouth opening (Legerstee, 1991; Meltzoff & Moore, 1983), hand opening and closing (Vinter, 1986), lip-protrusion (Meltzoff & Moore, 1977; Reissland, 1988), finger movements (Meltzoff & Moore, 1977; Nagy et al., 2005), lateral head movements (Meltzoff & Moore, 1989) and facial expressions of emotion (Field et al., 1983; Field, Woodson, Greenberg, & Cohen, 1982). However, failures to replicate these findings are more numerous than positive results (Ray & Heyes, 2011).

Where observed, reports of imitation of gestures other than tongue protrusion may reflect the particular scoring protocols or methodology employed. For example, mouth opening and lip-protrusion may be side-effects of tongue protrusion matching (Anisfeld, 1991). So-called ‘cross-target’ procedures record the frequency of mouth opening and tongue protrusion, during observation of either tongue protrusion or mouth opening. Where mouth opening responses are suppressed during the observation of tongue protrusion, the frequencies of mouth opening responses may appear elevated during the observation of mouth opening. Similarly, infants are known to track moving objects with their head and eyes. This instinctive ‘perceptual tethering’ behaviour might account for apparent imitation of lateral head movements (Ray & Heyes, 2011). Neonatal imitation of emotional expressions is only evident when scoring techniques are employed which fail to distinguish between imitation and any generic eye-gaze or facial behaviours which vary as a function of observed expression (Kaitz, Meschulachsarfaty, Auerbach, & Eidelman, 1988).

1.3.2.1.3 Improvement data

In addition to the basic tongue matching effect, there have also been reports that newborns’ tongue protrusion responses get progressively closer to the modelled action (Meltzoff & Moore, 1994; Soussignan et al., 2011). One study reported that initial responses to modelled sideways tongue protrusion started as small mid-line protrusions, and thereafter got progressively closer to the modelled action (Meltzoff & Moore, 1994). Similarly, it has been found that initial responses to modelled mid-line tongue protrusion are more often scored as ‘partial’ tongue protrusions whereas subsequent attempts were more frequently deemed ‘full’ tongue protrusions (Abravanel & Sigafoos, 1984; Jacobson, 1979; Soussignan et al., 2011). A similar finding was also reported for neonatal imitation of index finger extension (Nagy et al., 2005), however this is hard to interpret given the unreliability of the basic phenomenon.

Reports of incremental improvement are potentially important because evidence of imitative error-correction argues that neonates are intentionally replicating the topography of observed actions (Meltzoff & Moore, 1997). Reliable evidence of improvement in infant imitation would be hard to reconcile with an IRM or arousal account of tongue protrusion matching. Moreover, such evidence would imply that infants are born with a mechanism which not only allows the detection of disparities between observed and executed actions, but also reveals the nature of the disparity. While these reports are therefore consistent with the AIM hypothesis, and the innate mechanisms it describes, they are extremely challenging for associative solutions to the correspondence problem, which argue that accurate visual feedback is required to refine the accuracy of visuomotor links.

However, the interpretation of these data as evidence of ‘progressive improvement’ has also attracted criticism. Those studies which report incremental improvements in infant imitation all confound accuracy of imitation with response vigour; that is the optimal imitation was also the most exaggerated posture (Meltzoff & Moore, 1994; Nagy et al., 2005; Soussignan et al., 2011). Where repeated execution causes infants’ gestures to become more pronounced, trends may be misinterpreted as imitative error-correction (Ray & Heyes, 2011). The sole report of incremental improvement in imitation of sideways tongue protrusion (Meltzoff & Moore, 1994) also relies on unusual assumptions as to what constitutes monotonic improvement in accuracy (Ray & Heyes, 2011). For example, a large mid-line tongue protrusion was arbitrarily deemed more accurate than a small sideways protrusion.

1.3.2.2 Mirror system responses to visual stimuli that have not been contingently paired with action performance

Evidence from neuroimaging suggests that the MNS responds to visual stimuli which have not been contingently paired with action performance. For example, in a study reported by Buccino et al. (2004) human participants passively observed a human, a monkey, or a dog, either biting or performing oral communicative actions (speaking, lip-smacking or barking, respectively). The observation of biting activated the MNS (inferior parietal and ventral premotor regions) irrespective of the species of the actor. The observation of speech and lip-smacking also activated ventral premotor cortex. However, the observation of barking failed to elicit activation in any areas considered part of the human MNS. These results may indicate that, provided an action can be performed by the observer, a link with the corresponding visual representation may be established in the absence of correlated sensorimotor experience. Consequently, barking – the only action which the human participants could not perform, was the only action to which the MNS did not respond. Because the human participants were unlikely to have had experience of talking while observing monkeys lip-smacking, or biting while seeing dogs biting, direct links between these particular stimuli and the corresponding motor representations are unlikely to have been acquired through correlated sensorimotor experience.

A similar finding was reported by Gazzola, Rizzolatti, Wicker & Keysers (2007). Passive observation of goal-directed hand actions performed by human and robotic agents elicited comparable MNS responses in temporal, frontal and parietal regions for both types of stimuli. If the visuomotor correspondence knowledge represented within the MNS is acquired through correlated sensorimotor experience, one would expect to see stronger responses for naturalistic hand actions than for robotic stimuli.

An associative account can be offered to explain MNS responses to stimuli that have not been paired contingently with action performance, where those stimuli are broadly similar to other stimuli which have been paired with action performance. It is well established that a response originally conditioned to a specific stimulus, may be subsequently observed in the presence of other similar stimuli, in the absence of any additional training – a phenomenon known as stimulus generalisation (Guttman & Kalish, 1956). It is conceivable that the form of the robotic hand employed by Gazzola et al (2007) was sufficiently similar to a human hand to afford stimulus generalisation, and thereby excite direct visuomotor links acquired associatively during correlated sensorimotor experience. Similarly, the kinematics of the dog biting stimuli presented by Buccino et al. (2004) may be sufficiently similar to human biting to elicit stimulus generalisation. In contrast, dog barking may be too dissimilar to human communicative acts to elicit stimulus generalisation, and thus could not gain access to visuomotor links acquired associatively.

Moreover, while traditional fMRI paradigms may reveal that *some* motor representations are active during passive observation, it is difficult to determine *which* motor representations are active. In both of these studies it is unclear whether the motor activations reflect corresponding or complimentary representations. For example, the motor responses excited by the observation of biting might not reflect the motor representations of biting; but rather ‘withdraw’ motor responses. Similarly, passive observation of goal-directed robotic actions may reflect participants preparing to guide manually the grasping being executed by a crude effector toward the target object.

Consequently, it is unclear whether the motor responses observed in these studies required visuomotor correspondence knowledge.

1.3.2.3 Motor-to-visual transfer in the absence of correlated sensorimotor experience

Several findings have been cited as evidence that the motor-system contributes to the topographic description of observed actions - to the representation of the spatial configuration and trajectories of limbs and effectors - in situations where observers have not had correlated sensorimotor experience. This interpretation is challenging for associative accounts as it implies that visuomotor links exist which cannot have been derived through correlated sensorimotor experience.

1.3.2.3.1 Self-recognition effects

It has been reported that participants show superior recognition of their own whole-body movements relative to those of friends (Beardsworth & Buckner, 1981; Jokisch, Daum, & Troje, 2006; Loula, Prasad, Harber, & Shiffrar, 2005; Prasad & Shiffrar, 2009). In these studies, pairs of friends are first filmed performing actions, while dressed in attire and apparatus which allow motion cues to be isolated from form cues. Participants are subsequently shown samples of the recorded movements and required to identify the actors solely from their movement. Reports of superior self-recognition are striking because stimuli depict movements from third-person perspectives; viewpoints from which we rarely view ourselves, but from which we frequently observe friends. If perception was solely determined by visual experience, one would predict the opposite result; that participants would show superior recognition of friends' movements when viewed from third-person perspectives. Several authors therefore regard these counterintuitive self-recognition effects as evidence that the motor system contributes to the topographic description of observed actions (Blake & Shiffrar, 2007; Schutz-Bosbach & Prinz, 2007; M. Wilson & Knoblich, 2005). If this view is correct, such a motor-to-visual transfer cannot be mediated by visuomotor links acquired associatively during correlated sensorimotor experience.

If the motor contribution to perception were mediated by visuomotor links established during correlated sensorimotor experience, one would expect the self-recognition advantage to be even greater for stimuli depicting movements from first-person perspectives. This prediction was tested directly in a study which compared the ability of actors to recognise their own whole body actions (walking, running, dancing) from first and third-person perspectives (Prasad & Shiffrar, 2009). However, actors failed to recognise their own actions or the actions of friends, when point light displays depicted egocentric movements. This finding appears to argue that neither first-person visual experience, nor correlated sensorimotor experience, contribute substantially to self-recognition, as egocentric stimulus presentation should maximise the contribution of these sources of knowledge.

While the self-recognition paradigm is of significant theoretical interest, several methodological issues limit the extent to which existing findings address the origins of visuomotor correspondence knowledge. To study self-recognition from motion cues, techniques must be employed which isolate motion from form. If this process is imperfect, it allows participants to use residual form cues to identify their actions. Several studies of self-recognition (Beardsworth & Buckner, 1981; Cutting & Kozlowski, 1977; Loula et al., 2005; Prasad & Shiffrar, 2009) have employed traditional point-light techniques (Johansson, 1973) whereby actors are filmed against a dark background, whilst wearing dark clothes and small lights attached to their joints. However, this approach inevitably leaves some form cues present. For example, point-lights are typically attached to each shoulder. Thus, one could estimate the breadth of an actor’s shoulders from the distance between these point-lights, and thereby infer their identity.

One of the assumptions underpinning the logic of self-recognition experiments is that the stimuli are naturalistic; it is crucial that they depict accurately how individuals act under normal circumstances. After all, friends only have visual experience of an individual’s typical gait. However, there is ample reason to believe that the stimuli used previously depicted unnatural movements. Feelings of embarrassment elicited by filming, marker or point-light apparatus, or the requirement to perform, may cause unnatural performance. Moreover, the need to wear tight fitting clothes and experimental apparatus may restrict movements.

Studies seeking to assess observers’ ability to recognise self-produced and friends’ motion need to ensure that observers cannot use their episodic recall of the recording phase to achieve recognition. However, actors are likely to be aware of certain aspects of their performance. For example, they might realise that their gait was hurried; that their stride length was shortened; or that they swung their arms less than usual. If actors are able to verbally-code such performance-characteristics they may be subsequently used to identify self-produced movements. This problem may be particularly pronounced in two studies which used infrequently performed actions such as jumping, playing ping-pong, punching a punch bag, and dancing (Loula et al., 2005; Prasad & Shiffrar, 2009). Because actions such as boxing, dancing, and playing table-tennis are performed infrequently, the performance of these actions is likely to be extremely salient. Moreover, because these actions are poorly specified – that is subject to few constraints - performances are likely to be highly idiosyncratic. Interestingly, both of these studies report evidence of superior self-recognition for these infrequent actions, but not for walking (Loula et al., 2005; Prasad & Shiffrar, 2009).

The absence of a self-recognition effect for movements viewed egocentrically (Prasad & Shiffrar, 2009) is difficult to interpret given that neither self- or friend-recognition exceeded chance levels. If self-recognition were mediated by visuomotor links acquired through correlated sensorimotor experience, stronger self-recognition effects should be seen when actions are viewed from first-person perspectives. However, point light displays may be harder to ‘organise’ when viewed egocentrically, not least because observers have to integrate salient motion across three dimensions (across X,Y & Z planes) as opposed to the two dimensional integration (across X & Y planes) necessitated by allocentric presentation. The authors argue this is not the case because, despite being unable to recognise actor identity, participants could nevertheless identify which actions were depicted in first-person point-light displays. However, to the extent that action recognition is an easier attribution than identity recognition, a degraded percept may support the former, but not the latter.

1.3.2.3.2 Effects of non-visual motor training on action perception

It has been reported that practice performing actions improves visual discrimination of those actions, even when actors are prevented from observing their training movements (Casile & Giese, 2006; Hecht, Vogt, & Prinz, 2001). For example, in the study reported by Casile and Giese (2006) actors were blindfolded during a non-visual motor training phase, during which they learned to execute a walking gait where the arm and leg movements had a phase difference of 270º rather than the typical difference of 180º. Gait training was achieved through verbal and haptic feedback provided by the experimenter. Before and after training, participants completed a pre- and post-test where they were required to discriminate point-light stimuli of gaits with phase differences of 180º, 225º and 270º. Not only did the non-visual motor training cause a selective improvement in participants' discrimination of the 270º gait, but the visual recognition performance at post-test was found to correlate with the ability of the participants to perform the actions during the training. This finding is consistent with a motor contribution to perception – that expertise performing actions aids the perception of those actions. Because the actors were blindfolded during the motor training, any contribution from the motor system would appear be mediated by visuomotor links that cannot have been acquired through correlated sensorimotor experience.

However, the effect of the verbal instructions used to guide gait execution during training remains unclear. The authors specify that during training "only verbal and haptic feedback was provided about the appropriateness of the adopted postures and smoothness of the movements" (Casile & Giese, 2006, p73). Participants may have used the verbal instructions to form a series of verbal propositions describing each gait, for example "when one arm is at its forward peak, the other should be passing the adjacent leg." The use of such verbal strategies could explain the correlation between training and recognition performance - a detailed verbal description could improve both gait performance and gait recognition.

In addition, it is also unclear the extent to which the training enhanced the temporal or topographic description of the observed gait. Being blindfolded prevented the actors from benefitting from first-person visual experience of the trained gaits. However, actors had access to other sensory information during training in the form of proprioceptive, auditory, vestibular and somatosensory feedback. All of these sensory signals contain information about the temporal characteristics of the trained gait. It is possible that this temporal experience improved perception and discrimination of the rhythm of the gait, rather than aid its topographic description. The authors argue that this is not the case because, in an auxiliary experiment, periods of visual-only training where participants studied the trained gait, failed to improve discrimination of arbitrary rhythmic stimuli. However, given the considerable differences between the two forms of training (visual-only and motor-only) and the test stimuli used (point-light walking displays and alternating squares) it seems little can be inferred about the mechanism of the main effect.

1.4 Empirical chapters

1.4.1 Summary

This literature review began by posing a question; how do actors solve the visuomotor correspondence problem? Several putative solutions were then discussed. Associative theories were described which posit that visuomotor correspondence knowledge can be acquired only through correlated sensorimotor experience. The ASL model proposes that experience of this kind creates visuomotor associations through the same learning processes that mediate Pavlovian and instrumental conditioning in the laboratory. As its name suggests, Hebbian MN theory argues that visuomotor links are established though Hebbian learning. In contrast, two nativist theories have also been proposed which argue that correlated sensorimotor experience is not necessary to solve the correspondence problem. The AIM hypothesis proposes that we possess an innate means of matching the ‘seen but unfelt’ to the ‘felt but unseen’ supramodal representations. Innate MN theory proposes that MNs constitute an innate form of correspondence knowledge and that they are an adaptation for action understanding and imitation.

Considerable evidence indicating that correlated sensorimotor experience plays a crucial role in the acquisition of visuomotor correspondence knowledge has amassed from (i) neuroimaging studies of sensorimotor expertise; (ii) sensorimotor training studies; and (iii) neurophysiological studies of macaque MNs. Such evidence is therefore consistent with associative theories such as the ASL model and Hebbian MN theory. Nevertheless, evidence of (i) neonatal imitation; (ii) MNS responses to stimuli that have not been contingently paired with action performance; and (iii) motor to visual transfer in the absence of correlated sensorimotor experience, remains challenging for these theories.

1.4.2 Thesis plan

Despite the extensive research undertaken in this area, the nature and origins of visuomotor correspondence knowledge remain contentious. The aim of the work reported in this thesis was to contribute to this debate in two respects: First, it sought to address previous reports that appear to challenge associative accounts of the acquisition of visuomotor correspondence knowledge; namely evidence of motor-to-visual transfer in the absence of correlated sensorimotor experience and imitative improvement in the absence of visual feedback. These are regarded as the most challenging lines of evidence against associative accounts, given that reports of MNS responses to stimuli that have not been paired with action execution may reflect stimulus generalisation or excitation of complimentary responses (section 1.3.2.2). Second, it sought to determine which of the two associative accounts, ASL or Hebbian MN theory, best describes the learning responsible for the acquisition of correspondence knowledge.

The aim of the experiments reported in Chapter 2 was to determine what types of cues mediate superior recognition of self-produced actions from third-person perspectives (section 1.3.2.3.1). Existing accounts of these effects argue that the self-recognition advantage reflects a transfer of topographic information from the motor- to visual-systems. This account requires visuomotor links that are unlikely to have been acquired through correlated sensorimotor experience, and is therefore inconsistent with associative solutions to the correspondence problem. However, this account remains speculative – in practice little is known about the cues mediating the self-recognition advantage. A number of methodological issues raise the possibility that these effects may not reflect a motor contribution to perception, and may instead be due to recall of episodic cues from the recording context. Equally, it remains unknown whether knowledge of temporal action characteristics, which *can* be acquired through correlated sensorimotor experience, also contributes to self-recognition.

Chapter 3 tested whether adults can improve the accuracy of their imitative performance in the absence of contingent visual feedback (see section 1.3.2.1.3). According to the AIM hypothesis, humans are endowed with an innate mechanism which enables us to identify visual-proprioceptive correspondences. This mechanism is consistent with reports of incremental improvement in neonatal imitation, in the absence of visual feedback. In contrast, these findings are challenging for associative solutions to the correspondence problem, which posit that accurate visuomotor links can be established only through the provision of accurate third-person visual feedback. If the AIM hypothesis is correct, adults reliant solely on proprioceptive feedback should also be able to achieve closer and closer correspondence between observed and executed action. However, if reports of improvements in neonatal imitation rely on subjective scoring techniques, or reflect short-term learning effects within the infants' peripheral motor system, adults should show no improvement without visual feedback.

Having found little support for the nativist solutions to the correspondence problem in the first two empirical chapters, Chapters 4 and 5 sought to determine whether the learning mediating the acquisition of visuomotor links is Hebbian (as proposed by Hebbian MN theory - section 1.2.1.3) or associative (as proposed by the ASL model – section 1.2.1.1). Chapter 4 tests whether the acquisition of visuomotor links is sensitive to both sensorimotor contiguity and contingency, as predicted by the associative account, or just temporal contiguity, as predicted by the Hebbian account. Chapter 5 sought to test a further prediction of the ASL model; that learning which establishes rival visuomotor links, between the sight of an action and the execution of another action, should show context-specificity.

Chapter 2: Self-recognition is mediated by temporal, not topographic, cues

2.1 Introduction

People are better at recognizing their own walking gaits and whole body movements than those of friends, even when stimuli are viewed from third-person perspectives (Beardsworth & Buckner, 1981; Jokisch et al., 2006; Loula et al., 2005; Prasad & Shiffrar, 2009). For example, Loula et al. (2005) filmed three pairs of friends, dressed in tight black clothes and with 13 reflective markers attached to their joints, while they performed a variety of actions. Two-to-three months after filming participants were recalled and required to identify whether the actors depicted in the resulting point light displays were themselves, a friend or a stranger, in a 3-alternative-forced-choice (3-AFC) task. The authors found that participants successfully identified themselves (69%) more often than their friends (47%). Analysis of performance by action type suggests that participants were particularly good at recognising self-produced dancing, boxing, ping-pong, and jumping movements, but failed to recognise self-produced walking, running, laughing, and greeting.

This self-recognition advantage is surprising because whole body movements are ‘perceptually opaque’ (Heyes & Ray, 2000); they cannot be viewed directly by the actor from third person perspectives. While we sometimes view our movements in mirrors or in video recordings, it is likely that we see our friends’ movements from third-person perspectives more often than our own. Therefore, if action perception depended solely on visual experience (Bulthoff, Bulthoff, & Sinha, 1998; Giese & Poggio, 2003; Johansson, 1973; O'Toole, Roark, & Abdi, 2002), one would expect the opposite result - superior recognition of friends’ movements when viewed from third-person perspectives. However, while we lack visual experience of our own actions, we have extensive knowledge of executing our own actions. Superior self-recognition is therefore widely cited as evidence that the motor system contributes to action perception (Blake & Shiffrar, 2007; Blakemore & Frith, 2005; Schutz-Bosbach & Prinz, 2007; M. Wilson & Knoblich, 2005); that repeated performance of an action makes that action easier to recognize when viewed from the outside.

Existing accounts of the self-recognition advantage argue that the motor system aids perception of limb configurations through a prediction process. Actors are thought to simulate covertly observed actions within their motor system, and thereby predict future limb positions and trajectories of stimuli viewed from third-person perspectives (Blake

& Shiffrar, 2007; M. Wilson & Knoblich, 2005; Wolpert et al., 2003). For example, Wilson and Knoblich (2005) propose that:

> "The various brain areas involved in translating perceived human movement into corresponding motor programs collectively act as an emulator, internally simulating the ongoing perceived movement. This emulator bypasses the delay of sensory transmission to provide immediate information about the ongoing course of the observed action as well as its probable immediate future. Such internal modeling allows the perceiver to rapidly interpret the perceptual signal, to react quickly, to disambiguate in situations of uncertainty, and to perceptually complete movements that are not perceived in their entirety (Wilson & Knoblich, 2005, p468)."

When viewing self-produced motion, these perceptual predictions are most accurate and thus yield a self-recognition advantage (Knoblich & Flach, 2003; Schutz-Bosbach & Prinz, 2007; M. Wilson & Knoblich, 2005).

According to this account, the cues being transferred from the motor- to perceptual-system are topographic; relating to the precise spatial configuration of limb positions and trajectories. Such a motor contribution must be mediated by an endogenous motor-to-visual transfer process. In the absence of correlated sensorimotor experience third-person perceptual predictions need to be derived or extrapolated from 'first-person' motor experience. This ability would be inconsistent with associative solutions to the correspondence problem. For perceptually transparent self-produced actions, such as hand movements, third person visual descriptions might be derived from first-person visual experience via a mental rotation process. This may reflect a form of stimulus generalisation (see section 1.3.2.2). However, for perceptually opaque actions, such as facial expressions and whole-body actions, a rotation or generalisation process is unlikely because of the profound differences between first- and third-person visual descriptions. In order for actors to somehow 'work out' what their actions look like from the outside, in the absence of correlated sensorimotor experience, they would need to be endowed with innate correspondence knowledge (Rizzolatti & Craighero, 2004; Rizzolatti et al., 1996) or an innate ability to derive such knowledge (Meltzoff & Moore, 1997). Only the existence of i) innate links between motor programs and third-person visual representations or ii) an endogenous translation mechanism, would allow actors to extrapolate allocentric visual descriptions from egocentric motor experience, in the absence of correlated sensorimotor experience.

However, the topographic predication account remains speculative. An equally plausible alternative is that self-recognition is mediated by recognition of temporal cues. The repeated performance of actions is likely to give actors knowledge of the temporal characteristics of their actions, such as the frequency or rhythm of particular action segments. Numerous sensory signals carry information about the temporal characteristics of our actions, including first-person visual, proprioceptive, somatosensory, auditory, and vestibular feedback. A lifetime of exposure to these signals, may allow subtle temporal characteristics to be identified (e.g. “my head movements are smooth and constant”) and thereby compared with the characteristics of third-person displays. The importance of temporal cues is suggested by studies showing that participants can recognize their own clapping both from degraded visual stimuli depicting just two point-lights (Sevdalis & Keller, 2010) and from simple auditory tones matched to the temporal structure of actions (Flach, Knoblich, & Prinz, 2004). It has also been reported that velocity characteristics, including rates of acceleration and deceleration are essential for the self-recognition of kinematic displays of drawing (Knoblich & Prinz, 2001). Crucially, because the recognition of temporal characteristics does not require actors to infer third person visual descriptions of their own movements from first-person motor experience, this account is entirely consistent with associative solutions to the correspondence problem.

In addition, a number of methodological issues present within the existing reports raise the possibility that superior self-recognition is mediated, not by a contribution from the motor system, but by actors’ knowledge of their physical proportions or by episodic recall of the recording phase. Previous studies comparing recognition of self-produced and friends’ actions have employed traditional point-light methods (Johansson, 1973) to isolate motion cues (Beardsworth & Buckner, 1981; Cutting & Kozlowski, 1977; Loula et al., 2005; Prasad & Shiffrar, 2009). This technique is poorly suited to the study of self-recognition because point-light stimuli contain residual form cues indicating the actor’s build and, due to the unusual apparatus employed during filming, necessarily depict unnatural, idiosyncratic movements. In addition, little has been done to prevent participants encoding and recalling performance characteristics from the filming stage (e.g. “did I swing my arms less than usual?”), and using these episodic memory cues to identify their movements during the test phase.

The experiments reported in this chapter sought to test the existing account of the self-recognition advantage; that the motor system contributes topographic information via an endogenous mechanism. In particular, they sought to eliminate two alternative accounts: First, that previously reported self-recognition effects are methodological artefacts. Second, that superior self-recognition is mediated by temporal, and not topographic, cues.

To determine whether the self-recognition advantage remains when procedures are adopted to address the methodological problems present in previous reports, the experiments reported employ a markerless facial avatar technique. An avatar is a digital mannequin; a common form that may be animated with the motion derived from different actors. Avatar techniques are preferable to point-light approaches because they completely eliminate form cues (Daprati, Wriessnegger, & Lacquaniti, 2007; Hill & Johnston, 2001). Moreover, because the present technique does not require individuals to wear markers or point-light apparatus during filming, it is also better able to capture naturalistic motion than the methods used previously. Having identified a self-recognition advantage for inverted facial motion in Experiment 1, Experiment 2 sought to identify what type of information plays the crucial role in self-recognition. To this end, avatar stimuli were subject to manipulations which selectively disrupted either the topographic or temporal stimulus characteristics.

2.2 Experiment 1

The principal aim of Experiment 1 was to determine whether the self-recognition effects previously reported with whole-body actions reflect artefacts of the methodologies employed. To address this concern, Experiment 1 sought evidence of superior self-recognition for another form of perceptually opaque action – facial motion – while overcoming the methodological problems identified in previous studies of self-recognition.

In addition, the experiments sought to determine whether the self-recognition advantage varies with the orientation of the facial stimuli. The recognition of familiar others, from both static (Yin, 1969) and dynamic (Knight & Johnston, 1997) facial displays is known to be impaired by inversion. This effect is thought to be due to the disruption of ‘configural’ or ‘holistic’ cues (Maurer, Le Grand, & Mondloch, 2002; Tanaka & Farah, 1993; Young, Hellawell, & Hay, 1987). If the recognition of self-produced facial motion is mediated by configural topographic information it was reasoned that the self-recognition advantage should be greater for upright facial stimuli. If however, self-recognition ability is mediated by temporal cues extracted from local features e.g. rhythmic mouth or head movements, the self-recognition advantage should be greater for inverted stimuli.

2.2.1 Methods

2.2.1.1 Participants

Participants were 12 students (4 male, mean age = 23.2 years) from the University of London comprising six same-sex friend pairs. Friends were defined as individuals of the same sex, who had spent a minimum of 10 hours a week together during the 12 months immediately prior to the experiment (Loula et al., 2005). Participants were of approximately the same ages and physical proportions. All were right handed, had normal or corrected-to-normal vision, and were naive to the purpose of the experiment. Participants received a small honorarium for taking part. None of the participants had taken part in experiments described elsewhere in this thesis.

2.2.1.2 Stimuli

Five-to-six months prior to testing, each member of the friendship pairs was filmed individually while recalling and reciting question and answer jokes (Hill & Johnston, 2001). The demands of this task – to recite the jokes from memory, while aiming to sound as natural as possible – drew the participants’ attention away from their visual appearance. These driver sequences were filmed using a digital Sony video camera at 25 frames per second (FPS).

Avatar stimuli were produced from this footage using the Cowe Photorealistic Avatar technique (Berisha, Johnston, & McOwan, 2010; Cowe, 2003). Principle components analysis (PCA) is used to extract a multidimensional space from the image variation present within a given sequence of images. This allows any frame within that sequence to be represented as a mean-relative vector within a multi-dimensional space. If a frame vector from one sequence is projected into the space derived from another actor, a 'driver' expression from one individual may be projected on to the face of another individual (Figure 4). If this is done for an entire sequence of frames, it is possible to animate an avatar face with the motion derived from another actor. This technique was used to project the motion extracted from each actor's sequences onto an average androgynous head.

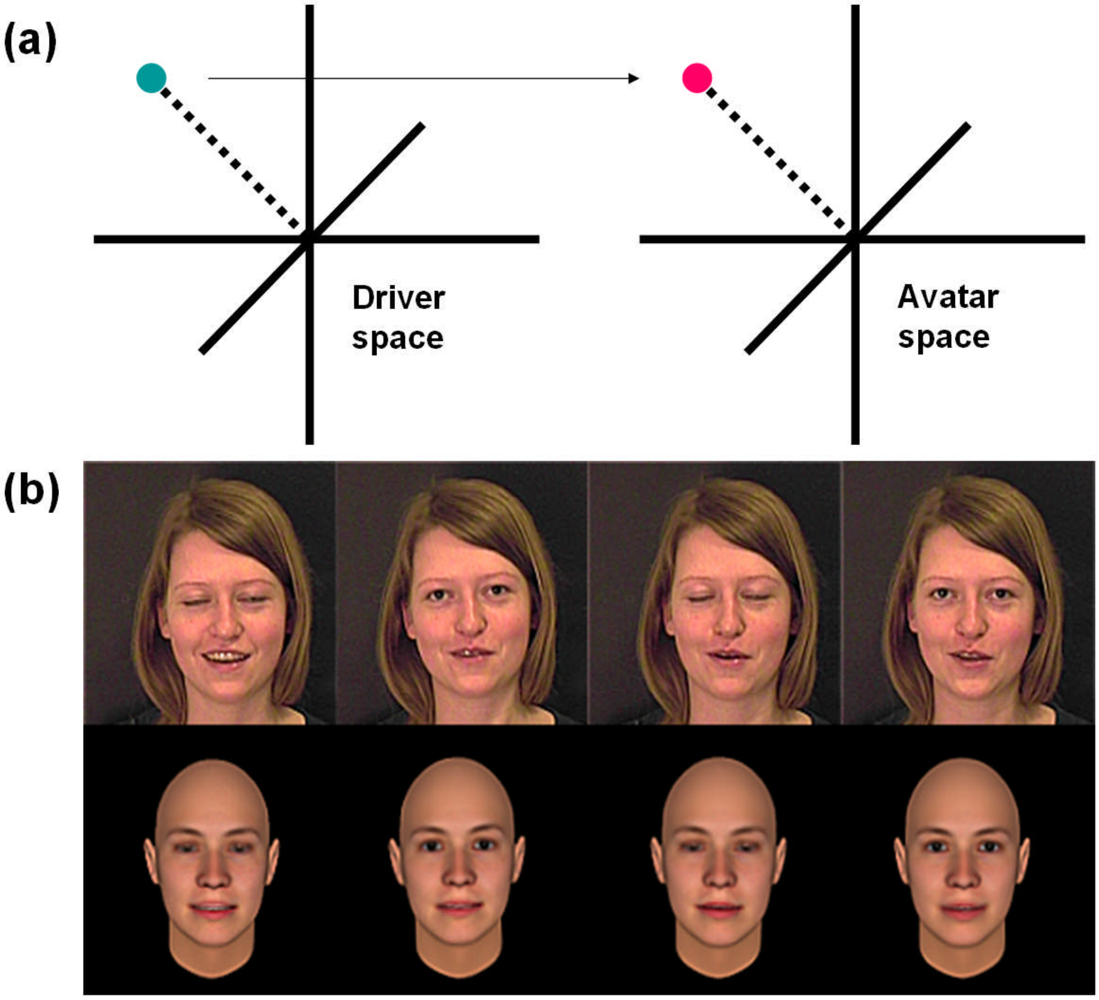

Figure 4: (a) Schematic representation of the animation process employed in the Cowe Photorealistic Avatar procedure. (b) Examples of driver frames (top) and the resulting avatar frames (bottom) when the driver vector is projected into the avatar space.

The avatar space was constructed from a cohort of 721 still images derived from Singular Inversions' FaceGen Modeller 3.0 by placing an approximately average, androgynous head in a variety of poses. These poses sampled the natural range of rigid and non-rigid facial motion, but were not explicitly matched to real images. The resulting image set included mouth variation associated with speech, variations of eye gaze, eye aperture, eyebrow position and blinking, variation of horizontal and vertical head position, head orientation, and apparent distance from camera. The avatar images were cropped and resized to 135 x 135 pixels.

Each image from the avatar cohort was first described in terms of its posture deviation from the cohort average, as well as its texture characteristics, using a biologically plausible optic flow algorithm (Berisha et al., 2010; Cowe, 2003; Johnston, McOwan, & Benton, 1999; Johnston, McOwan, & Buxton, 1992). A single frame was chosen to act as the reference. For the remaining 720 (n-1) frames, displacement fields were calculated which warped the facial features back to their position in the reference image, on a pixel-by-pixel basis. By averaging these flow-fields, the mean warp was calculated, which in turn was applied to the reference frame to derive the mean face shape. The displacement-fields were then adjusted so that they registered to the mean face shape rather than the original reference. The resulting fields described the variation in face shape present within the avatar cohort.

However, texture deviations reflecting changes in lighting or shadow, or iconic changes whereby a feature (e.g. teeth) is present in the target, but not in the mean face, or vice-versa, are not described within the displacement-field variance. In order to represent the texture deviations it was necessary to feature align all frames. Because the displacement fields had been adjusted to register to the warp mean rather than the original reference, they could be used to warp each frame such that the features were aligned in the mean position. Texture deviations were encoded by a Red-Green-Blue (RGB) triplet reflecting the texture of a given point on the mean face shape for any given frame. These values did not represent departures from the mean texture; they represented the colour of a pixel when that frame was feature aligned.

Thus for every pixel of every frame a five element vector was derived, describing optic flow deviation from the mean in the x and y dimensions, and an RGB triplet describing the texture variation for each point on the mean face shape. For each frame, these pixel vectors were concatenated to produce 721 image vectors of length 5 x *image width* x *image height*, to which PCA was then applied.

Having built the avatar space, it was then necessary to derive driver spaces from the footage of participants' facial motion. Suitable segments for stimulus generation were defined as sections of 92 frames (3.7 secs) containing salient facial motion, and in which the participant's gaze was predominantly fixated on the viewer. The majority of clips contained both rigid and non-rigid facial motion. Facial speech was also present in most, but exceptions were made when other salient non-rigid motion was evident. The driver frames were also cropped and resized to 135 x 135 pixels.

Driver spaces were created for each 92-frame sequence through the same process used to generate the avatar space (described above). Having derived 14 driver spaces for each participant, these were converted into 3.7 second avatar stimuli by projecting each of the 92 frame vectors of the driver sequence into the avatar space, and converting the resulting vector into movie frames (Figure 4). The resulting avatar stimuli were saved and presented in uncompressed AVI formats.

2.2.1.3 Procedure

Five to six months after the recording phase, participants were recalled to complete a 3-AFC recognition test. Testing was completed in a dimly lit cubicle. Each trial presented a single avatar stimulus, in an upright or inverted orientation, and required participants to indicate whether the motion used to animate the head had been taken from themselves, their friend or a stranger. The stimuli derived from each actor appeared once as 'self', once as 'friend' and once as 'stranger'. The experiment was completed over two sessions: In session 1, participants completed a block of upright trials followed by an inverted block; in session 2 block order was reversed. Different strangers were allocated across the first and second sessions to ensure that effects were not artefacts of the particular stranger allocations.

Trials began with a fixation dot presented for 750ms, followed by an 3.7 second avatar stimulus looped to play twice. Following stimulus offset "Self, Friend, or Other?" appeared at the display centre. Participants were required to press S, F or O keys to record their judgment. No feedback was provided during the experiment. Participants were informed that trial order was randomized, but a third of trials would present their own motion, a third the motion of their friend, and a third the motion of a stranger. Each stimulus was presented twice, making a total of 84 trials per block. Participants were seated at a viewing distance of approximately 60cm. The avatar head subtended 6° x 4° of visual angle.

Testing for Experiment 1 commenced 5-6 months after filming. The delay was longer than that typically imposed in studies of self-recognition (Beardsworth & Buckner, 1981; Jokisch et al., 2006; Loula et al., 2005; Prasad & Shiffrar, 2009) to minimise any risk that test performance would be influenced by episodic recall of idiosyncratic movements made during filming. As a further precaution, participants were informed only a few minutes prior to testing that they would be required to discriminate their own motion. These steps, together with the measures taken to prevent encoding of idiosyncrasies during filming, ensured that the effects observed were due to recognition of actors' motion signatures and not attributable to episodic recall of the filming session.

2.2.2 Results & Discussion

2.2.2.1 Data handling

For each condition, d-prime (*d'*) statistics were calculated to measure participants' ability to discriminate self-produced and friends' motion from the motion of strangers (Macmillan & Creelman, 1991). The analyses reported were conducted on the resulting distributions of *d*-prime values. Hits were correct identifications (self response to self stimulus / friend response to a friend stimulus) whereas false alarms were incorrect judgements of the stranger stimuli (self-response to stranger stimulus / friend response to stranger stimulus). The *d*-prime values were calculated by subtracting the normalised false alarm probabilities from the normalised hit probabilities. It was necessary to exclude self-responses to friend-stimuli and friend responses to self-stimuli in order to ensure the distributions of d-primes for self- and friend discrimination were independent. Including these trials in the analyses would introduce an inevitable dependency between the self and friend *d*-primes. For example, a friend-response to a self-stimulus detracts from the self sensitivity by being scored as a miss instead of a hit, but also from the friend sensitivity by increasing the false alarm rate. Although the approach outlined above excludes some trials from the analyses, it is necessary to ensures that the self-and friend distributions are free to vary independently.

2.2.2.2 Results

The mean *d*-primes from Experiment 1 are shown in Figure 5a. One-sample t-tests were used to determine whether the distributions of *d*-primes observed in each condition differed significantly from chance. Participants were able to successfully discriminate their own motion both in upright (*M* = .49, *SD* = .52) [*t*(11) = 3.25, *p* = .008] and inverted (*M* = .47, *SD* = .37) [*t*(11) = 4.34, *p* = .001] orientations, as well as their friends' motion when presented upright (*M* = .37, *SD* = .33) [*t*(11) = 3.95, *p* = .002]. However, recognition of friends' motion failed to exceed chance levels when stimuli were inverted (*M* = .11, *SD* = .22) [*t*(11) = 1.69, *p* > .10].

A mixed model ANOVA with actor (self, friend) and orientation (upright, inverted) as within subjects factors, revealed a significant main effect of orientation [*F*(1,11) = 6.25, *p* < .05, η^2 = .36]. However, the main effect of actor [F(1,11) = 2.83, p > .10, η^2 = .21] and the actor x orientation interaction [*F*(1,11) = 2.40, *p* > .10, η^2 = .18] were non-significant. Simple effects analysis (warranted by Howell, 1996, p.415) yielded evidence of superior self-recognition only when stimuli were inverted [*t*(11) = 2.84, *p* = .016]. When stimuli were presented upright, discrimination of self-produced and friends' motion were comparable [*t*(11) = .62, *p* > .50]. Whereas friend-recognition was substantially impaired by inversion [*t*(11) = 2.84, *p* = .016], self-recognition was not [*t*(11) = .24, *p* > .80].

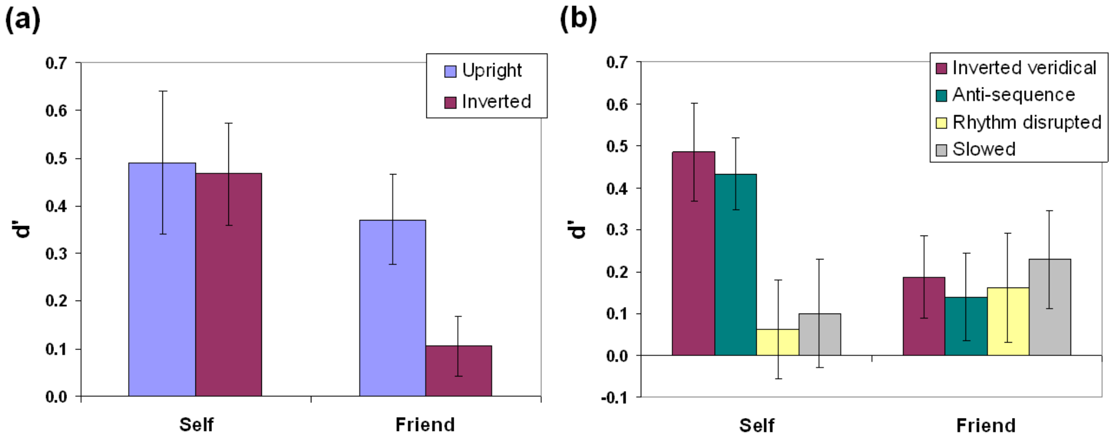

Figure 5: (a) Results from Experiment 1. Whereas discrimination of friends' motion showed a marked inversion effect, participants' ability to discriminate self-produced motion was insensitive to inversion. (b) Results from Experiment 2. Error bars denote standard error of the mean in both panels.

2.2.2.3 Discussion

Despite the absence of a significant main effect of actor, these results provide some interesting insights into the cues mediating self-recognition. These findings indicate that people are able to recognize their own facial motion under remarkably cryptic conditions - when it is mapped on to an inverted average synthetic head. They also indicate that, under these conditions, self-recognition is superior to friend recognition.

Stimulus inversion impaired friend recognition but not self-recognition, suggesting that people are not only better at self-recognition, but that they use different cues to identify self-produced and friends' motion. While configural topographic cues, known to be disrupted by inversion (Maurer et al., 2002; Tanaka & Farah, 1993; Young et al., 1987), may be necessary for the recognition of familiar others, such cues may play a less significant role, if any, in self-recognition.

2.3 Experiment 2

Having identified a self-recognition advantage for inverted facial motion in Experiment 1, the aim of Experiment 2 was to ascertain whether temporal or topographic cues mediate this advantage. The same participants completed the 3-AFC test while viewing inverted stimuli under three additional conditions: anti-sequence, rhythm disrupted and slowed. In the anti-sequence condition, stimuli were transformed in a way that selectively disrupted their topographic properties, whereas the rhythm disrupted and slowed manipulations disrupted the temporal characteristics of the avatar stimuli.

2.3.1 Methods

2.3.1.1 Participants

The same six friend pairs who participated in Experiment 1 also took part in Experiment 2 in return for a small honorarium.

2.3.1.2 Stimuli

For each actor, 14 anti-frame sequences were created which depicted the 'mirror' trajectory through avatar space (Figure 6). For a given frame, the corresponding anti-frame is the equivalent vector projected into the opposite side of the avatar space. Thus, each anti-frame was derived by multiplying the vector describing each veridical frame (from the original unaltered sequence) by -1. Because frames and anti-frames are equidistant from the mean avatar posture, sequences of anti-frames preserve the relative magnitude and velocity of the changes in expression space over time, but reverse the direction of the rigid and non-rigid changes, radically distorting their appearance. For example, a frame in which an actor is raising their eye-brows, pronouncing the phoneme /ooh/ and tilting their head to the front-right, becomes an anti-frame where the actor is frowning, pronouncing the phoneme /ee/ and tilting their head backwards towards the left. As a result the topographic cues contained within a sequence are grossly distorted, while leaving the temporal and rhythmic structure intact. It was anticipated that participants, who were naïve to the nature of the manipulation, would be unable to recover from the anti-sequences the topographic characteristics of an individual's facial motion.

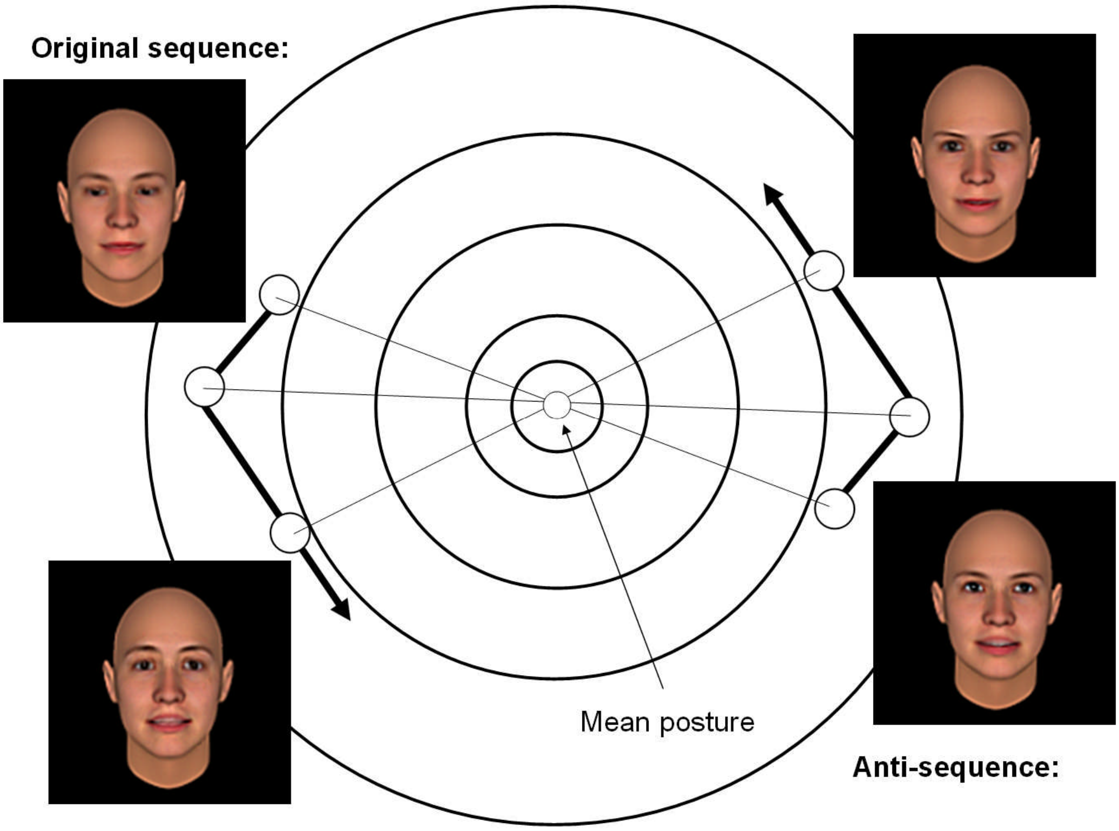

Figure 6: Schematic representation of three frames and their corresponding anti-frames within avatar space. Anti-frames are derived by projecting a veridical frame vector into the diametrically opposite side of the avatar space, across the mean posture.

Fourteen rhythm disrupted stimuli were also produced for each actor, by inserting 46 pairs of interpolated frames between 50% of the original frame transitions. Interpolations were calculated based on the loadings on each of the avatar components. If, for example, an original transition consisted of a change from frame n with a loading of 400 on PC1 to frame n+1 with a loading of 700, two interpolated frames would be inserted with loadings of 500 and 600. Runs of interpolated transitions were encouraged by biasing the decision to interpolate (chance +/- 25%) contingent on whether frames had or had not been inserted on the previous transition. The resulting 184 frames (twice the original number) were converted into uncompressed AVI movie files using Matlab and played at 50 FPS (twice the original rate). Inserting pairs of interpolated frames at half the transitions, and playing the rhythm disrupted stimuli at twice the original frame rate, ensured that they were of the same duration as the veridical stimuli. Moreover, biasing the insertions so that they clustered together ensured salient rhythmic disruption: Segments containing frequent interpolations appeared slower than the veridical; segments with few insertions appeared faster than the veridical.

A further set of 14 slowed stimuli were created for each actor by slowing the veridical stimuli by one of seven parameters applied consistently throughout the sequence. The resultant stimulus durations ranged from 120% of veridical to 180% in 10% intervals.

2.3.1.3 Procedure

Experiment 2 was completed in a single session, conducted 10-11 months after filming. The testing session comprised three blocks of 84 trials. Within each block, the stimuli presented were all subject to the same manipulation. The order in which blocks were completed was counterbalanced across participants. Each manipulated stimulus was presented twice in a randomised order. As in Experiment 1, trials presented a single stimulus centrally, looped to play twice. Participants were then required to make a 3-AFC recognition judgement, identifying the actor whose motion was used to animate the avatar. The stranger allocations were identical to those employed during the second session of Experiment 1. Data from the inverted condition in this session, where the stimuli were veridical rather than temporally or spatially distorted, were used for comparison with the results of Experiment 2 (Figure 5b).

For each condition, *d*-prime statistics were calculated to measure participants' ability to discriminate self-produced and friends' motion from the motion of strangers. Hits were again correct identifications (self response to self stimulus / friend response to a friend stimulus) whereas false alarms were incorrect judgements of the stranger stimuli (self-response to stranger stimulus / friend response to stranger stimulus). As in Experiment 1, the analyses reported were conducted on the resulting distributions of *d*-prime values.

2.3.2 Results & Discussion

The mean *d*-primes from Experiment 2 are shown in Figure 5b. The data were subject to ANOVA with actor (self, friend) and sequence manipulation (anti-sequence, rhythm disruption, slowing) as within subjects factors, which revealed a significant actor x manipulation interaction [$F(2,22) = 4.26$; $p < .05$, $\eta^2 = .28$]. To better understand this overall interaction, three further simple interaction analyses were conducted, revealing a marginally significant actor x manipulation (anti-sequence, rhythm disruption) interaction [*F*(1,11) = 4.66; *p* = .054, η^2 = .30] and a significant actor x manipulation (anti-sequence, slowed) interaction [*F*(1,11) = 9.58; *p* < .025, η^2 = .47], but not a significant actor x manipulation (rhythmic, slowed) interaction [*F*(1,11) = .80; *p* > .60, η^2 = .47]. These results suggest that the ability of participants to recognise stimuli subject to the anti-sequence manipulation, varied disproportionately as a function of actor, relative to the stimuli subject to the other manipulations.

This interpretation was confirmed through the application of paired and one-sample t-tests. If self-recognition is mediated by topographic cues one would expect participants to be unable to recognize themselves in the anti-sequence condition. However, despite the profound changes to the rigid and non-rigid topographic cues, a marginally significant self-recognition advantage was again observed [*t*(11) = 2.17; *p* = .053], replicating that seen in Experiment 1. Participants continued to show better than chance discrimination of their own motion (*M* = .43, *SD* = .30) [*t*(11) = 5.04, *p* < .001], comparable with the inverted veridical condition from the second session of Experiment 1 (*M* = .48, *SD* = .41) [*t*(11) = .50, *p* > .60], whereas friend recognition failed to exceed chance levels (*M* = .14, *SD* = .36) [*t*(11) = 1.33; *p* > .20]. Thus, participants continued to recognise their own motion when the feature trajectories and configurations were grossly distorted, suggesting that self-recognition does not rely on the identification of familiar topographic cues.

In contrast, changes to the temporal properties of the stimuli eliminated the self-recognition advantage, and reduced recognition of self-produced motion to chance levels. Participants could no longer discriminate their own motion in either the rhythm-disrupted (*M* = .06, *SD* = .40) [*t*(11) = .53; *p* > .60] or the slowed (*M* = .10, *SD* = .45) [*t*(11) = .76; *p* > .40] conditions. Self-recognition under both rhythm-disrupted [*t*(11) = 3.15, *p* = .009] and slowed [*t*(11) = 2.48, *p* = .031] conditions was poorer than under veridical conditions. These findings indicate that self-recognition is not mediated by cues such as frequency of eye-blinks or gross head movements, which are unaffected by rhythmic disruption and slowing. Taken together, the results of Experiment 2 suggest that self-recognition depends on the temporal characteristics of local motion. Friend recognition again failed to exceed chance levels in either rhythm disrupted (*M* = .16, *SD* = .45) [*t*(11) = 1.22; *p* > .20] or slowed manipulations (*M* = .23, *SD* = .40) [*t*(11) = 1.97;

p > .07]. This is not surprising given that participants could not discriminate friends' inverted veridical motion.

2.4 General Discussion

Two experiments have been described which sought to determine whether the self-recognition advantage extends from whole body movements such as walking gait, to another form of perceptually opaque actions - facial motion, and if so, to identify which cues mediate the effect. Experiment 1 failed to reveal a self-recognition advantage when stimulus orientation was upright: Recognition of both self-produced and friends' motion was above chance, and broadly equivalent. However, a self-recognition advantage was present when stimulus orientation was inverted. While the discrimination of friends' motion showed evidence of the typical face inversion effect, the discrimination of self-produced motion was relatively insensitive to inversion. Experiment 2 sought to identify whether such orientation insensitive self-recognition was mediated by temporal cues, topographic cues, or both. To this end, manipulations were employed that selectively impaired either temporal (slowing or rhythmic disruption) or topographic (anti-sequencing) cues. Discrimination of inverted self-produced motion was impaired only by stimulus manipulations which altered the temporal properties of the stimuli.

The results of Experiment 1 suggest that self-recognition may be more reliant than friend-recognition, on local motion cues. Inversion of faces is thought to impair perception by disrupting configural representation (Maurer et al., 2002; Tanaka & Farah, 1993; Young et al., 1987). That discrimination of friends' motion was impaired by inversion therefore suggests that configural 'motion signatures' (O'Toole et al., 2002), integrated from multiple features across the face, mediate the recognition of familiar others. In contrast, participants' ability to recognize their own motion was found to be insensitive to inversion. Recognition of self-produced motion may therefore depend on local cues derived from particular features. However, in the absence of a significant actor x orientation interaction, this remains a speculative suggestion.

The results of Experiment 2 suggest that recognition of self-produced motion is mediated by temporal information. Such cues might include the rhythmic structure afforded by the onsets and offsets of motion segments and characteristic variations in feature velocities. That self-recognition depends on temporal cues is consistent with previous reports of a self-recognition advantage for highly rhythmic actions such as walking (Beardsworth & Buckner, 1981; Jokisch et al., 2006; Prasad & Shiffrar, 2009). It is also consistent with the observation that participants cannot accurately discriminate self-produced and friends' motion when the stimuli depict walking or running on a treadmill (Loula et al., 2005). The artificial tempo imposed by a treadmill reduces natural variation in the temporal properties that define an individual's gait.

Existing accounts attribute the self-recognition advantage to an endogenous transfer of topographic information from the motor- to perceptual-systems. The covert simulation of actions within the motor system is thought to enable superior prediction of limb positions or configurations (Prasad & Shiffrar, 2009; Schutz-Bosbach & Prinz, 2007; M. Wilson & Knoblich, 2005). This account is inconsistent with associative solutions to the visuomotor correspondence problem, as it relies on actors being able to derive a third-person visual description from first-person motor experience. We rarely see our own actions from third-person perspectives. Therefore we have little opportunity to learn what our bodies look like from the outside as we act. Since such visuomotor correspondence knowledge is unlikely to be learned through correlated experience of observing and executing the same action, it would have to be inferred; a complex but unspecified process would be needed to generate view-independent visual representations of actions from motor programmes.

However, that self-recognition depends on temporal rather than topographic cues, indicates that an endogenous process is unnecessary. We have the opportunity to learn the temporal signatures of our actions via first-person visual, proprioceptive, somatosensory and vestibular feedback. Consequently, whereas self-recognition based on an endogenous transfer of topographic information is inconsistent with associative accounts, self-recognition based on temporal cues is not. Self-recognition mediated by temporal features may be achieved without an online contribution from the motor system. Participants may have encoded verbally the temporal characteristics of their actions prior to the experiment, and simply compared this description of their own actions, with the temporal properties of third-person stimulus displays. However, self-recognition mediated by temporal features does not exclude the possibility of an online contribution from the motor system during the experiment. A covert simulation process may serve to excite temporal representations, which may thereafter be compared with temporal representations excited by third-person displays. Crucially however, this simulation contribution could be mediated by links acquired associatively during correlated sensorimotor experience.

Self-recognition effects have been cited as evidence that topographic information is transferred from the motor to the perceptual system, in the absence of correlated sensorimotor experience. In this respect, self-recognition effects are conceptually similar to reports that non-visual motor training, during which the participant is blindfolded, improves action recognition (Casile & Giese, 2006; Hecht et al., 2001); an effect also cited as evidence of topographic transfer. As discussed above, the finding that self-recognition is dependent on temporal cues argues against this view, indicating that an endogenous transfer is not required. The present findings therefore raise the possibility that the effects of non-visual motor training might also be mediated by temporal cues. Although being blindfolded deprives actors of first-person visual experience of a newly trained gait, participants can still use proprioceptive, somatosensory, vestibular and auditory feedback to identify the temporal characteristics of the novel gait. Such temporal expertise may thereafter allow recognition of the gait from third-person perspectives.

In summary, the present findings argue that the self-recognition advantage is mediated by temporal and not topographic cues. Crucially, an endogenous process, whereby third-person topographic representations are derived from first-person motor experience, is therefore unnecessary. Insofar as temporal information is readily available during correlated sensorimotor experience, self-recognition mediated by temporal cues is in no way inconsistent with associative solutions to the visuomotor correspondence problem.

Chapter 3: Contingent visual feedback is necessary to refine visuomotor correspondence knowledge

3.1 Introduction

The associative account, embodied within the ASL model (section 1.2.1.1), argues that the correspondence problem is solved through the acquisition of associations which link directly visual and motor representations of actions. Where the sight of an action reliably predicts the execution of that action, the visual and motor representations will become associated. Thereafter the sight of an action may excite the corresponding motor representation. ASL states that this learning is sensitive to both temporal contiguity and the sensorimotor contingency. Temporal contiguity describes the extent to which action observation and execution occur at the same time – where a delay is imposed between action execution and observation, ASL predicts that less learning will occur. Sensorimotor contingency describes the strength of the predictive relationship between action execution and observation.

In contrast, the AIM model (section 1.2.2.1) proposes that humans are born with the ability to use proprioceptive feedback to identify correspondence between observed and executed actions. AIM proposes that an innate ‘supramodal representational system’ encodes visual and proprioceptive representations of observed and executed actions in terms of ‘organ relations’ using a ‘supramodal code’. Within the AIM framework, this supramodal code is the common metric which allows the correspondence between observed and executed acts to be established. Once the actor’s current posture and the target organ configuration are described in a common supramodal metric, imitation is achieved through an iterative, match-to-target process, mediated by a proprioceptive feedback loop. Where a mismatch is detected between the visually-derived and the proprioceptively-derived supramodal representations, corrective actions are inferred and executed. Thereafter, the proprioceptive feedback from the new body posture is processed and, where necessary, a new iteration initiated.

It has proved remarkably difficult to test these two models against one another (Leighton & Heyes, 2010). Crucially, however, these two putative solutions to the correspondence problem make different predictions about the conditions in which imitative performance will show progressive improvement. Both models allow that improved imitative performance could result from independent perceptual- and motor-learning. Perceptual learning potentially enhances the quality of the sensory input into the core mechanism responsible for resolving the correspondence problem. Thus, repeated observation of action may yield more complete visual representation, affording stronger activation of a direct visual-motor link (ASL) or a more accurate supramodal representation of organ relations (AIM). Equally, motor learning may improve imitative performance through the modulation of output mechanisms. Repeated performance of an action may produce stronger connections between supramodal representations of action and the motor commands controlling movements (AIM) or allow more fluent action execution of associatively activated motor programs by the peripheral motor system (ASL).

However, the two models make different predictions about the kind of experience that can generate improvement by enhancing the functioning of the core mechanisms thought to solve the correspondence problem. The ASL model suggests that these core mechanisms are direct associations between visual and motor representations, and that they can be refined only through contingent experience in which the performance of an action is paired with accurate third-person visual consequences. Sources of accurate third-person visual experience include reflections in mirrors or other surfaces, and being imitated accurately by others. In contrast, AIM suggests that the core mechanism is a system that explicitly compares vision-derived and proprioception-derived supramodal representations of action, and that its functioning can be enhanced by action execution in the absence of third-person visual feedback. According to AIM, proprioceptive feedback from action execution updates a supramodal representation of what the imitator has just done. This is compared with a vision-derived supramodal representation of the other person's action, and an error-correction procedure initiates remedial action (Meltzoff & Moore, 1997).

While it has long been recognised that identifying the conditions necessary for imitative improvement could resolve the correspondence problem (Meltzoff & Moore, 1994), few studies have attempted directly to exploit this source of information. Consistent with AIM, Meltzoff and Moore (1994) reported that, over trials without third-person visual feedback, infants produced progressively more accurate imitations of sideways tongue protrusion. Similarly, other studies of infant imitation of mid-line tongue protrusion have scored early responses as 'partial' matches, but later responses as 'full' or 'complete' matches (Abravanel & Sigafoos, 1984; Jacobson, 1979; Soussignan et al., 2011). However, these observations are not decisive for a number of reasons. First, existing evidence of improvement depends on subjective judgements of similarity between the model's and the infant's behaviour. For example, Meltzoff and Moore (1994) controversially deemed a large mid-line tongue protrusion a closer match of side-ways tongue protrusion, than small sideways protrusions. Second, any improvement could have been due to perceptual or motor learning, rather than to enhancement of the core mechanisms that solve the correspondence problem. Third, it is not clear that the tongue protrusion matching exhibited by young infants reflects the presence of mature correspondence knowledge (section 1.3.2.1).

The experiments described in Chapter 3 used improvement in facial imitation to test the AIM and ASL solutions to the correspondence problem, using methods that overcome each of these problems. In order that scoring was not influenced by subjective assumptions, the accuracy of facial gesture imitation was measured for the first time using a precise, objective procedure based on automated calculation of Euclidean distances. The likelihood that improvement would be based on perceptual or motor learning was minimised by requiring participants to imitate their own, spontaneously produced actions. This ensured that the model's face was familiar, and that the target actions could be performed fluently. Finally, typically developing adults were used as participants, ensuring that mature cognitive mechanisms were studied.

Using these methods, Experiments 3 and 4 examined performance over successive attempts to imitate a facial expression. In each of two experiments, one group of participants was given contingent, third-person visual feedback across successive training attempts, while the other group was not. If imitation is mediated by associatively acquired links which form a direct connection between visual and motor representations of action, one would expect improvement only in the visual feedback group. Only in this condition do participants receive experience of the kind that, according to the ASL model, is necessary to enhance the core mechanisms of imitation. However, according to AIM, proprioceptive feedback is sufficient to enhance the resolution of the comparison process at the heart of imitation. If imitation is mediated by a supramodal representational system, one would therefore expect all groups to show improvement.

3.2 Experiment 3

According to the ASL model, the accuracy of visuomotor links may be refined only by contingently pairing the execution of an action with its precise third person visual consequences. In contrast, AIM predicts progressive improvement both with and without contingent visual feedback, due to the availability of proprioceptive feedback. To test these rival predictions, Experiment 3 compared two groups' self-imitation accuracy before, during and after self-imitation training. During training, one group ("visual feedback") received contingent visual feedback from their latest attempt as soon as it had been recorded, while the second group ("no visual feedback") were reliant solely on proprioception throughout. According to the AIM hypothesis, both groups should show evidence of improvement because both receive proprioceptive feedback following each attempt. In contrast, ASL predicts that only the visual feedback group should improve.

3.2.1 Methods

3.2.1.1 Participants

Twenty healthy adults (6 male) with a mean age of 23.9 years participated in the experiment in return for a small honorarium. Participants were assigned randomly to the two groups in equal numbers. All were right handed, had normal or corrected-to-normal vision, and were naive to the purpose of the experiment. None of the participants had taken part in experiments described elsewhere in this thesis.

3.2.1.2 Stimuli

Target and feedback expression stimuli were colour images, saved and presented as uncompressed bitmap files, subtending approximately 10° of visual angle.

3.2.1.3 Procedure

The experiment took place in a small testing cubicle lit by an overhead light and a spotlight, both of which were diffused. The self-imitation paradigm necessitated a two stage procedure. In the first phase participants were filmed for 16 seconds while they recalled and recited question and answer jokes, a method used previously to elicit natural facial expressions (Hill & Johnston, 2001). The purpose of this initial sequence was two-fold: First, the variation present within the sequence was used to construct a measurement-space within which the accuracy of self-imitation was subsequently assessed. Second, four frames were selected from the sequence to serve as target expressions. Sequences were captured at 20 frames-per-second using a Microsoft Lifecam High-definition webcam. Throughout each experiment participants sat with the back of their head resting within a U-shaped head guide.

In the second phase participants attempted to imitate each of the four target expressions. Targets were selected semi-randomly by the experimental program, with selection biased to sample different parts of the joke sequence. Target expressions were presented centrally. Participants were free to take as long as they wished to pose their head and face. When participants reached what they perceived to be the optimal pose, they recorded their attempt with a mouse click. Participants made 18 successive attempts to imitate each target expression: four pre-test attempts, ten training attempts, and four post-test trials. No visual feedback was given after pre- and post-test attempts. Imitative attempts were blocked such that participants made 18 attempts at the first target, followed by 18 attempts at the second target, and so on.

Both groups completed the first four pre-test attempts and the last four post-test attempts without visual feedback. Following the capture of each training attempt participants in the visual feedback group were immediately presented with an acknowledgement display for five seconds depicting a still image of their attempt alongside the target expression. Members of the visual feedback group were told that the feedback was from their most recent attempt. Participants in the no visual feedback group were presented with the same acknowledgment display, without the attempt image, to ensure both groups received equal visual experience of the target. Participants in both groups were instructed to replicate the target image as closely as possible both in terms of the rigid head position and non-rigid facial expression. Participants were not explicitly informed that the experiment was studying imitative improvement, and were encouraged to do their best on each attempt irrespective of the provision of feedback. Both groups remained naive to the pre-test, training, post-test design throughout the experiment.

3.2.1.4 Scoring

The accuracy of self-imitation was assessed by projecting targets and attempts into a multi-dimensional space derived from PCA (Figure 7, top panel). First, an algorithm was applied to each frame from the joke sequence, and to the self-imitation attempts, which recovered the displacement vector field and texture variation between that frame and the average facial image (Berisha et al., 2010; Johnston et al., 1999; Johnston et al., 1992). Next, PCA was used to extract the 50-dimensional structure that most efficiently described the image variance present in the 320 frame joke sequence. Each principal component (PC) therefore represented a mode of facial variation. The application of PCA permitted each frame from the joke sequence to be represented as a point or 'vector' defined by its loadings on the 50 PCs. Through the application of the same algorithm, it was also possible to represent each self-imitation attempt as a vector within the same space. Thereafter, the accuracy of an attempt was quantified by calculating its Euclidean distance to the target in PC space (denoted by '*d*PC'). The closer the two points, the better the attempt (Figure 7, bottom panel).

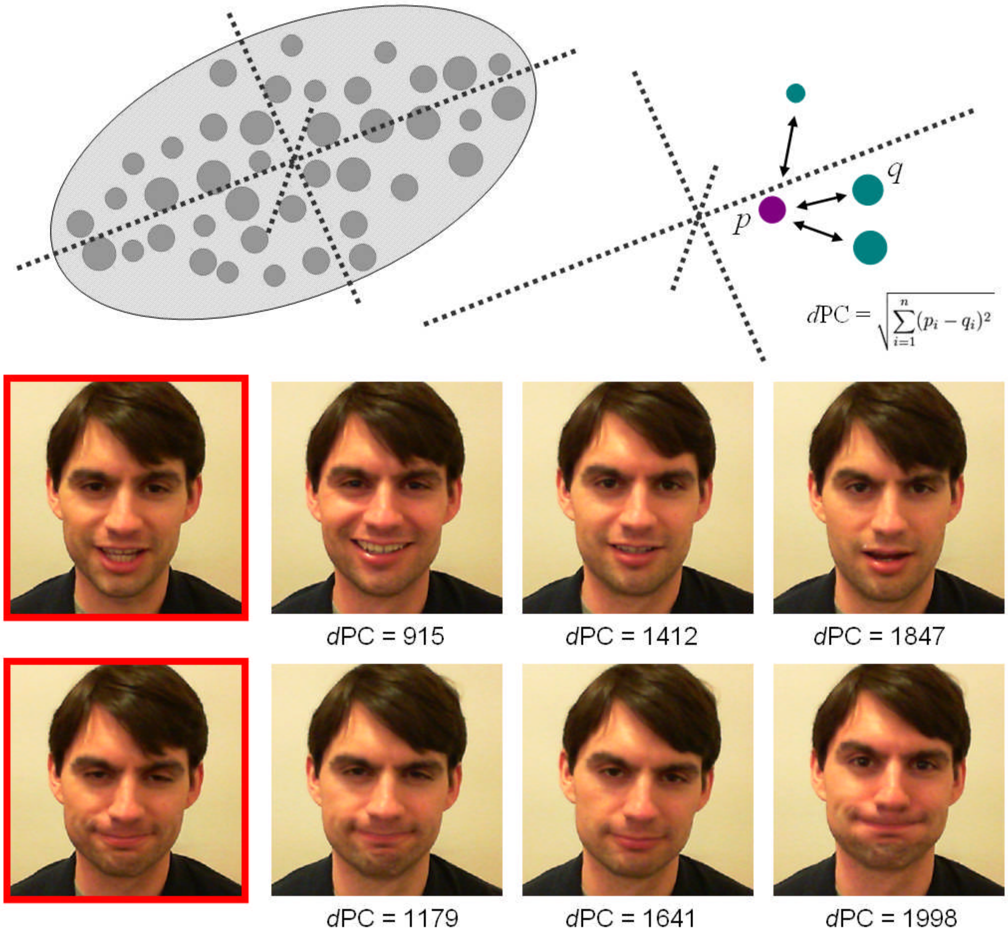

Figure 7: Having described each of the frames of a movie sequence in terms of their spatial and texture deviation from the sequence mean, PCA was used to extract the orthogonal dimensions which represented the variation in the facial images most efficiently (top left panel). The application of PCA allowed target expressions (purple) and imitative attempts (green) to be represented as points within a common multi-dimensional space (top right panel). Within this framework, accuracy of attempts is indicated by the Euclidean distance between the two points: The closer the attempt vector is to the target, the better the attempt. The application of this scoring is demonstrated in the bottom panel. The targets are presented within a red border, with attempts to the right. The Euclidian distance to the target is provided under each attempt.

3.2.2 Results & Discussion

3.2.2.1 Training

The mean Euclidian distances observed across the ten training attempts (Figure 8a) were analysed using ANOVA with training attempt (1-10) as a within subjects factor and group (visual feedback, no visual feedback) as a between subjects factor. No main effects of attempt ($p > .90$) or group ($p > .50$) were observed. However the analysis revealed a highly significant interaction between the linear trend observed across training attempts and group [$F(1,18) = 17.01$, $p < .001$, $\eta^2 = .49$]. The visual feedback group showed a significant linear improvement across attempts [$F(1,9) = 6.02$, $p < .05$, $\eta^2 = .40$]. Whereas the mean distance from target was 1320 *d*PC (*SD* = 250 *d*PC) on the first training attempt, this fell to a mean distance of 1210 *d*PC (*SD* = 260 *d*PC) by the final training attempt. In contrast, the no visual feedback group showed a significant linear deterioration across training attempts [$F(1,9) = 14.82$, $p < .01$, $\eta^2 = .62$]. The mean distance from the target increased from 1336 *d*PC (*SD* = 286 *d*PC) on the first training attempt to 1396 *d*PC (*SD* = 292 *d*PC) on the final training attempt.

3.2.2.2 Test

The mean distances observed at pre- and post-test (Figure 8a) were also analysed using ANOVA with test (pre-test, post-test) as a within subjects factor and group (visual feedback, no visual feedback) as a between subjects factor. A highly significant group x test interaction was observed [$F(1,18) = 18.02$, $p < .001$, $\eta^2 = .50$]. Simple effects analysis revealed that the attempts of the visual feedback group were closer to their targets at post-test (*M* = 1254 *d*PC, *SD* = 257) than at pre-test (*M* = 1331 *d*PC, *SD* = 300) [$t(9) = 3.02$, $p < .025$]. In contrast, the attempts of the no visual feedback group were further from their targets at post-test (*M* = 1416 *d*PC, *SD* = 320) than at pre-test (*M* = 1353 *d*PC, *SD* = 279) [$t(9) = 3.02$, p < .025]. No main effects of test ($p > .60$) or group ($p > .40$) were observed.

3.2.2.3 Attempt durations

Because the task was self-paced, it was possible to assess participants' motivation to succeed by examining attempt duration; the interval between target presentation and participant-initiated response capture. These data were subjected to ANOVA with phase (pre-test, training, and post-test) as a within subjects factor and group (visual feedback, no visual feedback) as a between subjects factor. There was a trend for participants to spend longer on pre-test attempts (*M* = 7.72 secs, *SD* = 4.49 secs) than either training (*M* = 6.55 secs, *SD* = 4.33 secs) or post-test attempts (*M* = 6.13 secs, *SD* = 4.50 secs). However the main effect of phase did not reach significance [*F*(2,36) = 2.95, *p* = .065]. Neither a main effect of group (*p* > .40) nor a group x phase interaction (*p* > .35) was observed. The absence of a main effect of group suggests that participants who did and did not receive visual feedback were equally motivated to succeed in the task.

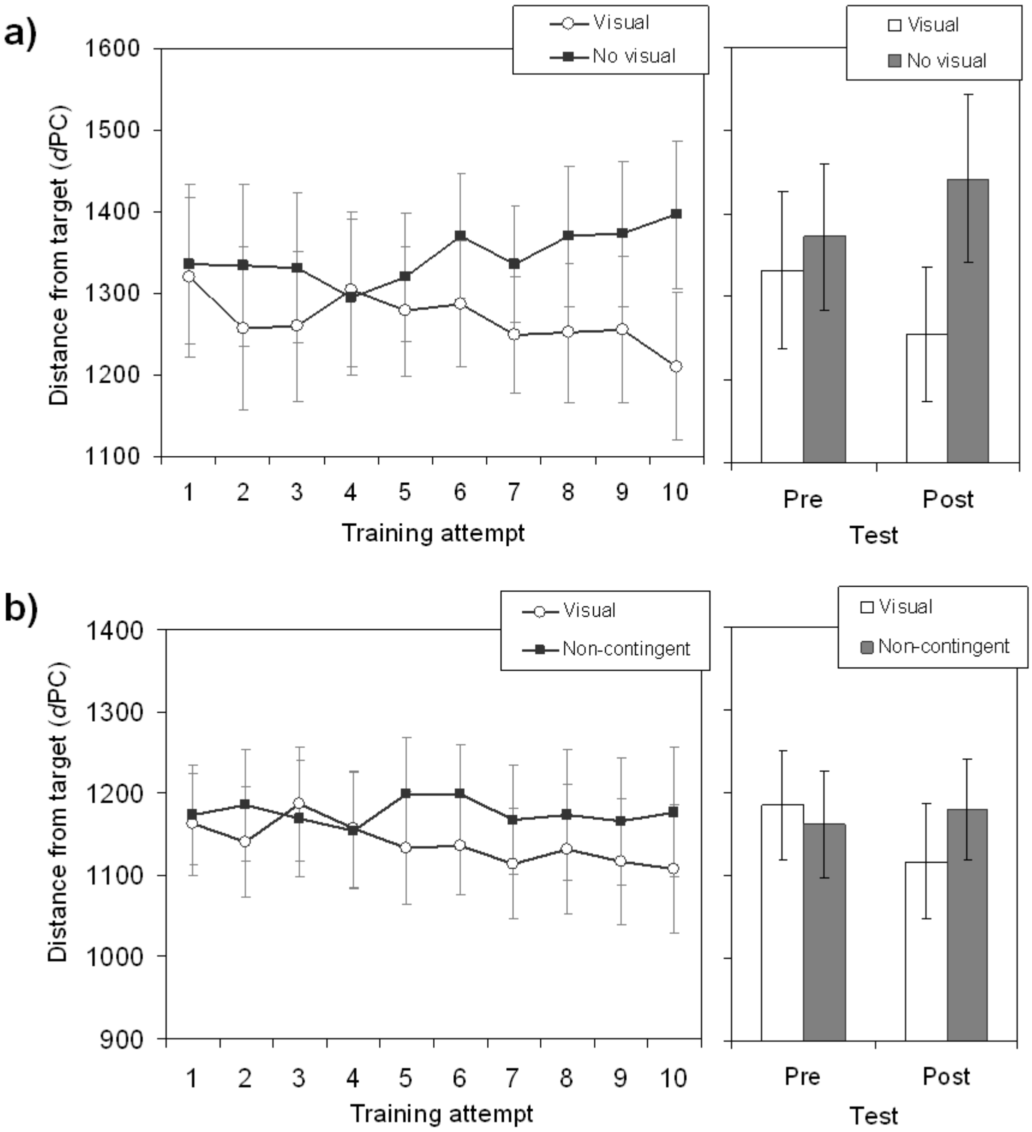

Figure 8: Mean distances achieved (a) by the visual feedback and no visual feedback groups during Experiment 3 and (b) by the visual feedback and non-contingent feedback groups during Experiment 4. Left panels depict mean distances achieved across the ten training attempts. Right panels depict mean distances achieved during the pre- and post-test attempts. In both cases smaller distances reflect better self-imitation. Error bars denote standard error of the mean. Distances were generally closer in Experiment 4 than in Experiment 3, due to a change in the distance between the camera and the participant.

3.2.2.4 Discussion

Participants in the visual feedback group improved the accuracy of their self-imitation during training, whereas participants in the no visual feedback group, who were reliant solely on proprioception, showed a significant linear deterioration. These data argue against the view proposed by the AIM hypothesis, that proprioceptive feedback may be transformed into a supramodal code for the purposes of guiding imitative error-correction. Instead, consistent with the prediction of the ASL model, Experiment 3 suggests that third-person visual experience, which corresponds closely to the gesture executed, is necessary to improve the accuracy of visuomotor correspondence.

3.3 Experiment 4

The ASL model suggests that the image of the recent attempt improved performance in the visual feedback group by virtue of its contingent relationship with the recently executed response; this contingency strengthened excitatory links between specific visual and motor representations of task-relevant facial gestures. However, the AIM hypothesis can be sustained if the visual feedback improved performance by a different route - through perceptual learning. On each training trial in Experiment 3, participants in the visual feedback group observed a novel attempt alongside the target expression. This additional experience of the variation in their facial expressions may have improved participants' ability to represent the target. Without this enhanced visual experience, participants in the no visual feedback group may have failed to perceive subtle features of the target expressions, and consequently derived an inadequate supramodal representation.

To test this perceptual learning account, Experiment 4 compared the performance of a second visual feedback group, with a non-contingent feedback group, which received visual feedback that was not derived from the most recent attempt executed. If the group difference observed in Experiment 3 was due to perceptual learning, this difference should disappear in Experiment 4 because both groups saw the target alongside a novel attempt and therefore had equal opportunity for perceptual learning. However, if the group difference was due to the strengthening of associations between specific visual and motor representations of facial expressions, it should be sustained in Experiment 4.

3.3.1 Methods

3.3.1.1 Participants

A further twenty healthy adults (6 male) with a mean age of 24.2 years participated in the experiment in return for a small honorarium. Participants were assigned randomly to the two groups in equal numbers. All were right handed, had normal or corrected-to-normal vision, and were naive to the purpose of the experiment. None of the participants had taken part in experiments described elsewhere in this thesis.

3.3.1.2 Procedure

The visual feedback group completed the experiments under the same conditions employed in Experiment 3. Participants in the non-contingent feedback group also received feedback, but from a previous trial, not the current training attempt. On the first training attempt (the fifth attempt in total) the feedback presented was from the fourth attempt (n-1). The feedback presented on the second training attempt (the sixth attempt in total) was from the third attempt (n-3). Thereafter, the feedback presented alternated between the *n*th-1 and *n*th-3 attempt. This manipulation served to degrade the sensorimotor contingency because expression execution was no-longer predictive of the visual feedback presented.

Participants in both groups were informed that they would receive feedback “from a recent trial” and received identical instructions in all respects. Participants were again told to replicate the head position and facial expression as closely as possible and encouraged to do their best on each attempt irrespective of the provision of feedback. Again, both groups remained naive to the pre-test, training, post-test design and were not told that the study was investigating imitative improvement until after the experiment. As in Experiment 3, neither group received any feedback on pre- or post-test trials. Participants were seated slightly further away from the camera / monitor during Experiment 4 (approximately 65-70cm) than in Experiment 3 (approximately 60-65cm) to accommodate tall participants more easily.

3.3.2 Results & Discussion

3.3.2.1 Training

The mean Euclidian distances observed across the ten training attempts (Figure 8b) were analysed using ANOVA with training attempt (1-10) as a within subjects factor and group (visual feedback, non-contingent feedback) as a between subjects factor. No main effects of attempt ($p > .25$) or group ($p > .60$) were observed. However the analysis revealed a significant interaction between the linear trend observed across training attempts and group [$F(1,18) = 4.59$, $p < .05$, $\eta^2 = .20$]. The visual feedback group showed a significant linear improvement in their attempts [$F(1,9) = 15.76$, $p < .01$, $\eta^2 = .64$]. The mean distance from target fell from 1162 *d*PC (*SD* = 218 *d*PC) on the first training attempt, to 1107 *d*PC (*SD* = 207 *d*PC) on the final training attempt. In contrast, the non-contingent feedback group showed no linear trend across training attempts ($p > .65$).

3.3.2.2 Test

The mean distances observed at pre- and post-test (Figure 8b) were analysed using ANOVA with test (pre-test, post-test) as a within subjects factor and group (visual feedback, non-contingent feedback) as a between subjects factor. A significant group x test interaction was observed [$F(1,18) = 6.59$, $p < .025$, $\eta^2 = .27$]. Simple effects analysis revealed that the attempts of the visual feedback group were closer to their targets at post-test (*M* = 1117 *d*PC, *SD* = 218 *d*PC) than at pre-test (*M* = 1184 *d*PC, *SD* = 210 *d*PC) [$t(9) = 3.99$, $p < .01$]. In contrast, there was no significant difference ($p > .70$) between the mean distance from target achieved by the non-contingent feedback group at pre-test (*M* = 1175 *d*PC, *SD* = 203 *d*PC) and post-test (*M* = 1184 *d*PC, *SD* = 196 *d*PC). No main effects of test ($p > .20$) or group ($p > .40$) were observed.

Euclidean distances tended to be smaller in Experiment 4 compared to those calculated in Experiment 3. ANOVAs conducted with Experiment (3, 4) as a between subjects factor revealed the trend to be significant for both the training [$F(1,38) = 4.26$, $p < .05$] and test distances [$F(1,38) = 5.02$, $p < .05$]. It is likely that this difference reflects the fact that participants were slightly further away from the camera in Experiment 4. As a result, expression variation accounted for a greater proportion of the total image variance in Experiment 3.

3.3.2.3 Attempt durations

Participants' mean attempt durations were analysed using ANOVA with phase (pre-test, training, and post-test) as a within subjects factor and group (visual feedback, non-contingent feedback) as a between subjects factor. A significant main effect of condition was observed [$F(2,36) = 16.2$, $p < .001$]. Participants spent longer on the pre-test attempts (*M* = 6.14 secs, *SD* = 1.97 secs) than either the training (*M* = 4.70 secs, *SD* = 1.43 secs) [$t(19) = 4.76$, $p <.001$] or post-test attempts (*M* = 4.48 secs, *SD* = 1.28 secs) [$t(19) = 4.33$, $p <.001$]. As in Experiment 3, neither a main effect of group ($p > .80$) nor a group x condition interaction ($p > .65$) was observed.

3.3.2.4 Discussion

Participants in the visual feedback group improved the accuracy of their self-imitation during training, whereas participants in the non-contingent feedback group showed no evidence of improvement. The absence of improvement in the no visual feedback group in Experiment 3 could conceivably be attributed to poor perceptual representation due to limited opportunity for perceptual learning. However, such an account cannot explain the differential learning effects observed in Experiment 4, as participants in the visual feedback and non-contingent feedback groups observed the same number of displays in which a target appeared alongside an attempt.

3.4 General Discussion

The imitation of facial gestures presents the cognitive system with a particularly challenging correspondence problem because they are perceptually opaque (section 1.1). The AIM model suggests that this correspondence problem is solved by an inborn, indirect mechanism that uses a dedicated supramodal representational system to compare vision-derived and proprioception-derived representations of action. This account predicts that, when opportunities for perceptual- and motor-learning are controlled, imitative performance can be improved using proprioceptive feedback alone. The present findings provide no support for this hypothesis. In two experiments, adults reliant solely on proprioceptive feedback failed to show improvement in facial self-imitation performance. In contrast, and consistent with the solution to the correspondence problem advanced by ASL, participants in both experiments who received visual feedback contingent upon facial gesture execution showed incremental improvement in the accuracy of their self-imitation.

In principle, the group differences observed in Experiments 3 and 4 could reflect weaker motivation in participants who did not receive contingent visual feedback. However, the attempt duration data argue against this view. The groups that did not receive contingent visual feedback spent as long as the other groups trying to match their executed facial expression to the visual target.

The ASL model suggests that the correspondence problem is solved by associatively acquired links that connect directly visual and motor representations of the same action. Therefore, the accuracy of a person's imitative performance depends on the sensorimotor experience to which they have been exposed. Within this framework, the improvement observed in the visual feedback groups can be explained in terms of a discrimination learning process (Figure 9). Experience in noisy sensorimotor environments, where the execution of an action predicts the observation of a range of similar actions, results in the formation of a number of links between broadly matching visual and motor representations. While this is likely to afford ready generalisation (broadly imitative responses may be elicited by a range of approximately similar visual inputs, as they were at pre-test), such imprecise links will detract from the accuracy of imitation. The discrimination hypothesis suggests that contingent visual feedback improves the accuracy of imitation both by strengthening links between precisely matching visual and proprioceptive representations, and by weakening links between only approximately matching representations.

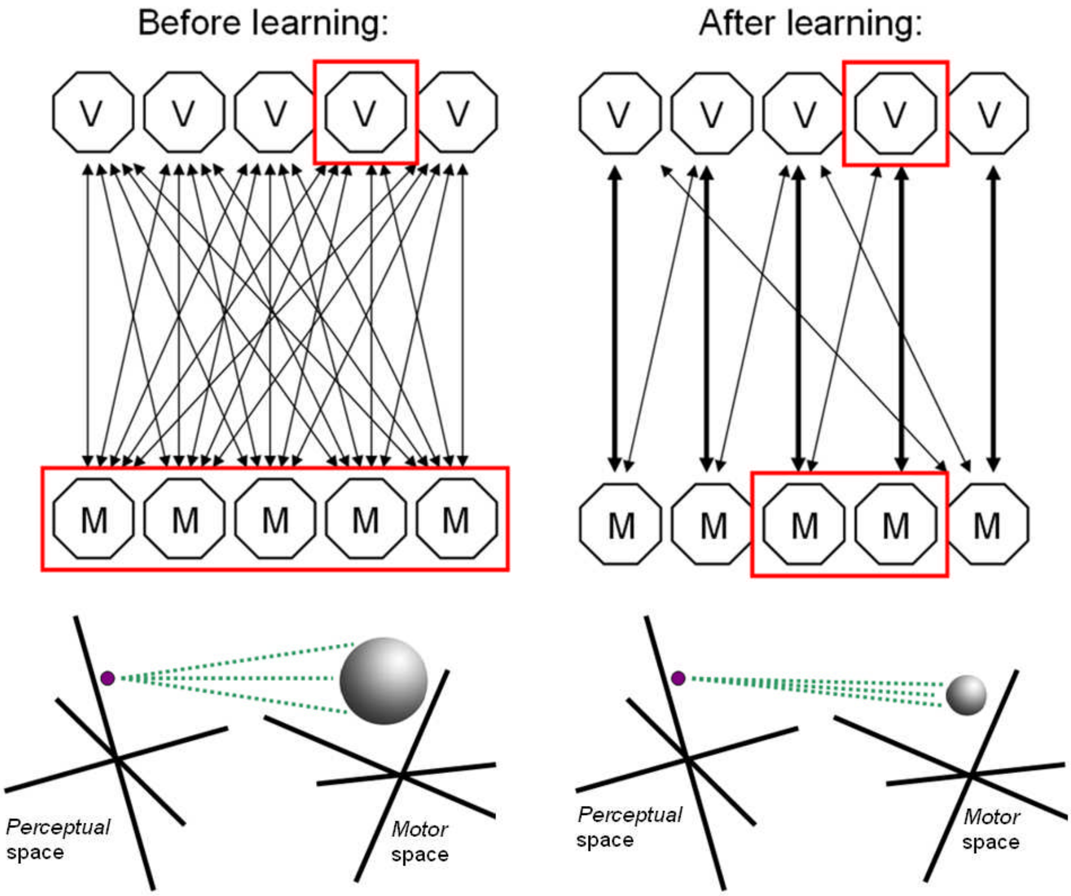

Figure 9: Illustration of the discrimination learning process using schematic representations of the sensorimotor associations in place before (left) and after (right) the provision of contingent visual feedback in this study. In the top panel, vertical arrows denote accurate visuomotor links between visual (V) and the motor (M) representations of facial expressions. Diagonal arrows reflect visuomotor links between broadly similar, but not precisely matching, visual and motor representations. Contingent feedback of the visual consequences of action execution serves to both strengthen the precise vertical links and weaken the imprecise diagonal correspondences. Consequently, the range of 'candidate' motor programs excited by action observation is reduced following experience of precise third-person visual feedback (bottom panels).

Progressive improvement in infants’ imitation of midline (Abravanel & Sigafoos, 1984; Johansson, 1973; Soussignan et al., 2011) and sideways (Meltzoff & Moore, 1994) tongue protrusion has been reported in the absence of contingent visual feedback. The contrast between these reports and the present findings could be due to several factors. First, the aforementioned studies failed to use objective methods to score performance. For example, Meltzoff and Moore (1994) arbitrarily deemed a large mid-line tongue protrusion a closer match of side-ways tongue protrusion, than small sideways protrusions. Second, apparent improvement may reflect motor-learning. Those studies which report incremental improvements in infants’ tongue protrusion matching all confound accuracy of imitation with response vigour (Meltzoff & Moore, 1994; Nagy et al., 2005; Soussignan et al., 2011). Where repeated execution causes infants’ gestures to become more pronounced, scorers may interpret changes over time as increasing accuracy (Ray & Heyes, 2011). Third, the different pattern of findings might reflect the use of adult participants in the present study. It is possible that the mechanisms responsible for tongue protrusion matching in infants are distinct from the sensorimotor associations mediating adults’ facial gesture imitation. Consistent with this possibility, studies have shown that tongue protrusion matching ceases to occur in infants at around three-months, and reappears around six-months together with a broader range of imitative responses (Abravanel & Sigafoos, 1984; Fontaine, 1984; Jacobson, 1979).

While the present findings argue against the AIM hypothesis, they do not help to distinguish between the different associative solutions to the correspondence problem – they are consistent with both the ASL model, which specifies that learning is sensitive to contiguity and contingency, and the Hebbian hypothesis of the origins of MNs, which specifies that learning is determined solely by temporal contiguity. In the non-contingent visual feedback condition employed in Experiment 4, the contingency between execution and feedback was degraded by presenting feedback, not from the current training attempt, but from a previous trial. Specifically, the feedback presented alternated between the *n*th-1 and *n*th-3 attempt. This manipulation successfully disrupted the sensorimotor contingency; expression execution became a poor predictor of the feedback presented. However, this manipulation also eliminated the temporal contiguity between expression execution and visual feedback. The comparison made in Experiment 4 was thus between a contingent, contiguous condition and a non-contingent, non-contiguous condition. The findings from Experiments 3 and 4 therefore accord with both the ASL model and the Hebbian hypothesis.

In summary, the experiments reported in Chapter 3 indicate that adults find it very difficult to use proprioception to inform imitative error correction. In contrast, the provision of third-person visual feedback readily affords imitative improvement. Previous reports that infants' imitation improves in the absence of visual feedback have provided important empirical support for nativist solutions to the correspondence problem. They argue against the view that neonatal matching of tongue protrusion reflects generic arousal or the operation of an IRM, and instead suggest that these responses are intentional. Critically, they also suggest that humans are endowed with an innate means of indentifying disparities between our own postures and expressions and those we observe, and inferring the correction necessary to match the two. However, the present findings indicate that adults do not possess an innate means to match the seen-but-unfelt to the felt-but-unseen, and suggest that previous reports may have been artefacts of the particular scoring procedures employed.

Chapter 4: The acquisition of visuomotor links is sensitive to contingency

4.1 Introduction

The experiments reported in Chapters 2 and 3 indicate that the correspondence problem is solved associatively; that our ability to match observed and executed actions is mediated by visuomotor links acquired through correlated sensorimotor experience. However, none of the experiments described so far help to distinguish between the two leading associative accounts – the ASL model (section 1.2.1.1) and the Hebbian account of MN origins (section 1.2.1.3). The experiments reported in this chapter sought to test these theories by determining whether the acquisition of links between visual and motor representations is sensitive to sensorimotor contingency.

The ASL model proposes that the correspondence problem is solved by phylogenetically ancient mechanisms of associative learning; the very same learning mechanisms that produce Pavlovian and instrumental conditioning in humans and animals. Studies of instrumental learning have identified two main determinants of the acquisition of associations between stimuli (Ss) and responses (Rs): the temporal contiguity between stimulus and response and their contingency.

By imposing a delay between a stimulus and a response, it is possible to reduce their temporal contiguity - the extent to which they overlap in time. Under these conditions, learning is known to decrease in proportion to the delay in both animals (Dickinson, Watt, & Griffiths, 1992) and humans (Shanks & Dickinson, 1991). However, associative learning mechanisms are thought to be sensitive, not only to the probability of the response given the stimulus [P(R/S)], but also to the probability of the response in the absence of the stimulus [P(R/~S)] or, in other words, to the contingency between the stimulus and response. If a response is just as likely to occur in the absence of the stimulus as in its presence, the stimulus does not predict an increase in the likelihood of the response above its base rate and is therefore uninformative about the occurrence of the response. Contingency sensitivity is thought to prevent the acquisition of ‘superstitious’ associations, ensuring that only those stimuli and responses that genuinely co-vary become associated.

Both animal (Dickinson & Charnock, 1985) and human instrumental learning (Elsner & Hommel, 2004; Shanks & Dickinson, 1991) is sensitive to the contingency between stimulus and response. In these studies the probability of S-R pairings was fixed to ensure a constant number of contiguous pairings, while the contingency was degraded by increasing the probability of unpaired presentations. Learning was observed when the probability of unpaired presentations was low, and therefore the contingency high. However little or no learning occurred in a non-contingent condition in which the probabilities of paired and unpaired presentations were the same.

Hebbian MN theory is conceptually similar to ASL. Both theories posit that the correspondence problem is solved by direct links between sensory and motor representations, acquired through experience. However, insofar as it is based solely on the principles advanced by Hebb (1949), the Hebbian hypothesis implies that the contiguous activation of sensory and motor representations is sufficient for sensorimotor learning to occur. Temporal contiguity is the fundamental principle of Hebbian learning. For example, Hebb posits “that any two cells or systems of cells that are repeatedly active at the same time will tend to become ‘associated,’ so that activity in one facilitates activity in the other” (p70). A similar emphasis on temporal contiguity is present in the Hebbian account of the origins of MNs:

> “When they are young, monkeys and humans spend a lot of time watching themselves. Each time, the child’s hand wraps around an object, and brings it towards him, a particular set of neural activities overlaps in time. Neurons in the premotor cortex responsible for the execution of this action will be active at the same time as the audio-visual neurons in the STS responding to the sight and sound of grasping. Given that STS and F5 are connected through PF, ideal Hebbian learning conditions are met (Keysers & Gazzola, 2006, p391-392).”

The two experiments reported in this chapter manipulated the contingency between stimulus and response in the counter-mirror training paradigm (section 1.3.1.2.2), in order to evaluate the role of temporal contiguity in the acquisition of visuomotor correspondences. During counter-mirror training participants are required to execute a response (e.g. hand open) in the presence of the incompatible counter-mirror stimulus (e.g. hand close). During this training, participants are thought to acquire novel counter-mirror visuomotor links conceptually equivalent to the mirror links thought to mediate correspondence knowledge. The magnitude of counter-mirror learning is quantified through reductions or reversals of behavioural or electrophysical markers of MNS functioning (Catmur et al., 2008; Catmur et al., 2007; Heyes et al., 2005). For example, it has been shown that automatic imitation – the facilitation of imitative responses and interference with non-imitative responses by a task irrelevant action stimulus – may be eliminated by periods of counter-mirror training (Heyes et al., 2005). Experiment 5 sought to determine whether counter-mirror learning is sensitive to variations in contingency by measuring automatic imitation effects following perfect-contingency (contingent training) or zero-contingency (non-contingent) counter-mirror training. Having observed an effect of contingency in Experiment 5, Experiment 6 attempted to clarify the mechanism by which contingency acts upon sensorimotor learning through the addition of a third ‘signalled’ training variant.

4.2 Experiment 5

In order to determine whether the acquisition of visuomotor links is sensitive to variations in sensorimotor contingency, residual automatic imitation effects were measured following two sessions of either contingent or non-contingent counter-mirror training. During training, participants were given a choice reaction-time (RT) task in which hand-open responses were made to closed-hand stimuli and hand-close responses made to open-hand stimuli. This was achieved through the use of a compound stimulus comprising a task irrelevant hand stimulus (depicting either an open- or closed-hand) and a numerical imperative stimulus (either 1 or 2). Participants thus responded to the number presented, but were exposed to incompatible action stimuli. The number of contiguous sensorimotor pairings was held constant across the training groups - both groups made equal numbers of open and close responses in the presence of the incompatible (and irrelevant) hand stimuli (hereafter ‘paired trials’). However, the sensorimotor contingencies between the incompatible hand stimuli and the responses were degraded for the non-contingent group through the addition of trials in which responses were made in the absence of an incompatible hand stimulus (hereafter ‘unpaired’ trials’).

Residual automatic imitation was subsequently assessed using a simple RT task. Starting from a neutral hand gesture, participants were required to make a pre-specified response (e.g. open) to the onset of compatible (open) and incompatible (closed) stimuli. The magnitude of the residual imitation effect was calculated by subtracting the mean RT on compatible trials (open responses to open stimuli and close responses to closed stimuli) from the mean RT on incompatible trials (open responses to closed stimuli and closed responses to open stimuli).

If visuomotor links are acquired through associative learning, the effects of sensorimotor training should be sensitive to contingency. ASL therefore predicts that the contingent training condition will produce greater counter-mirror learning, and hence less automatic imitation, than non-contingent training. Conversely, if equivalent learning is observed in the contingent and non-contingent groups, this would indicate that the acquisition of visuomotor links is sensitive to contiguity only.

4.2.1 Method

4.2.1.1 Participants

Twenty-four healthy adults (5 male) with a mean age of 23.0 years served as participants in the experiment in return for a small honorarium. All were right handed, had normal or corrected-to-normal vision, and were naive to the purpose of the experiment. Participants were assigned randomly to either the contingent or non-contingent training groups in equal numbers. Two were replacements for participants who responded slowly during the test phase (RT > 400ms). None of the participants had taken part in experiments described elsewhere in this thesis.

4.2.1.2 Stimuli

The stimuli used in the training phase consisted of 12 digital images; eight for use on paired trials, and four for use on unpaired trials (Figure 10). Each depicted a model's right hand in naturalistic tones against a black background. Of the eight images used on the paired trials, four showed an open hand, and four a closed hand. The open stimulus showed fingers and thumb splayed. The male stimulus subtended approximately 17° x 20° and the female stimulus 16° x 16°. The closed gesture was a fist. The male stimulus subtended approximately 11° x 14° and the female stimulus 8° x 12° vertically. Four images depicted a male hand and four a female hand. The four stimuli used on the unpaired trials each presented a neutral hand gesture, with fingers together and pointing upwards in parallel with the thumb. Two of the images depicted a male hand and two a female hand. The imperative stimulus was a white number, either a 1 or a 2, presented centrally.

Four similar digital images were used during the test phase. The same male and female hands were shown, depicting either an open or closed gesture. However, in contrast to the images used during training, the test stimuli did not have a number superimposed on them.

A further two images were employed during training and test phases as warning stimuli. These images depicted the same neutral hands, either male or female, as the stimuli presented on the unpaired trials, but without a superimposed numerical stimulus.

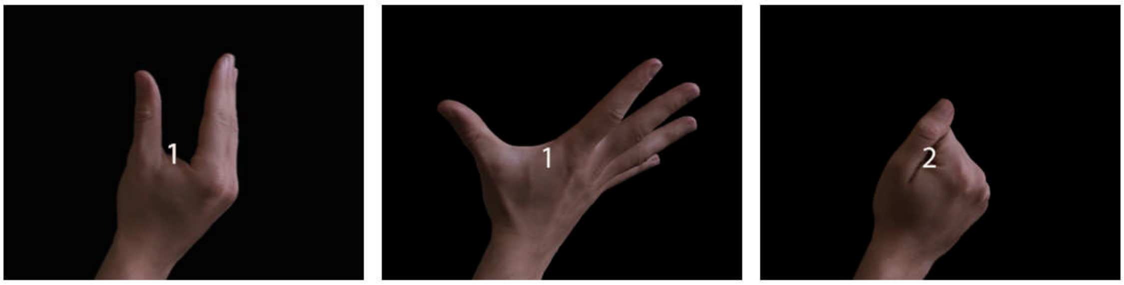

Figure 10: Examples of the stimuli used during training. Male compound stimulus presented on unpaired trials (left); male compound stimuli presented on paired trials (central and right).

4.2.1.3 Data recording and analysis

Both the training and test procedures took the form of RT tasks where electromyography (EMG) was used to establish response onset. Recordings were taken from the first dorsal interosseus (FDI) muscle using disposable Ag/AgCl surface electrodes. The EMG signal was amplified, mains-hum filtered at 50 Hz and digitized at 2.5 kHz. They were rectified and smoothed using a dual-pass Butterworth filter, with cut-off frequencies of 20 Hz and 1000 Hz. The EMG signal was registered for 100 ms before the onset of the imperative stimulus to define the baseline. A window of 20 ms was then shifted incrementally over the raw data in 1-ms steps. Response onset was defined as the start of the earliest 20-ms window, following presentation of the imperative stimulus, in which the standard deviation for that window, and for the following 20-ms epoch, was greater than 2.75 times the standard deviation of the baseline. That this criterion reasonably defined gesture onset was verified by sight for every training and test trial.

4.2.1.4 Procedure

The experiment comprised a training phase (sessions 1 and 2) and a test phase (session 3). In total participants, completed three hour-long sessions, each completed approximately 24-hours apart. Throughout each session, participants sat with their shoulders and right forearm parallel to the stimulus display. Their elbow and forearm were supported by an armrest, with both the hand and forearm occluded from view. The participant's hand and wrist were positioned such that their fingers moved upwards during open responses and downwards during close responses. Stimulus postures were presented in the lateral plane (left–right), thus ensuring response movements were orthogonal to stimulus postures throughout. This feature of the design allows automatic imitation to be isolated from left-right spatial compatibility.

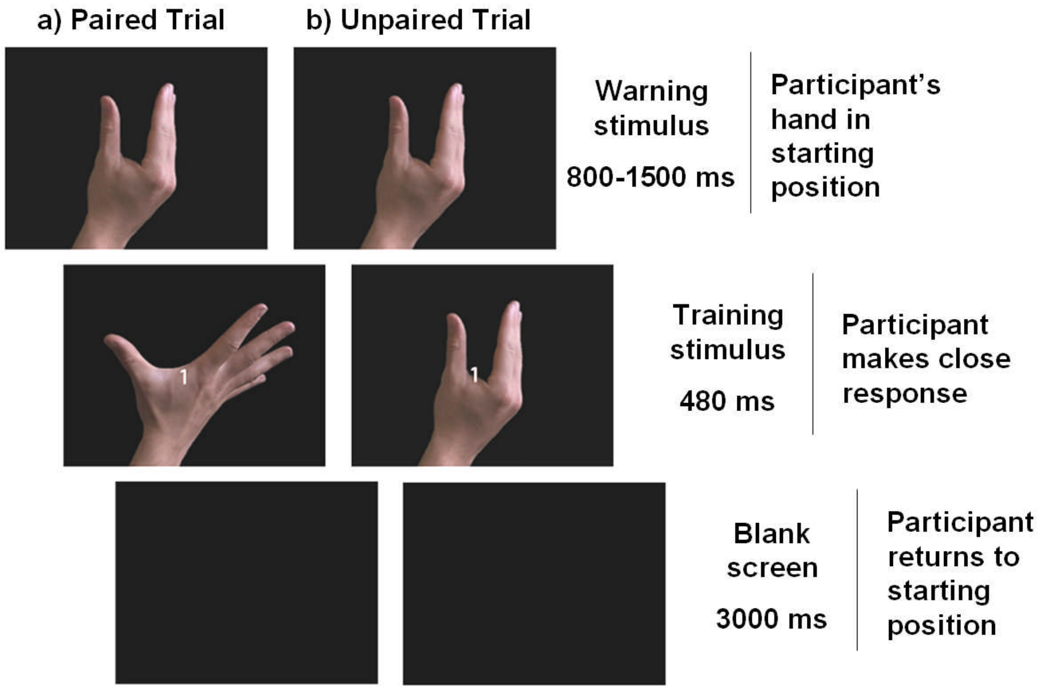

Figure 11: (a) Display sequence for paired trials employed in both the contingent and non-contingent training variants, for response mappings 1 = make close hand response; 2 = make open hand response. (b) Display sequence for an unpaired trial employed in the non-contingent training condition to degrade S-R contingency.

The training procedure took the form of a choice RT task (Figure 11). The contingent and non-contingent training variants each consisted of six blocks of incompatible training trials spread equally over two sessions, completed approximately 24-hours apart. On each trial, participants were required to make one of the two speeded response movements (open or close) as soon as the numerical imperative stimulus appeared, and then to return to the starting position ready for the next trial. Whether participants were required to make open or close responses to either 1 or 2 was counterbalanced across training conditions. The order of stimulus presentation during the training trials was randomised. The beginning of each trial was indicated by the appearance of the warning stimulus, which was then replaced by one of the training stimuli, presented for 480ms. Stimulus onset asynchrony (SOA) was varied randomly between 800ms and 1500ms in

50ms increments. The hand depicted in the warning stimulus was identical to that in the subsequent training stimulus, thus giving rise to apparent motion. On paired trials, the compound training stimulus consisted of an imperative stimulus (a ‘1’ or a ‘2’) superimposed on a task-irrelevant action image (an open or closed hand). On unpaired trials, the imperative stimulus was superimposed on the image of the neutral hand stimulus.

For the contingent group, training blocks comprised 72 paired trials (Figure 11a). On 36 trials participants were required to make close responses having seen an open stimulus, and on 36 trials they made open responses having seen a closed stimulus. Thus, for the contingent group, there was a perfect contingency between the irrelevant stimuli and the counter-mirror responses [P(R/S) = 1; P(R/~S) = 0].

For the non-contingent group, each training block comprised 144 trials (72 paired, 72 unpaired). On 36 of the paired trials participants were required to make close responses having seen an open stimulus, and on 36 trials they made open responses having seen a closed stimulus. Consequently, the number of temporally contiguous sensorimotor pairings was held constant across the training groups. However, the non-contingent training contingency was degraded through the addition of 72 unpaired trials in which responses were elicited from participants in the absence of an incompatible irrelevant stimulus (Figure 11b). On 36 of the unpaired trials participants had to make open responses to the imperative stimulus, and on 36 of the unpaired trials they had to make close responses. Consequently, during non-contingent training, there was no contingency between the irrelevant stimuli and the responses because the responses were equally likely to occur in the presence and in the absence of the incompatible stimuli [P(R/S) = P(R/~S) = 1]. Open and closed stimulus trials were again presented in random order. Whether trials were paired or unpaired was also randomised. Because the non-contingent blocks were longer, participants were given an opportunity to take a short break after the first 72 trials.

Approximately 24-hours after completion of the second training session, both groups were given identical test procedures to assess the impact of the two training variants on participants’ compatibility effects. The test was a simple RT task, in which EMG recording was again used to establish response onset. The test comprised two blocks of 120 trials during which participants made speeded pre-specified movements (open responses in one block, close in the other) in response to the onset of open and closed stimuli presented in a randomised order (Figure 12). This factorial manipulation constitutes a SRC design whereby responses can be either compatible with the observed stimulus (e.g. hand-open response to hand-open stimulus) or incompatible (e.g. hand-open response to hand-close stimulus). On each stimulus trial, participants started in the neutral hand position, were required to make their response as soon as a gesture stimulus appeared and then return to the neutral position ready for the next trial. The gesture stimulus was present for 480ms until replaced by a blank display for 3000 ms, prior to the warning stimulus for the next trial. SOA was varied randomly between 800ms and 1500ms in 50ms increments. The order in which participants completed the open and close blocks was counterbalanced.

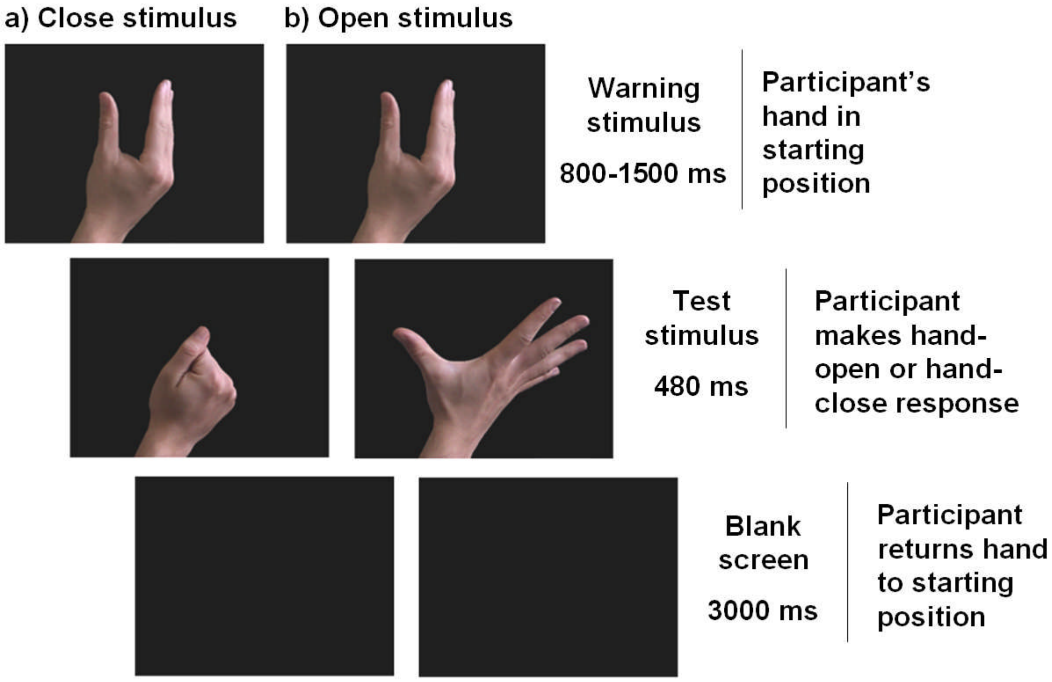

Figure 12: Display sequences for test trials presenting (a) close-hand and (b) open-hand test stimuli. In one block of test trials participants made open-hand responses to stimulus onset; in the other they made close-hand responses.

Twenty-four catch trials were included in each test block, in which the warning stimulus was displayed throughout the trial to which the participants had been instructed to make no response whatsoever. Because participants were making speeded pre-specified responses, catch trials were included to prevent habitual, anticipatory responding and to encourage participants to continue to monitor what the stimulus was doing and to remain engaged with the task. On catch trials, the warning stimulus was presented for 1,980 ms before the 3,000 ms inter-trial interval.

4.2.2 Results and Discussion:

4.2.2.1 Data handling

Training and test trials in which participants made incorrect responses, no response at all, or where the time of movement onset was equivocal were excluded from all further analyses (2.5% of training trials; 1.6% of test trials). Any remaining data points beyond 2.5 standard deviations of a participant's mean response latency for a given block were also excluded (2.1% of training trials; 1.8% of test trials). Participants' EMG signals on catch trials were examined to ensure that they obeyed task instructions, but were excluded from all further analyses. During the test sessions, participants initiated movements on only 5.8% of catch trials, implying that they were not making anticipatory responses on the stimulus trials.

4.2.2.2 Training

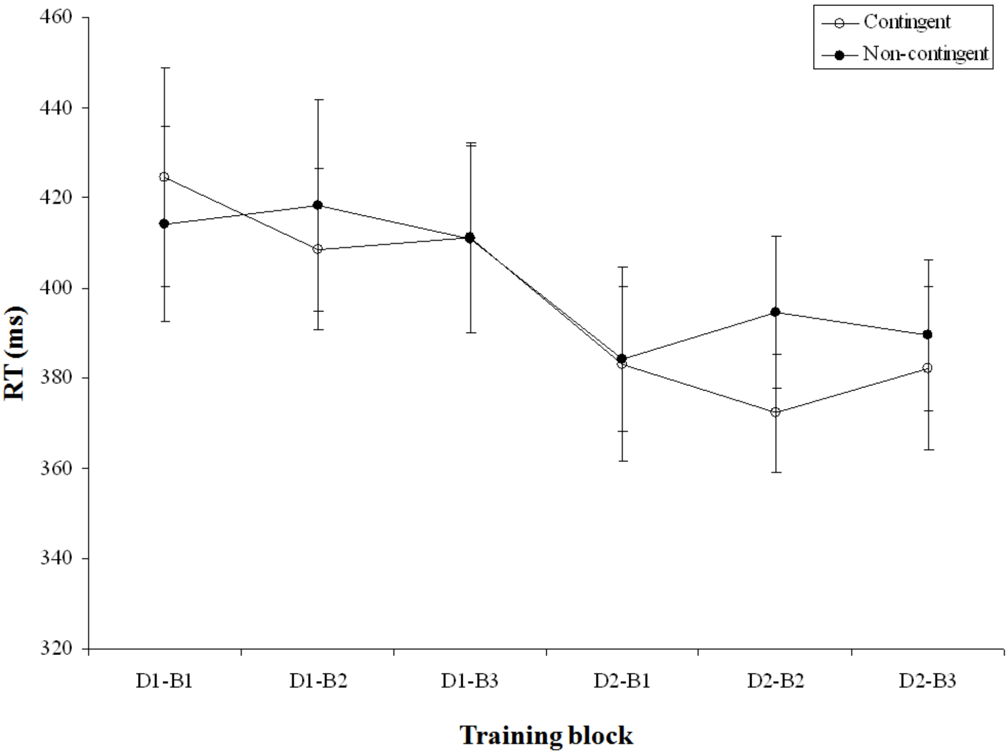

Figure 13: Mean RTs observed in Experiment 5, plotted across the six training blocks where D1-2: day1 to day-2; and B1-B3: block-1 to block-3. Open circles represent the performance of the contingent training group, and filled circles represent the non-contingent group. Error bars represent the standard error of the mean.

Figure 13 shows mean RTs for the contingent and non-contingent groups observed during training. A mixed model ANOVA with training block (B1-B6) as a within subjects factor and training group (contingent, non-contingent) as a between subjects factor revealed a significant linear decline across the six training blocks [$F(1,22) = 8.6$,

$p < .01$, $\eta^2 = .28$] suggestive of learning. No group x block interaction was observed [$F(1,22) = 0.4$, $p > .51$, $\eta^2 = .02$], indicating that the improvement in training task performance was comparable across groups. Equally, no main effect of group was present [$F(1,22) = 0.04$, $p > .84$, $\eta^2 = .00$] implying that overall RTs during the training task were also equivalent. Simple effects analysis revealed that the difference between the groups at day-1 block-1 was not significant [$t(22) = 0.3$, $p > .75$], indicating no difference in baseline performance.

4.2.2.3 Test

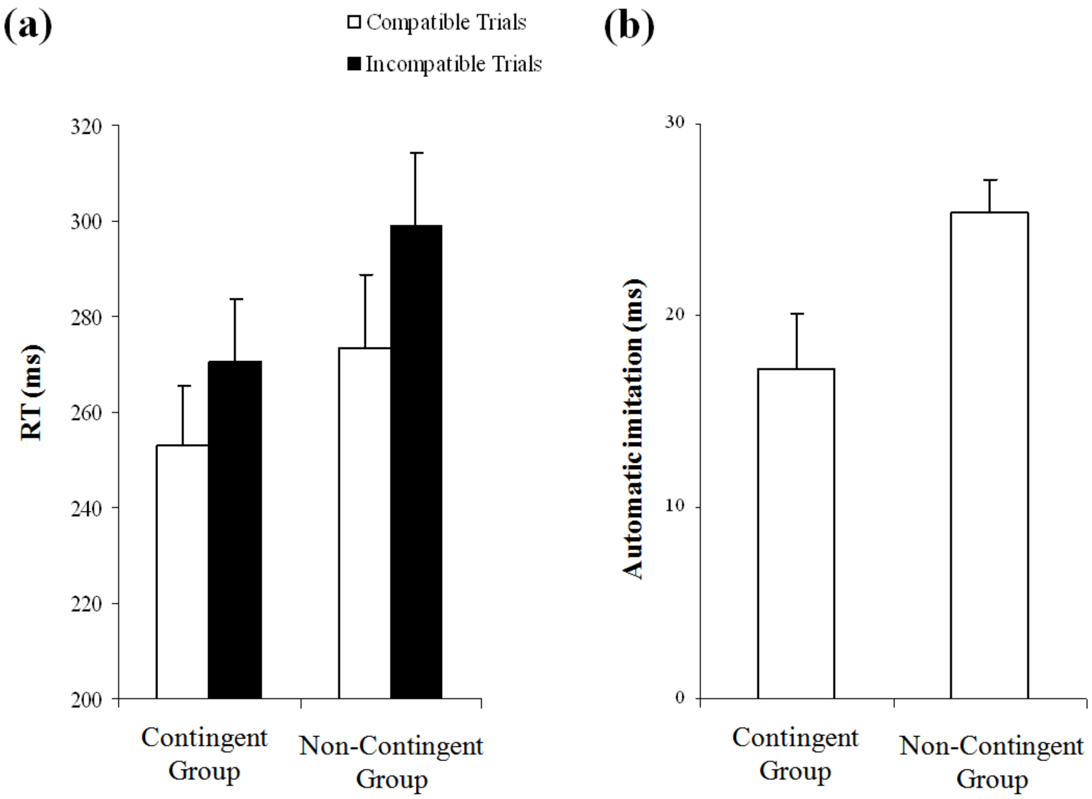

Figure 14: (a) Mean RTs on compatible and incompatible trials during the test phase for the contingent and non-contingent training groups. (b) Residual automatic imitation effects for the contingent and non-contingent training groups calculated by subtracting mean RTs on incompatible trials from mean RTs on compatible trials. Error bars represent the standard error of the mean.

Figure 14 shows (a) the mean RTs on compatible and incompatible trials and (b) the mean automatic imitation effects for the two training groups. These data were analysed using a mixed-model ANOVA with trial type (compatible, incompatible) as a within subjects factor and training group (contingent, non-contingent) as the between subjects factor. The analysis revealed a significant main effect of trial type [$F(1,22) = 159.9$, $p < .001$, $\eta^2 = .88$], indicating that compatible responses ($M = 263.4$ ms, $SD = 48.3$ ms) were executed faster than incompatible responses ($M = 284.7$ ms, $SD = 50.9$ ms). Crucially, the interaction between trial type and training condition was also significant [*F*(1,22) = 5.8, $p < .05$, $\eta^2 = .21$]. This indicates that the effect of trial compatibility (automatic imitation) was greater in the non-contingent training group (*M* = 25.3 ms, *SD* = 5.9 ms) than in the contingent training group (*M* = 17.2 ms, *SD* = 10.1 ms). No main effect of group was observed [*F*(1,22) = 1.5, $p > .23$, $\eta^2 = .06$], indicating that the small disparity in mean RTs between the contingent group (*M* = 261.8 ms, *SD* = 44.5 ms) and non-contingent group (*M* = 286.3 ms, *SD* = 52.8 ms) was not significant.

4.2.2.4 Discussion

Given that all participants received counter-mirror training – training which, if effective, would tend to reduce automatic imitation - the smaller automatic imitation effects found in the contingent group implies that more learning occurred in this group than in the non-contingent group. Associative accounts suggest that, as a result of training, the contingent group acquired stronger incompatible associations (between the open stimulus and close response; and between the close stimulus and open response) than the non-contingent group. Stronger incompatible associations overcome more completely the effects on behaviour of pre-existing, compatible associations, and therefore result in more complete abolition of the automatic imitation effect. Insofar as mirror and counter-mirror links are thought to be conceptually equivalent, these findings suggest that acquisition of the mirror visuomotor links that allow us to solve the correspondence problem is sensitive to sensorimotor contingency.

4.3 Experiment 6

The results of Experiment 5 argue that the acquisition of links connecting visual and motor representations of action is sensitive to sensorimotor contingency. However, the procedure used in Experiment 5 confounded the S-R contingency with the number of times the participants performed the responses. Members of the non-contingent group performed each response twice as often as those in the contingent group. If this additional response repetition caused habituation of the motor representations, the impaired learning observed in the non-contingent group could have been due to enhanced habituation rather than the degradation of the S-R contingency. Counter-mirror learning on paired trials may have been reduced by the additional habituation produced by unpaired trials. The second experiment sought to test this habituation account of the contingency effect observed in Experiment 5. The best evidence that contingency sensitivity in both animal (Dickinson & Charnock, 1985) and human instrumental learning (Shanks, 1989) is not due to habituation comes from studies investigating the effect of ‘signalling’ unpaired trials.

Although the various associative theories explain contingency sensitivity through different mechanisms, they all appeal to the idea that when two or more stimuli are present on a trial, they compete for the acquisition of associative strength (Mackintosh, 1975; Pearce & Hall, 1980; Rescorla & Wagner, 1972; Wagner, 1981). In the case of the counter-mirror paradigm, the target stimuli, the opening and closing of the hand, should have suffered competition from the warning stimulus, the static neutral hand. Because the warning stimulus was present on unpaired trials the associative strength of this warning stimulus for each of the responses should have been enhanced in the non-contingent group. This, in turn, should have augmented the ability of the warning stimulus to compete with the target opening- and closing-hand stimuli on the paired trials, and thereby reduced control over the response by these stimuli in the non-contingent condition.

This competition can be ameliorated, however, by signalling the responses on the unpaired trials with a warning stimulus that differs from that employed on the paired trials. In this signalling condition, the unpaired trials do not enhance learning about the warning signal employed on the paired trials. Therefore, the competition on these trials is effectively the same as that experienced in the contingent condition, without any unpaired trials. By contrast, the habituation account predicts that the impact of unpaired trials should be equivalent in the non-contingent and signalled conditions because the two responses are performed exactly the same number of times in the two conditions.

Experiment 6 examined the contrasting predictions of the habituation and associative accounts by adding a signalling condition in which the warning stimulus on unpaired trials differed from that on paired trials. According to the simple habituation account, the inclusion of unpaired trials should degrade counter-mirror learning as effectively in the signalled condition as in the non-contingent condition. In contrast, associative competition predicts that the presence of a different warning stimulus on the paired and unpaired trials should enable the signalling condition to sustain learning equivalent to that observed in the contingent condition.

4.3.1 Method

4.3.1.1 Participants

A further 36 healthy adults (13 male) with a mean age of 25.4 years served as participants in the experiment in return for a small honorarium. All were right handed, had normal or corrected-to-normal vision, and were naive with respect to the purpose of the experiment. Participants were assigned randomly to either the contingent, non-contingent or signalled training groups in equal numbers. Five were replacements for participants who responded slowly during the test phase (RT > 400ms). None of the participants had taken part in experiments described elsewhere in this thesis.

4.3.1.2 Stimuli

The training stimuli used in the contingent and non-contingent groups were the same as those used in Experiment 5. However, the addition of the signalled training condition required a further four stimuli to act as contextual signals on the unpaired trials. Each of these images depicted a neutral hand (either male or female) coloured blue against a red background (Figure 15). Again, a numerical imperative stimulus (either a 1 or a 2) was superimposed centrally. The unpaired training stimulus was followed by an audible 880-Hz tone, presented for 250 ms immediately following the imperative stimulus. A further two warning stimuli were also employed depicting the male and female hands in blue against a red background. The stimuli used during the test phase were identical to those described in Experiment 5.

Figure 15: Examples of the stimuli used for the signalled training group. Male warning stimulus (left); male irrelevant stimulus with numerical imperative stimulus (right). Hands were coloured blue and presented against a red background.

4.3.1.3 Procedure

The training procedure used for the contingent and non-contingent groups was the same as in Experiment 5. The procedure employed in the signalled condition was virtually identical to that for the non-contingent group, the only difference being the contexts presented on the unpaired trials (Figure 16). In the signalled condition, the warning- and training stimuli depicted in Figure 15 were presented on unpaired trials instead of the neutral hand presented in naturalistic tones, used in the non-contingent condition. As in Experiment 5, whether participants were required to make open or close responses to either 1 or 2 was counterbalanced across training groups, as was the order in which participants completed the two test blocks. The test procedure was identical to that employed in Experiment 5.

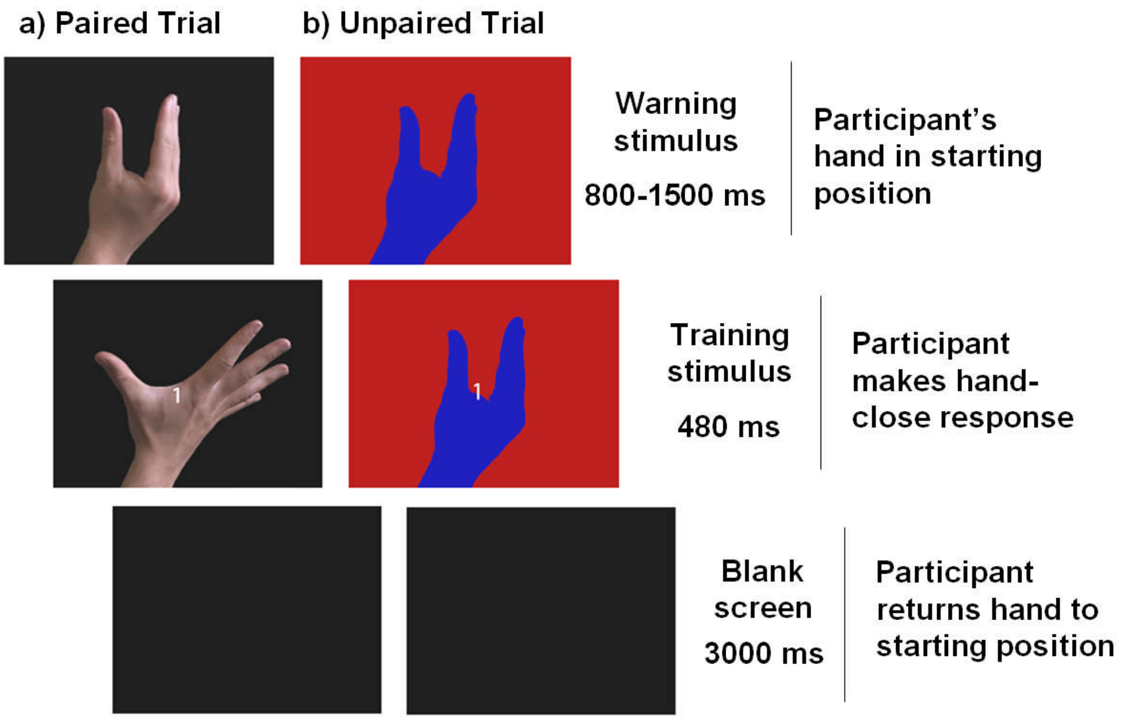

Figure 16: (a) Display sequence for a paired trial employed in the signalled training condition, for participants instructed 1 = close hand; 2 = open hand. (b) Display sequence for an unpaired trial employed in the signalled training condition.

4.3.2 Results and discussion

4.3.2.1 Data handling

Training and test trials where participants made incorrect responses, made no response or where the precise movement onset was equivocal were again excluded from all further analyses (3.3% of training trials; 2.6% of test trials). Any remaining data points beyond 2.5 standard deviations of a participant's mean response latency for a given block were also excluded (2.0% of training trials; 2.1% of test trials). During the test sessions, participants initiated movements on only 7.5% of catch trials.

4.3.2.2 Training

Figure 17 shows mean training RTs for the contingent, non-contingent and signalled groups. A mixed model ANOVA with training block (B1-B6) as a within subjects factor and training group (contingent, non-contingent, signalled) as a between subjects factor revealed a significant linear decline across the six training blocks [*F*(1,22) = 4.3, $p < .05$, $\eta^2 = .12$] suggestive of learning. No group x block interaction was observed [*F*(10,165) = 0.6, $p > .83$, $\eta^2 = .03$] indicating that this improvement in performance was comparable across groups. Similarly, no main effect of group was present [*F*(2,33) = 2.0, $p > .14$, $\eta^2 = .11$] implying that overall RTs during the training task were equivalent. Simple effects analysis revealed that the difference between the groups at day-1 block-1 was not significant [*F*(2,33) = 1.0, $p > .36$], indicating no pre-existing difference in baseline performance.

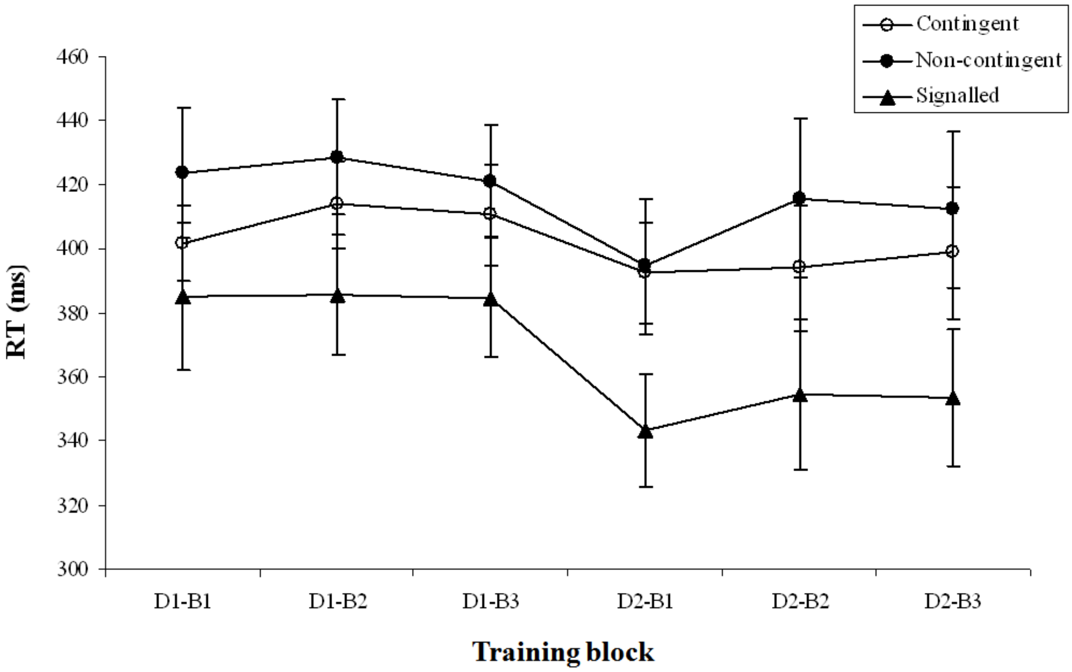

Figure 17: Mean RTs observed in Experiment 6, plotted across the six training blocks where D1-D2: day-1 to day-2, and B1-B3: block-1 to block-3. Open circles represent the performance of the contingent training group; filled circles represent the non-contingent group; and triangles represent the signalled group. Error bars represent the standard error of the mean.

4.3.2.3 Test

Figure 18 shows (a) the mean RTs on compatible and incompatible trials and (b) the mean automatic imitation effects for the three training groups. These data were analysed using a mixed-model ANOVA with trial type (compatible or incompatible) as a within subjects factor and training variant (contingent, non-contingent and signalled) as the between subjects factor. A significant main effect of trial type was observed [*F*(1,33) = 112.7, $p < .001$, $\eta^2 = .77$] indicating that compatible responses (*M* = 246.2 ms, *SD* = 42.3 ms) were executed faster than incompatible responses (*M* = 270.6 ms, *SD* = 48.0 ms). The analyses revealed no significant main effect of group, indicating that the small differences in mean RTs between the contingent (*M* = 262.8 ms, *SD* = 45.8 ms), non-contingent (*M* = 267.4 ms, *SD* = 40.5 ms), and signalled groups (*M* = 245.0 ms, *SD* =

47.7 ms) were not significant. As in Experiment 5, the crucial interaction between trial type and training condition was significant [$F(2,33) = 4.2$, $p < .025$, $\eta^2 = .20$], indicating that the effect of compatibility (automatic imitation) differed across groups. In order to interpret this 2 x 3 interaction, three simple 2 x 2 interactions were calculated. A significant interaction was observed between trial type and training (contingent, non-contingent) [$F(1,33) = 5.2$, $p < .05$] indicating greater automatic imitation in the non-contingent condition ($M = 33.6$ ms, $SD = 16.9$ ms) than in the contingent condition ($M = 20.8$ms, $SD = 12.0$ ms). A significant interaction was also observed between trial type and training (signalled, non-contingent) [$F(1,33) = 7.2$, $p < .025$] indicating greater automatic imitation in the non-contingent condition than in the signalled condition ($M = 18.6$ ms, $SD = 11.8$ ms). However, no significant interaction was observed between trial type and training (contingent, signalled) indicating that automatic imitation effects were comparable for the contingent and signalled conditions.

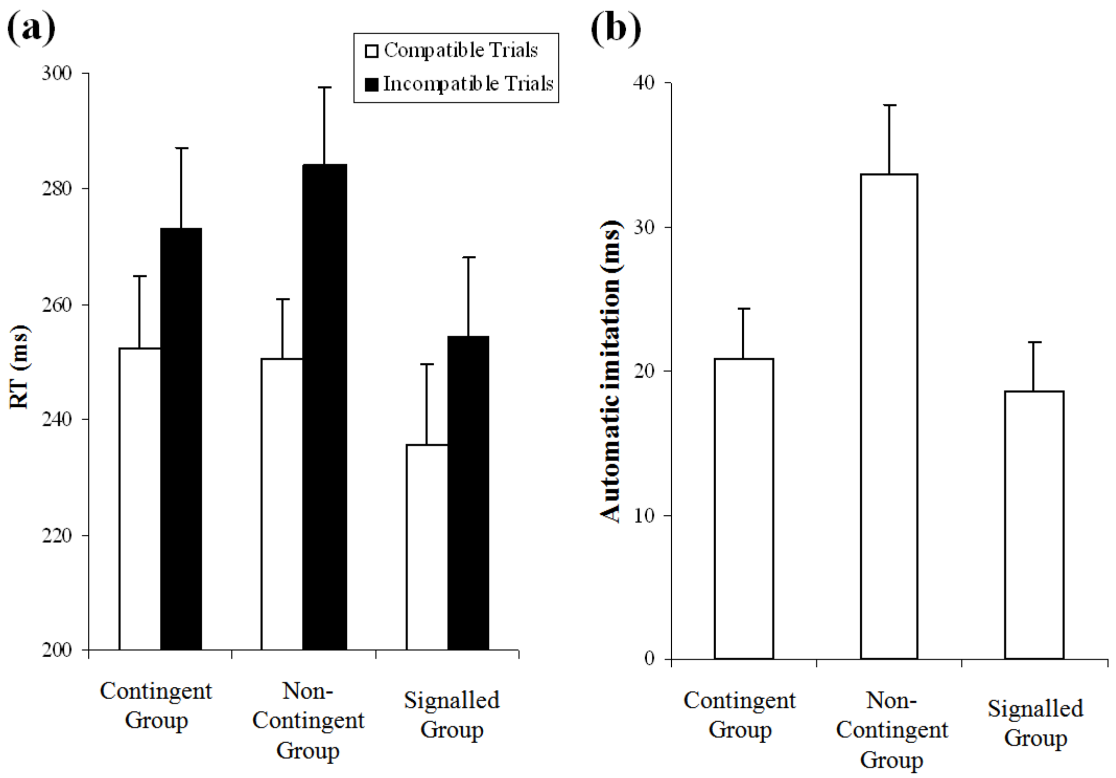

Figure 18: (a) Mean RTs on compatible and incompatible trials during the test phase for the contingent, non-contingent and signalled training groups. (b) Mean automatic imitation effects for the contingent, non-contingent, and signalled training groups. Error bars represent the standard error of the mean.

4.3.2.4 Discussion

This pattern of results replicates that observed in Experiment 5, in that participants given contingent training showed smaller automatic imitation effects, indicative of more counter-mirror learning, than those given non-contingent training. Moreover, the signalled group showed automatic imitation effects equivalent to the contingent training group, but significantly smaller than the non-contingent group. The fact that the contingent and signalled groups showed smaller automatic imitation effects suggests that the participants in these groups learned more during training than the participants in the non-contingent group, thereby acquiring stronger incompatible associations. Thus, when the context of unpaired trials was differentiated from that on paired trials, they did not detract from learning.

That signalling the response on unpaired trials attenuates their detrimental effect suggests that the basic contingency effect observed in Experiment 5, and replicated in Experiment 6, was not a product of response habituation. Rather the contingency effect appears to be due to the type of stimulus competition or ‘overshadowing’ observed in conditioning paradigms. In the non-contingent condition, the unpaired trials served to enhance the associative strength of the warning stimulus, which then competed more effectively with the target opening- and closing-hand stimuli for control over the respective responses. The use of a different warning stimulus on unpaired trials in the signalled condition alleviated this competition and restored learning to the level seen in the contingent condition.

4.4. General Discussion

In Experiment 5, residual automatic imitation effects were measured following either contingent or non-contingent counter-mirror training. Whereas the number of temporally contiguous S-R pairings was held constant across training groups, the S-R contingency was degraded in the non-contingent group through the addition of unpaired trials, in which participants made motor responses in the absence of the counter-mirror stimulus. A contingency effect was observed: participants given non-contingent training showed less counter-mirror learning than those given contingent training. In Experiment 6, this contingency effect was replicated: participants given non-contingent training again showed less counter-mirror learning than those who received contingent training. Experiment 6 also assessed learning in a third ‘signalled’ training group for whom the warning stimulus of the unpaired trials was differentiated from that of the paired S-R trials. The signalled group showed automatic imitation effects equivalent to the contingent training group, but significantly smaller than the non-contingent group. This implies that participants in the signalled group exhibited counter-mirror learning comparable with the contingent group; they acquired stronger links between the open and close motor responses and the incompatible visual stimuli. These links subsequently facilitated faster responding on the incompatible test trials, thus resulting in smaller automatic imitation effects.

The ASL model proposes that the acquisition of visuomotor links is mediated by the same mechanisms of associative learning that mediate conditioning in human and nonhuman animals. It has been known for at least 40 years that these mechanisms depend on contingency (Rescorla, 1968). ASL therefore implies that the acquisition of mirror visuomotor links depends on contingency. Hebbian MN theory also suggests that visuomotor links are acquired through some kind of associative process. However, as the name of this model indicates, it is based on the principles advanced by Hebb (1949), and these principles assume that contiguity is sufficient for associative learning.

The contingency effects observed in Experiments 5 and 6 do not accord with this view. A model based solely on Hebbian principles predicts equivalent learning across training conditions where the number of contiguous S-R pairings is held constant. However, in both Experiments 5 and 6, the contingent group showed greater sensorimotor learning than the non-contingent group, despite both receiving the same number of temporally contiguous S-R pairings during training. This finding suggests that the acquisition of visuomotor links is sensitive, not only to sensorimotor contiguity, but also to sensorimotor contingency.

A purely Hebbian account of the contingency effect observed in Experiment 5 can be sustained if it is assumed that extinction occurs on unpaired trials due to an anti-Hebbian process. This account predicts that signalling the unpaired trials should have no influence on the strength of learning, because, even when such trials are signalled, they involve the execution of responses in the absence of their incompatible visual stimuli. In other words, in spite of signalling, the anti-Hebbian process should continue to act against the learning taking place on the paired trials. However, this possibility was not supported by the signalling effect observed in Experiment 6: the signalled group showed more learning than the non-contingent group. Therefore, it seems that a sensorimotor learning account based solely on Hebbian principles is unable to explain the contingency effects observed.

When applied to the present sensorimotor paradigm, theories of associative learning assume that when two or more stimuli are paired with a response, these stimuli compete for association with the response, either directly (Rescorla & Wagner, 1972; Wagner, 1981) or through attentional processes (Mackintosh, 1975; Pearce & Hall, 1980). Therefore, in the counter-mirror paradigm the warning signal competes with the irrelevant action stimuli for association with the hand-opening and hand–closing responses on the paired trials. In the non-contingent condition, the competition from the warning stimulus on the paired trials is enhanced by the extra associative strength acquired by the warning stimulus on the unpaired trials, thereby reducing the strength of the associations formed between the action stimuli and the responses. This extra source of competition is removed in the signalling conditioning because the warning stimulus on the unpaired trials differs from that on the paired trials.

In summary, the findings from Experiments 5 and 6 indicate that the acquisition of visuomotor links is sensitive to the degree of sensorimotor contingency. This is challenging for the Hebbian hypothesis, because the learning principles advanced by Hebb (1949) state that the strength of learning is determined solely by temporal contiguity. Instead, the findings described in this chapter indicate that visuomotor link acquisition is mediated by the same phylogenetically ancient mechanisms of associative learning that are responsible for conditioning. The findings reported in this chapter represent the first empirical evidence to distinguish between the ASL and Hebbian solutions to the correspondence problem.

Chapter 5: Second-learned visuomotor links are modulated by context

5.1 Introduction

The aim of the experiments reported in this chapter was to provide a further test of the ASL model, by determining whether counter-mirror learning demonstrates the same context-specific properties as counter-conditioning. The associative literature – and thus by extension the ASL model - makes a number of testable predictions regarding the acquisition of visuomotor links, that might serve to distinguish the learning mechanisms hypothesised by ASL from the simple Hebbian account of MN origins. However, the context-specificity of learning was chosen for this purpose, in order to better understand previous reports of contextual modulation of the MNS (Bonini et al., 2010; Fogassi et al., 2005; Iacoboni et al., 2005). Through the study of conditioning phenomena, associative principles have been developed which may explain these reports. In contrast, the Hebbian account of MN origins, in its current form, offers no explanation of the modulation of visuomotor links by action context.

Counter-conditioning is an associative learning procedure in which responses to a stimulus are reversed across successive training phases. Typically human or animal participants are first placed on a training schedule where they learn that a stimulus signals one outcome (e.g. a tone warns of the imminent delivery of mild electric shock). They are then placed on a second conditioning schedule where the same stimulus predicts a different outcome (e.g. the same tone signals the delivery of food). Interestingly, an asymmetry is often observed between the learning that occurs in the first and second phases. Whereas the conditioning that occurs in the first phase is relatively insensitive to changes in context, the learning that occurs in the second phase is often far more context-specific. For example, rats trained on an initial conditioning schedule in context-A and a second counter-conditioning schedule in context-B, will exhibit the original conditioned response when returned to context-A and when placed in a novel context-C (Peck & Bouton, 1990). Thus, counter-conditioning experiments have shown that, when a stimulus predicts different outcomes in sequential phases of a training experiment, 'second-learned' associations are subject to greater contextual control than the original 'first-learned' associations acquired when the stimulus is novel (Bouton, 1993, 1994; Nelson, 2002).

A similar principle has been demonstrated in studies of renewal. Participants are first placed in context-A where they learn that a stimulus is predictive of a certain outcome. Once the initial association has been acquired, participants are transferred to context-B, where they learn that the stimulus no longer predicts the outcome. However, the extinction learning that takes place during this second phase is subject to contextual control: When transferred back to context-A (“ABA design”) or placed in a novel context-C (“ABC design) a renewal of responding occurs, whereby participants exhibit the conditioned response acquired during the first phase (Bouton & King, 1983; Nelson, Sanjuan Mdel, Vadillo-Ruiz, Perez, & Leon, 2011).

Applying the terminology used in these studies of counter-conditioning and renewal, the ASL model implies that the correspondence problem is solved by first-learned ‘mirror’ associations. Each of these associations connects a sensory representation of an action with a motor representation of the same action, and is acquired through everyday experience of a predictive relationship between observation and execution of the same action. The ASL model further implies that, during counter-mirror training, participants acquire a set of second-learned ‘counter-mirror’ associations, connecting sensory representations of actions with motor representation of different actions. Therefore, if the ASL model is correct in suggesting that the correspondence problem is solved by the same phylogenetically ancient mechanisms of associative learning that mediate conditioning, counter-mirror learning, like counter-conditioning, should be context-specific. In contrast the Hebbian account of MN origins, in its current form, offers no mechanism through which contextual modulation may be achieved and therefore predicts that counter-mirror learning ought to manifest equally in both trained and untrained contexts.

These predictions were tested in Experiment 7, where automatic imitation was measured in red and blue contexts, both before and after two sessions of counter-mirror training completed in either red or blue contexts. The purpose of Experiment 8 was to confirm that any context-specificity observed in Experiment 1 was due, as the ASL model suggests, to conflict between pre-experimental learning and counter-mirror training, and not due to the use to the use of the pre-test procedure.

5.2 Experiment 7

The first experiment tested whether counter-mirror learning, like counter-conditioning, is context-specific. In the first of four sessions, two separate pre-tests were conducted to establish participants’ baseline automatic imitation effects in distinctive red and blue contexts. During the second and third sessions, participants received counter-mirror training in either the red or the blue contexts. In the final session two separate post-tests were conducted to establish participants' residual automatic imitation effects in both the red and blue contexts. Context-specificity was indexed by comparing the magnitude of the pre- to post-test reduction in automatic imitation when participants were tested in the context in which they received counter-mirror training and in their untrained context. If the ASL model is correct, counter-mirror learning, like counter-conditioning, should show context-specificity due to the conflict between the first-learned mirror associations and the second-learned counter-mirror associations.

5.2.1 Method

5.2.1.1 Participants

Sixteen healthy adults (five males) with a mean age of 22.4 years served as participants in the experiment in return for a small honorarium. All were right handed, had normal or corrected-to-normal vision, and were naive to the purpose of the experiment. Participants were assigned randomly to either the train-red or train-blue groups in equal numbers. None of the participants had taken part in experiments described elsewhere in this thesis.

5.2.1.2 Stimuli

The stimuli used in Experiment 7 consisted of 12 digital images of a model's right hand in naturalistic tones (Figure 19). Six of the images were presented against a red background, and six against a blue background. Each colour set comprised images of male and female neutral hands; male and female closed hands; and male and female open hands. The open stimulus showed fingers and thumb splayed; the closed stimulus depicted a fist; and in the neutral hand stimulus fingers were shown together, pointing upwards in parallel with the thumb. The size of the male and female stimuli were identical to those used in Experiments 5 and 6 (section 4.2.1.2).

5.2.1.3 Contexts

Four elements were varied to provide two distinctive 'red' and 'blue' contexts. As described above, hand stimuli were presented against either red or blue backgrounds. In addition, the testing cubicle was lit by either red or blue light, provided by a Eurolite PAR-38 RGB LED spotlight. The experimental procedure required the use of an armrest to support the participants' hands and an occluder to prevent them observing their own responses. Distinctive armrests and occluders were constructed for use in the red and blue contexts. The occluders were identical in all features except their colour (either red or blue). The armrests were identical in size (15cm x 35cm x 25cm) but differed in both colour and surface texture. The red armrest was covered in a coarse red woollen fabric. The blue armrest was covered with a fine-grain plastic material. Display backgrounds (Nelson et al., 2011) and lighting changes (Van Gucht, Vansteenwegen, Beckers, & Van den Bergh, 2008; Vansteenwegen et al., 2005) have been shown to be effective context manipulations in single-room procedures.

5.2.1.4 Data recording and analysis

Both the training and test procedures took the form of RT tasks in which EMG was used to establish response onset. The recording and analysis procedures were the same as described in chapter 4 (section 4.2.1.3).

5.2.1.5 Procedure

The experiment was conducted over four sessions each completed approximately 24 hr apart. During the first session participants completed two test procedures to establish their baseline automatic imitation effects in the red and blue contexts. In sessions two and three, participants completed six blocks of counter-mirror training spread evenly over the two days, in either the red or blue context. In the final session, participants' residual automatic imitation effects were again measured in the red and blue contexts. The order in which participants completed the red and blue tests was counterbalanced, but was held constant across sessions one and four. Participants were seated in an identical manner to that described in Chapter 4. The use of a pre-test represents a departure from the design employed in Experiments 5 and 6. Measuring reductions in automatic imitation between pre- and post-test was intended to take account of individual differences in existing automatic imitation effects, and therefore provide a more sensitive measure of counter-mirror learning.

The test procedure was a simple RT task (Figure 19). Tests comprised two blocks of 80 trials during which participants made speeded pre-specified movements (hand-open responses in one block, hand-close in the other) in response to the onset of open and closed stimuli. Fewer test trials were included in each block than in Experiments 5 and 6 (where blocks comprised 120 trials) due to the need to complete two test procedures –

one in each context – rather than a single test in one context. Each test trial started with the participant's hand in the neutral starting position, and with a neutral hand warning stimulus on the screen. Thereafter participants were required to make the pre-specified response as soon as an action stimulus appeared, before returning to the neutral position ready for the next trial. The temporal parameters of each test trial were identical to those described in Chapter 4 (section 4.2.1.4). The order in which participants completed the open and close blocks was counterbalanced across groups. Twenty catch trials were again included in each test block, in which the warning stimulus was displayed throughout the trial during which the participants had been instructed to make no response whatsoever.

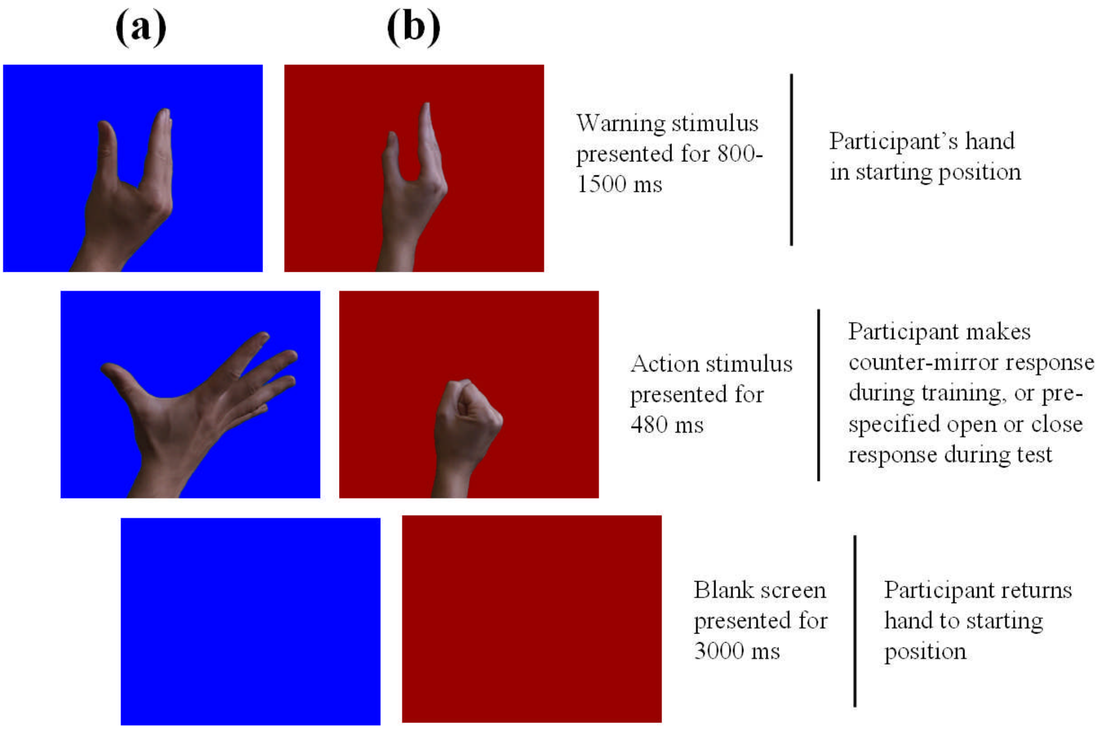

Figure 19: Display sequences for (a) male-hand open-stimulus trials in the blue context and (b) female-hand close-stimulus trials in the red context. During training participants made the counter-mirror response to the onset of the action stimulus (open responses to close stimuli; close responses to open stimuli). On test trials participants made pre-specified responses to the onset of the action stimulus; open-hand responses in one block, close-hand responses in the other.

The training procedure took the form of a choice RT task (Figure 19). Each trial required the participant to make either an open response to a close-hand stimulus or a close response to an open-hand stimulus. Having made each response, participants returned to the starting position ready for the next trial. In a departure from the training procedure used in Chapter 4, participants responded to the action itself, and not to a superimposed imperative stimulus. In Chapter 4 it was necessary to use the numerical stimulus in order to vary the contingency between the action stimuli and the responses.

However, using the action stimulus as the imperative stimulus is preferable as an additional imperative stimulus may detract from counter-mirror learning through associative competition (overshadowing). The beginning of each trial was indicated by the appearance of the warning stimulus, which was then replaced by one of the action stimuli, presented for 480 ms. SOA was varied randomly between 800 ms and 1,500 ms in 50 ms increments. The hand depicted in the warning stimulus was identical to that in the subsequent action stimulus, giving rise to apparent motion. Following the offset of the action stimulus, the screen went blank for 3,000 ms, until the warning stimulus for the subsequent trial was presented. Each training block comprised 144 counter-mirror training trials. Stimulus order was randomised during the training trials. Half of the participants completed the counter-mirror training task in the red context and half in the blue context.

5.2.2 Results and Discussion

5.2.2.1 Data handling

Training and test trials in which participants made incorrect responses, no response, or where the point of movement onset was equivocal were excluded from all further analyses (2.8% of training trials and 4.1% of test trials). Thereafter, any remaining data points beyond 2.5 standard deviations of a participant's mean response latency for a given block were also excluded (2.0% of training trials and 2.9% of test trials). EMG signals on catch trials were examined to ensure that the participants obeyed task instructions, but were excluded from all further analyses. During the test sessions, participants initiated movements on only 5.4% of catch trials, indicating that they were not making anticipatory responses on the stimulus trials.

5.2.2.2 Training

Figure 20 shows mean RTs for the groups trained in red and blue contexts. These training data were analysed using a mixed-model ANOVA with training block (B1-B6) as a within subjects factor and training group (train-red, train-blue) as a between subjects factor. Trend analysis revealed a highly significant linear decline across the six training blocks [$F(1,14) = 19.83$; $p < .001$; $\eta^2 = .59$] suggestive of learning. This trend did not vary as a function of group [$F(1, 14) = .02$; $p > .80$; $\eta^2 = .00$] indicating that the improvement in training task performance was comparable. Although Figure 20 suggests that responses were faster in the red context than in the blue, no main effect of group was present [$F(1, 14) = 2.33$; $p > .14$; $\eta^2 = .14$], implying that overall RTs during the training task were broadly equivalent. Simple effects analysis revealed that the difference between the groups at Day 1, Block 1 was not significant [$t(14) = .87$; $p > .40$] indicating that baseline performance was comparable. No higher order trends or other trend x group interactions were observed.

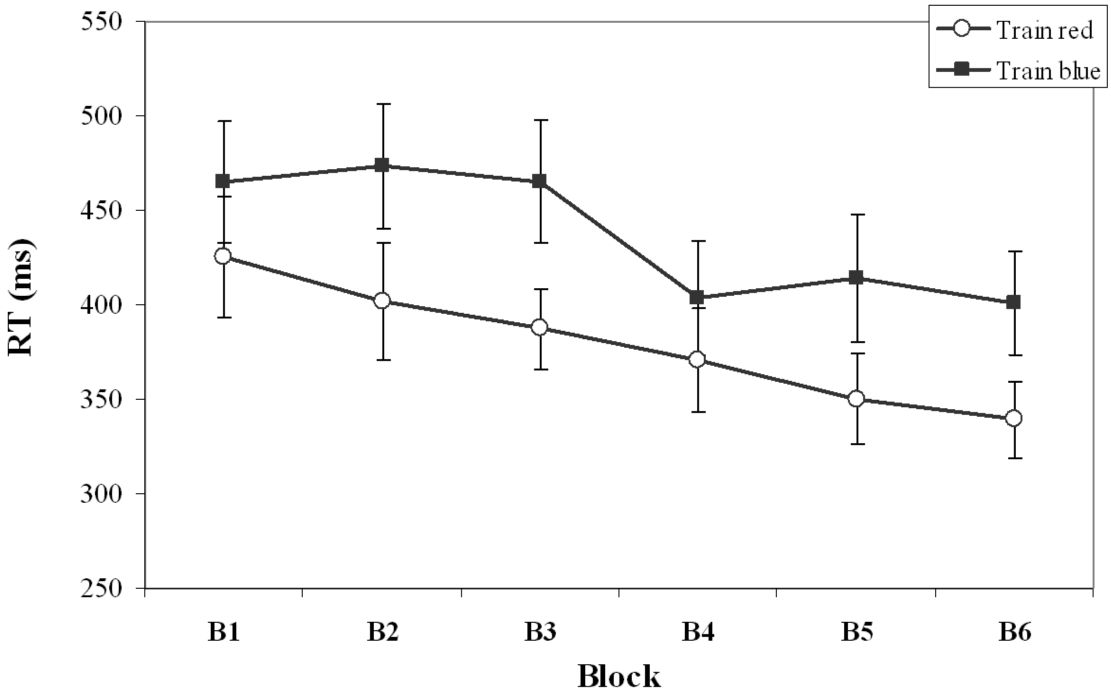

Figure 20: Mean RTs observed during training in Experiment 7, plotted across the six training blocks. Open circles represent the performance of the group trained in the red context, and filled squares represent the group trained in the blue context. Error bars represent the standard error of the mean. RT = reaction time; B = block.

5.2.2.3 Tests

Figure 21a shows the mean RTs observed at pre- and post-test, in the trained and untrained contexts, on compatible and incompatible trials. Figure 21b depicts the mean automatic imitation effects observed on the four tests (pre-test in the trained context; pre-test in the untrained context; post-test in the trained context; post-test in the untrained context). The magnitude of each automatic imitation effect was calculated by subtracting the mean RT on compatible trials (open responses to open stimuli; close responses to close stimuli) from the mean RT on non-compatible trials (open responses to close stimuli; close responses to open stimuli).

The RT data were analyzed using a mixed-model ANOVA with test (pre-test, post-test), context (trained, untrained) and S-R compatibility (compatible, incompatible) as within subjects factors and group (train-red, train-blue) as a between subjects factor. The analysis revealed a highly significant main effect of S-R compatibility [$F(1,14) = 98.23$;

$p < .001$; $\eta^2 = .88$] whereby participants were slower to make incompatible responses (*M* = 303.4ms; *SD* = 71.6ms) than compatible responses (*M* = 282.2ms; *SD* = 68.9ms) indicative of automatic imitation. In addition, a marginally significant test x compatibility interaction was observed [*F*(1,14) = 3.94; $p = .067$; $\eta^2 = .22$] indicating that automatic imitation effects were generally smaller at post-test (*M* = 16.3ms; *SD* = 15.1ms) than at pre-test (*M* = 26.2ms; *SD* = 16.4ms). Crucially, this test x compatibility interaction varied as a function of context [*F*(1,14) = 7.46; $p < .025$; $\eta^2 = .35$]. Simple effects analysis indicated that there was a significant reduction in automatic imitation between the pre-test (*M* = 26.9 ms; *SD* = 14.5 ms) and post-test (*M* = 12.6 ms; *SD* = 15.3 ms) when tested in the trained context [*t*(15) = 2.89; $p < .025$]. In contrast, the reduction in automatic imitation between the pre-test (*M* = 25.5 ms; *SD* = 18.6 ms) and post-test (*M* = 20.1 ms; *SD* = 14.5 ms) when tested in the untrained context was not significant [*t*(15) = 1.05; $p > .30$]. This test x context x compatibility interaction did not vary as a function of group [*F*(1, 14) = .40; $p > .5$; $\eta^2 = .03$].

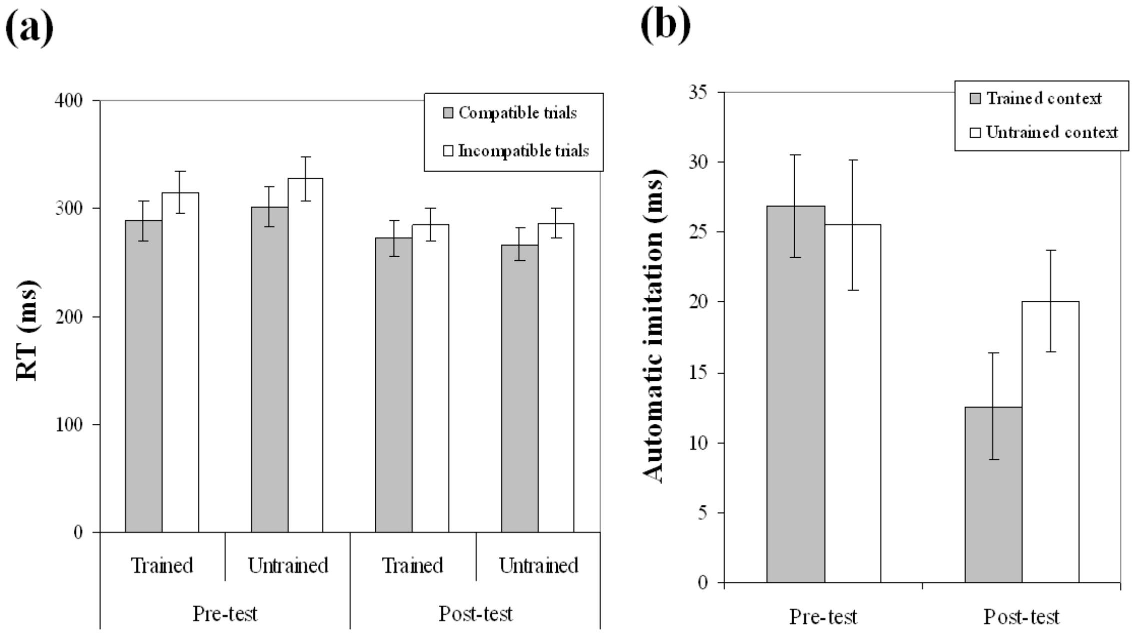

Figure 21: (a) Mean RTs observed at pre-test and post-test, in the trained and untrained contexts, on compatible and incompatible trials and (b) automatic imitation effects observed at pre-test and post-test in the trained and untrained contexts, calculated by subtracting mean RTs on compatible trials from mean RTs on incompatible trials. Error bars represent the standard error of the mean. RT = reaction time.

In addition to the effects of principal interest described above, a significant group x compatibility interaction was observed [*F*(1,14) = 10.25; $p < .025$ $\eta^2 = .42$]. Those participants given counter-mirror training in the red context generally showed greater automatic imitation (*M* = 28.1 ms; *SD* = 14.5 ms) than those trained in the blue context (*M* = 14.4 ms; *SD* = 15.5 ms). A marginally significant main effect of test was also observed [*F*(1,14) = 3.94; p = .067; η^2 = .22] indicating that participants generally responded faster at post-test (*M* = 277.5 ms; *SD* = 60.5 ms) than at pre-test (*M* = 308.2 ms; *SD* = 77.3 ms). None of the other main effects or interactions approached significance.

5.2.2.4 Discussion

The results of Experiment 7 indicate that greater reductions in automatic imitation were seen when participants were tested in their trained context than in their untrained context. This finding suggests that counter-mirror learning is similar to counter-conditioning, in that both show context-specificity (Bouton & Peck, 1992). Associative accounts argue that counter-conditioning comes under contextual control in order to resolve ambiguity (Bouton, 1993, 1994; Nelson, 2002). Rather than simply overwriting the original learning, there is considerable evidence that subsequent conditioning to the same stimulus sets up parallel, second-learned associations (Bouton & Peck, 1992; Brooks, Hale, Nelson, & Bouton, 1995). However, the conflict between the first- and second-learned associations renders the stimulus ambiguous: it is not clear what the appropriate behaviour is in the presence of the stimulus. To resolve this ambiguity, the excitability of both sets of associations is modulated by the context. Second-learned associations become active only in the trained context, whereas first-learned associations are inhibited by the cues present in the trained context, but remain active in all other contexts.

5.3 Experiment 8

The ASL model proposes that the correspondence problem is solved by visuomotor associations acquired during development, e.g. through self-observation, synchronous activity and while being imitated. This model therefore implies that the context-specificity observed in Experiment 7 was due to conflict between the second-learned counter-mirror associations established by experimental training, and first-learned mirror associations acquired long before the experiment began.

However, this is not the only potential interpretation of the context-specificity observed in Experiment 1. The conflict-generating learning could have taken place, not during normal development, but during the pre-test. At pre-test, participants made pre-specified open- and close-hand responses with equal frequency to the onset of both open- and close-hand stimuli in a simple RT task. There was thus no contingency between the action stimuli observed and the responses executed by participants, with the action stimulus serving simply as a go-signal. However, during the counter-mirror training sessions participants were transferred to a choice RT task where the same action stimuli predicted perfectly execution of the counter-mirror responses. During training the onset of the action stimulus not only served as a go-signal but also indicated which response participants should execute. Therefore, it could be argued that the 'meaning' of the action stimuli differed between the pre-test and training procedures (Bouton, 1994), and that it was this conflict, rather than the conflict between counter-mirror associations and pre-experimental learning, that promoted context-specific learning in the training phase.

Experiment 8 sought to distinguish these two accounts of the context-specificity observed in Experiment 7, by testing whether the same procedure could render first-learned associations to a novel stimulus context-specific. Novel S-R learning should generalise well to untrained contexts because there is no conflict with pre-experimental learning, and consequently the stimuli remain unambiguous. Evidence of context-specific learning in Experiment 8 would therefore indicate that the context-specificity observed in Experiment 7 was not due to conflict with pre-experimental learning, and may instead be a product of the pre-test procedure.

We used a four session design, identical to that employed in Experiment 7, to train arbitrary S-R mappings with abstract geometric shapes. Participants again made open- and close-hand responses, but this time to the onsets of geometric trapezoid forms. In the first session they completed simple RT pre-test procedures in the red and blue contexts to confirm the absence of any pre-existing compatibility effects. They then completed two training sessions with a choice RT task, either in the red or blue contexts, where they learned to make open and close responses to the onset of top- and bottom-heavy trapezoids, respectively. In the final session they were tested again in both the red and blue contexts to determine the magnitude of their newly acquired SRC effect. Due to the lack of conflict with prior learning, it was predicted that the S-R learning would manifest equally in both the trained and untrained contexts.

5.3.1 Method

5.3.1.1 Participants

A further 16 healthy adults (4 males) with a mean age of 21.9 years served as participants in the experiment in return for a small honorarium. All were right handed, had normal or corrected-to-normal vision, and were naive to the purpose of the experiment. Participants were assigned randomly to either the train-red or train-blue training groups in equal numbers. None of the participants had taken part in experiments described elsewhere in this thesis.

5.3.1.2 Stimuli

The stimuli used in Experiment 8 consisted of 12 digital images of grey quadrilaterals (Figure 22). Two sets of lighter and darker stimuli, each with different aspect ratios were employed to parallel the pigmentation and scale variation in the male and female hand stimuli. Six of the images were presented against a red background, and six against a blue background. Each colour set included two rectangles, one lighter and one darker, and four Isosceles trapezoids bisected by a horizontal black line, one top-heavy in lighter tones, one top-heavy in darker tones, one bottom-heavy in lighter tones and one bottom-heavy in darker tones. The lighter grey rectangle (160 on a 0-255 scale) had an aspect ratio of 1:1.23 and subtended approximately 12° horizontally. The darker rectangle (140 on a 0-255 scale) had an aspect ratio of 1:1.57 and subtended approximately 9° horizontally. The lighter grey trapezoids subtended 14° at the wider end (170 on a 0-255 scale) and 10° at the narrower end on a (150 on a 0-255 scale). The darker grey trapezoids subtended 11° at the wider end (150 on a 0-255 scale) and 7° at the narrower end on a (130 on a 0-255 scale).

5.3.1.3 Procedure

As in Experiment 7, the order in which participants completed the red and blue tests was counterbalanced, but held constant across pre- and post-test. The test procedure employed in both contexts again took the form of a simple RT task. Participants made speeded pre-specified movements (open responses in one block, close in the other) in response to the onset of top-heavy and bottom-heavy trapezoid stimuli (Figure 22). Each test trial started with the participant’s hand in the neutral starting position, and with a rectangular warning stimulus on the screen. Thereafter participants were required to make the pre-specified response as soon as a trapezoid stimulus appeared and then return to the neutral position ready for the next trial. The rectangle-trapezoid display sequence gave rise to apparent motion, either of the top or bottom of the rectangle moving forwards, depending on the trapezoid presented. The trapezoid stimulus was present for 480 ms until replaced by a blank display for 3,000 ms, prior to the warning stimulus for the next trial. SOA was varied randomly between 800 ms and 1,500 ms in 50 ms increments. Twenty catch trials were included in each test block. On catch trials, the warning stimulus was presented for 1,980 ms before the 3,000 ms inter-trial interval. The order in which participants completed the open and close blocks was counterbalanced across groups. The test procedure was therefore identical to that employed in Experiment 7.

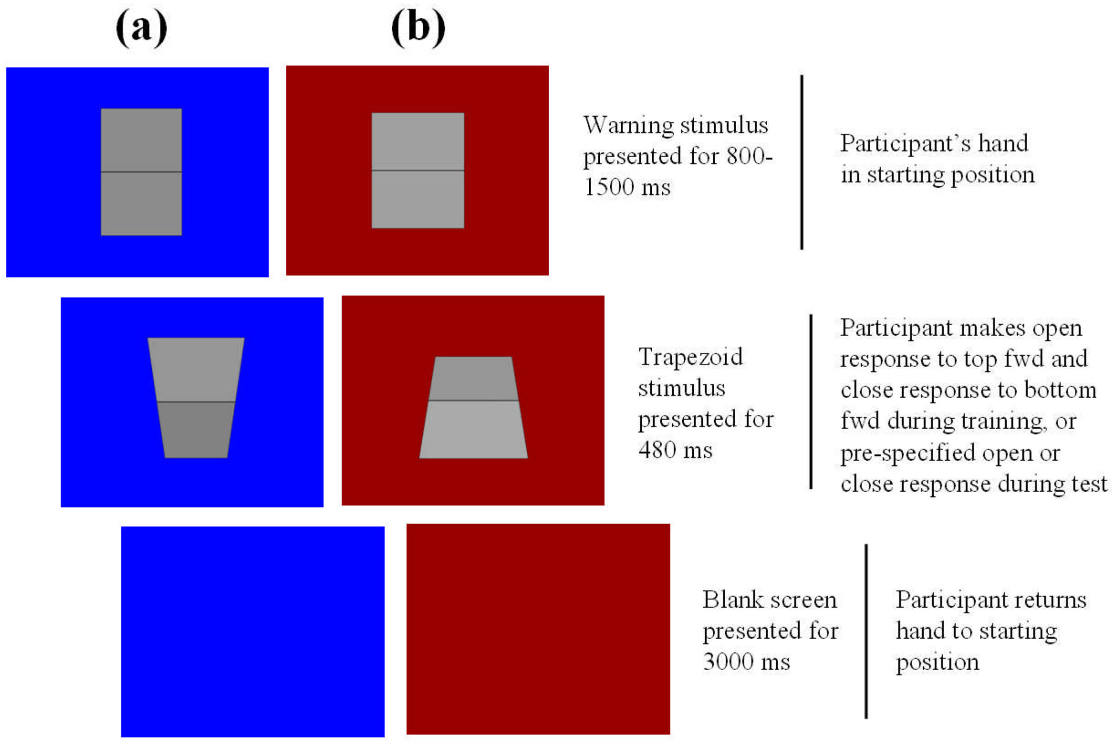

Figure 22: Display sequences for (a) larger aspect ratio top-heavy sequence in the blue context and (b) smaller aspect ratio bottom-heavy sequence in the red context. In each case the presentation of a top- or bottom-heavy trapezoid gave the appearance of the rectangle either falling towards or away from the participant. During training participants made open-hand responses when the top appeared to move forwards and close-hand responses when the bottom appeared to move forwards. On test trials participants made pre-specified responses to the onset of the trapezoid stimulus; open-hand responses in one block, close-hand in the other.

The training procedure again took the form of a choice RT task (Figure 22). Each trial required the participant to make either an open response to a top-forwards stimulus or a close response to a bottom forwards stimulus. To ensure that the design and analyses were comparable with those of Experiment 7, this arbitrary S-R mapping was not counterbalanced. The order of stimulus presentation during the training trials was randomized. The beginning of each trial was indicated by the appearance of the rectangle warning stimulus, which was then replaced by one of the trapezoid stimuli, presented for 480ms. SOA was varied randomly between 800 ms and 1,500 ms in 50 ms increments. Following the offset of the trapezoid stimulus, the screen went blank for 3,000 ms, until the warning stimulus for the subsequent trial was presented. Each training block comprised 144 training trials. Half of the participants completed the training task in the red context and half in the blue context.

5.3.2 Results and Discussion

5.3.2.1 Data handing

Training and test trials in which participants made incorrect responses, no response, or where the point of movement onset was equivocal were excluded from all further analyses (3.4% of training trials and 4.5% of test trials). Thereafter, any remaining data points beyond 2.5 standard deviations of a participant's mean response latency for a given block were also excluded (3.0% of training trials and 3.3% of test trials). EMG signals on catch trials were examined to ensure that participants obeyed task instructions, but were excluded from all further analyses. During the test sessions, participants initiated movements on only 5.7% of catch trials, indicating that they were not making anticipatory responses on the stimulus trials.

5.3.2.2 Training

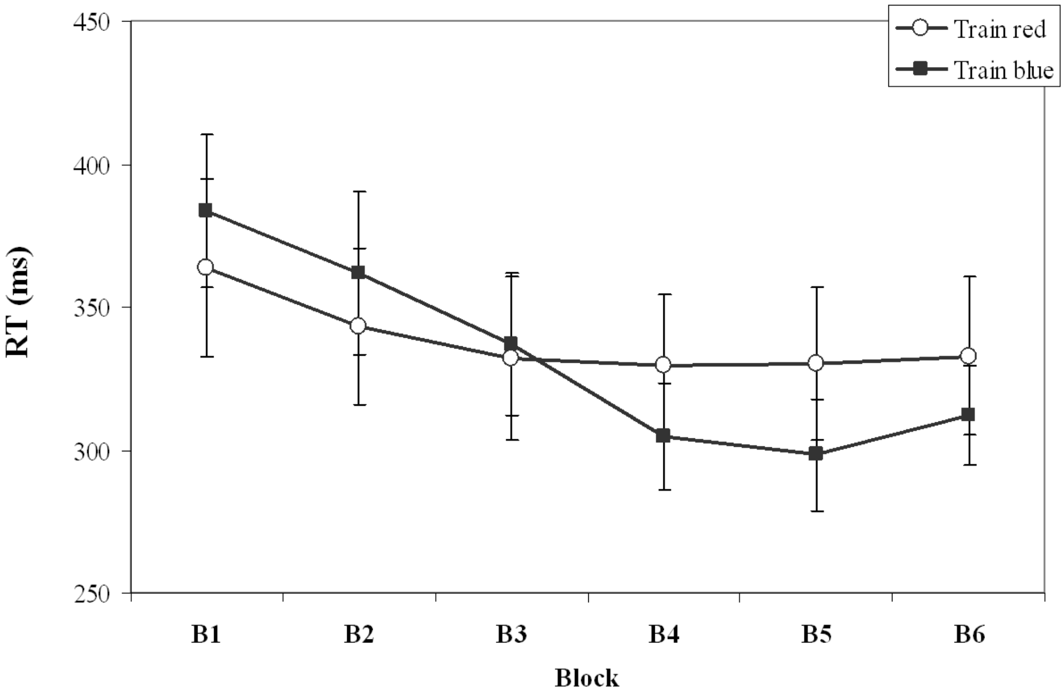

Figure 23: Mean RTs observed during training in Experiment 8, plotted across the six training blocks. Open circles represent the performance of the group trained in the red context, and filled squares represent the group trained in the blue context. Error bars represent the standard error of the mean. RT = reaction time; B = block.

Figure 23 shows mean RTs for the groups trained in red and blue contexts. A mixed-model ANOVA with training block (B1-B6) as a within subjects factor and training group (train-red, train-blue) as a between subjects factor, revealed a highly significant linear decline across the six training blocks [*F*(1,14) = 14.19; *p* < .01; η^2 = .50] suggestive of learning. The group x linear trend interaction failed to reach significance [*F*(1, 14) = 3.47; *p* > .08; η^2 = .19] indicating that the linear decline was broadly comparable across groups. Trend analysis also revealed a significant quadratic trend [*F*(1,14) = 14.07; *p* < .01; η^2 = .50]. However, this did not vary as a function of group [*F*(1, 14) = .66; *p* > .40; η^2 = .05]. No main effect of group was present [*F*(1, 14) = .03; *p* > .80; η^2 = .00], implying that overall RTs during the training task were also broadly equivalent. Simple effects analysis revealed that the difference between the groups at Day 1, Block 1 was not significant [*t*(14) = .49; *p* > .60] indicating that baseline performance was comparable.

5.3.2.3 Tests

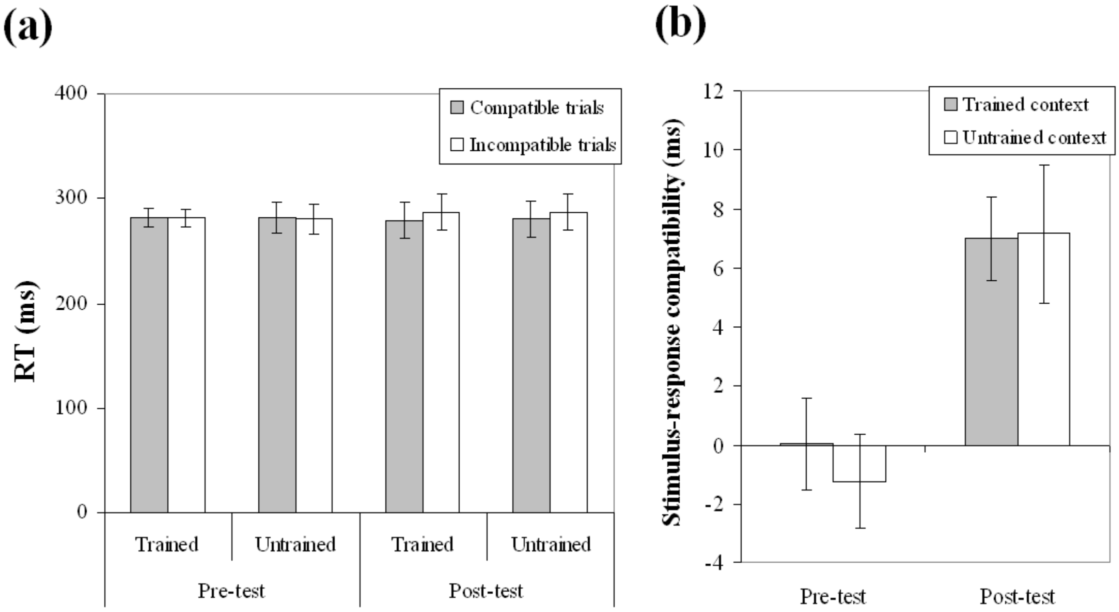

Figure 24: (a) Mean RTs observed at pre-test and post-test, in the trained and untrained contexts, on compatible and incompatible trials and (b) S-R compatibility effects observed at pre-test and post-test in the trained and untrained contexts, calculated by subtracting mean RTs on compatible trials from mean RTs on incompatible trials. Error bars represent the standard error of the mean. RT = reaction time.

Figure 24 shows the mean RTs observed at pre- and post-test, in the trained and untrained contexts, on compatible and incompatible trials. Figure 24 also depicts the mean compatibility effects seen on the four tests (pre-test in the trained context; pre-test in the untrained context; post-test in the trained-context; post-test in the untrained context). Because the trained S-R mappings were open to top forwards and close to bottom forwards, the size of each SRC effect was calculated by subtracting the mean RT on compatible trials (open to top forwards; close to bottom forwards) from the mean RT on incompatible trials (open to bottom forwards; close to top-forwards). The greater the RT difference, the stronger the participant's tendency to execute the trained response in the presence of the test stimuli. These SRC effects are conceptually equivalent to the automatic imitation effects calculated in Experiment 7.

The RT data were analyzed using a mixed-model ANOVA with test (pre-test, post-test), context (trained, untrained) and S-R compatibility (compatible, incompatible) as within subjects factors and group (train-red, train-blue) as a between subjects factor. The analysis revealed a significant main effect of S-R compatibility [$F(1,14) = 12.60$; $p < .01$; $\eta^2 = .47$] whereby participants were faster to make responses compatible with the trained mapping ($M = 280.8$ ms; $SD = 52.3$ ms) than incompatible responses ($M = 284.1$ ms; $SD = 51.9$ ms) when collapsed across test. In addition, a significant test x compatibility interaction was observed [$F(1,14) = 9.37$; $p < .01$; $\eta^2 = .40$] indicating that S-R compatibility effects were larger at post-test ($M = 7.08$ ms; $SD = 6.62$ ms) than at pre-test ($M = -0.61$ ms; $SD = 5.43$ ms). Crucially, this test x compatibility interaction did not vary as a function of context [$F(1, 14) = .27$; $p > .60$; $\eta^2 = .02$] indicating that the change in S-R compatibility was equivalent in trained and untrained contexts. Moreover, the test x compatibility x context interaction was comparable in both groups [$F(1, 14) = .18$; $p > .65$; $\eta^2 = .01$]. No other main effects or interactions approached significance.

5.3.2.4 Discussion

The results from Experiment 8 show that novel S-R learning generalizes to untrained contexts. In contrast, the counter-mirror learning observed in Experiment 7 was found to be context-specific; the associations acquired manifested more strongly in the trained than in the untrained context. Given that identical designs were used in the two experiments, this contrast indicates that the context-specificity observed in Experiment 7 was not an artefact of pre-test procedure. Rather, the results of Experiment 8 accord with the hypothesis that, in Experiment 7, context-specificity was generated by a conflict between second-learned counter-mirror associations and first-learned mirror associations established during ontogeny.

5.4 General Discussion

The present study tested the hypothesis, advanced by the ASL model, that when the sight of an action is associated with rival mirror and counter-mirror responses, this ambiguity is resolved through contextual modulation – the same process thought to resolve ambiguity resulting from conflicting associations in counter-conditioning experiments (Bouton, 1993, 1994; Nelson, 2002). The results were affirmative. When open- and close-hand stimuli were rendered ambiguous through periods of counter-mirror training, the second-learned counter-mirror associations manifested more strongly in the trained context, resulting in less automatic imitation. The second experiment confirmed that this context-specificity was due to conflict between second-learned counter-mirror associations and first-learned mirror associations established during the course of normal, pre-experimental development by demonstrating that analogous first-learned arbitrary sensorimotor associations show generalisation across the context manipulation.

Counter-conditioning and renewal effects demonstrate that second-learned associations show greater context specificity than the first associations formed with a novel stimulus (Bouton, 1994; Bouton & King, 1983; Nelson, 2002; Nelson et al., 2011; Peck & Bouton, 1990). That counter-mirror learning is modulated by context argues that the counter-mirror visuomotor links are equivalent to second-learned associations acquired during counter-conditioning. This finding suggests a clear parallel between the mirror visuomotor links acquired during development and first-learned associations established during conditioning. Given that contextual modulation was a prediction derived from conditioning experiments, these effects support the view that the acquisition of links between visual and motor representations of action is mediated by phylogenetically ancient mechanisms of associative learning.

Associative accounts argue that the context specificity of second-learned associations emerges as a means to resolve ambiguity (Bouton, 1993, 1994; Nelson, 2002). Rather than over-write or unlearn existing associations, there is considerable evidence that extinction and counter-conditioning schedules establish additional parallel associations (Bouton & Peck, 1992; Brooks et al., 1995). However, being associated with two opposing responses renders the stimulus ambiguous. Because of the conflict between the first- and second-learned associations, it is not clear what the appropriate behaviour is in the presence of the stimulus. To resolve this ambiguity, S-R relationships may be modulated by the context.

Several possible modulating mechanisms have been proposed within the associative literature. First, it has been suggested that context specificity may reflect the acquisition of a configural S-R relationship (Brandon & Wagner, 1998; P. N. Wilson & Pearce, 1992). Ambiguity may increase the likelihood that the stimulus – in the present case the opening and closing hand stimuli – and the context are encoded as a single configuration, and it is this configuration which is associated with the counter-mirror response. When the stimulus is presented by itself outside of the trained context, it therefore fails to excite the configural stimulus → counter-mirror response associations. A second possibility is that second-learned associations are gated such that the stimulus can only excite the response in the presence of the contextual cues present during training (Bonardi, 1998; Bouton, 1994; Holland, 1983). Such a gate mechanism could either serve to suppress the activation of the stimulus through input modulation or inhibit propagation by modulating the excitability of the association itself.

These findings suggest that the temporal order in which we are exposed to sensorimotor contingencies is an important factor in determining which responses generalise to novel contexts. In particular, the nature of early sensorimotor experience may be crucial. The results suggest that whichever associations are acquired first, be they mirror or counter-mirror, will generalise readily across contexts, whereas associations acquired subsequently will show greater context-specificity. Parents and caregivers imitate newborns, and these interactions are likely to provide some of the earliest experience of contingencies between action execution (e.g. neonate executes smile) and action observation (e.g. neonate observes caregiver smiling) (Ray & Heyes, 2011). The early acquisition of mirror associations may cause 'mirroring' to become the default, context-general response which manifests in novel situations. If, however, newborns were first exposed to predominantly non-matching sensorimotor contingencies, automatic counter-mirror or complementary responses might become prepotent in novel contexts.

Contextual modulation may help to explain why MNs appear to code abstract motor goals. MNs respond differently to the sight of an experimenter grasping a food item depending on whether the goal of the action is to eat the food or to place it in a shoulder-mounted cup (Bonini et al., 2010; Fogassi et al., 2005). That MN responses differed before the observed action trajectories could be differentiated is taken as evidence that MNs help us understand the motor intentions of others. However, the present data demonstrate that a simple associative account can be advanced to explain these data, without postulating any role of MNs in higher-order action understanding. Prior to the experiment, it is reasonable to assume that the macaques had ample opportunity to acquire S-R associations between the sight of grasping and grasp-to-eat motor programs. During the experiment the animals acquired further S-R associations such that the sight of grasping could also excite grasp-to-place motor programs. Because these second-learned associations served to render the sight of grasping ambiguous they became subject to contextual control. Crucially, whenever the monkeys were required to grasp to place, the target cup was always present, either in front of the subjects or on their shoulder. The second-learned grasp-to-place associations are likely to have been modulated by the presence of this salient contextual cue. When the macaques subsequently observed grasping, either in the presence or absence of the cup, MNs were either excited or inhibited by this context, thus giving the impression that they were coding ‘action intention’.

In summary, while counter-mirror learning shows context-specificity, the acquisition of arbitrary S-R associations generalizes well to untrained contexts. Associative accounts argue that counter-mirror learning comes under contextual control because it establishes second-learned associations which render the sight of actions ambiguous. In contrast, novel S-R associations generalise to untrained contexts because there is no conflict with previous learning. That the visuomotor links established during counter-mirror learning are modulated by context indicates that counter-mirror links are equivalent to second-learned associations acquired during counter-conditioning experiments. The results therefore suggest a parallel between the mirror links acquired during development and first-learned excitatory associations established during conditioning. Consistent with the view of the ASL, it appears that the acquisition of the visuomotor links, thought to solve the correspondence problem, conforms to the associative principles formulated through the study of conditioning. In contrast, the Hebbian account of MN origins offers no account of these effects.

Chapter 6: General Discussion

6.1 Overview

The present chapter will consider the findings from the experiments presented in Chapters 2-5, within the context of the central theoretical issue – the mechanisms that allow us to solve the visuomotor correspondence problem. The findings and theoretical significance of each experiment will first be described. This will be followed by a discussion of the limitations of each experiment, and where relevant those questions that remain unanswered will be identified. Section 6.2 will review Experiments 1 and 2 (Chapter 2), which investigated whether recognition of self-produced facial motion was mediated by topographic or temporal cues. Section 6.3 will review Experiments 3 and 4 (Chapter 3), which sought to identify the conditions necessary for the refinement of visuomotor correspondence knowledge. Section 6.4 will review Experiments 5 and 6 (Chapter 4), which tested whether the acquisition of visuomotor links is sensitive to sensorimotor contingency. Section 6.5 will review Experiments 7 and 8 (Chapter 5), which determined whether second-learned visuomotor links are subject to contextual control.

6.2 Self-recognition is mediated by temporal, not topographic, cues

6.2.1 Summary and interpretation

The experiments reported in Chapter 2 sought a better understanding of the mechanisms mediating superior recognition of third-person displays of self-produced, perceptually opaque movements, relative to displays of comparable movements produced by friends (Beardsworth & Buckner, 1981; Jokisch et al., 2006; Loula et al., 2005; Prasad & Shiffrar, 2009). Specifically, Experiments 1 and 2 tested the widely held view that the self-recognition advantage reflects the transfer of topographic information from the motor to visual systems (section 1.3.2.3.1).

The aim of Experiment 1 was to determine whether methodological artefacts were responsible for previous reports of the self-recognition advantage. To address this concern, Experiment 1 broke from the tradition of using point-light methods to study body actions, and instead sought evidence of superior self-recognition for another form of perceptually opaque actions – facial motion – using a markerless avatar technique. This approach completely eliminates form cues by animating a common facial form with the motion derived from different actors. Moreover, because this technique does not require individuals to wear markers or point-light apparatus during filming, it is better able to capture naturalistic motion than the methods used previously. Steps were also taken to prevent the encoding and episodic recall of performance characteristics. Despite these methodological improvements, evidence of superior self-recognition was observed for inverted facial motion, confirming that this effect is not an artefact.

Having identified a self-recognition advantage for inverted facial motion in Experiment 1, Experiment 2 sought to identify what type of information mediates superior self-recognition. Through the use of selectively disrupted avatar motion, Experiment 2 revealed that self-recognition for inverted facial motion was achieved through knowledge of the temporal characteristics of one's own actions. Inverted self-recognition was unaffected by anti-framing; a technique which disrupts topographic stimulus properties (feature configurations and trajectories), but which preserves the temporal characteristics of a stimulus. In contrast, self-recognition was eliminated by rhythmic disruption and global slowing; manipulations which distort the temporal characteristics of stimuli, but preserve topographic cues.

The ability to recognise third-person displays of self-produced, perceptually opaque movements better than those of friends is important because it suggests that 'first-person' performance expertise somehow makes that action easier to recognize when viewed from third-person perspectives. If recognition were dependent solely on visual experience, one would expect the opposite result - superior recognition of friend's movements. Existing accounts argue that covert simulation within the motor system enhances the topographic description of observed movements, by aiding the prediction of feature configurations and limb trajectories. When we observe our own actions the contribution from this prediction process is maximised (Blakemore & Frith, 2003; Schutz-Bosbach & Prinz, 2007; M. Wilson & Knoblich, 2005). Crucially, because stimuli depict movements from allocentric perspectives - view points from which we rarely see our own perceptually opaque actions, this motor-to-visual transfer could not excite third-person visual representations of their own actions directly via associatively acquired visuomotor links. Rather, a motor-to-visual transfer of topographic information would, in this context, require an endogenous process - actors would have to 'work out' what their actions look like from third-person perspectives.

However, the findings from Experiments 1 and 2 argue against the existing account of the self-recognition advantage, and instead indicate that the effect is mediated by temporal, not topographic information. That self-recognition depends on temporal, rather than topographic cues, indicates that such an endogenous process is unnecessary. While the topographic characteristics of facial action are perceptually opaque, the temporal characteristics are perceptually transparent. The information required for self-recognition can therefore be acquired during correlated sensorimotor experience. We have the opportunity to learn the temporal signatures of our actions via first-person visual, proprioceptive, somatosensory, auditory and vestibular feedback. Once acquired, this temporal knowledge may mediate recognition of self-produced perceptually opaque movements viewed from third-person perspectives. Consequently, while self-recognition of perceptually opaque movements based on their topographic characteristics is inconsistent with associative solutions, self-recognition mediated by temporal cues is not.

Self-recognition effects have been cited as evidence that the motor system may transfer topographic information to the perceptual system, in the absence of correlated sensorimotor experience. Because these effects were thought to require an endogenous transfer process they have been cited as evidence against associative solutions to the correspondence problem (section 1.2.1.1). However, the finding that self-recognition is mediated by temporal and not topographic cues, challenges this objection to associative solutions to the correspondence problem.

6.2.2 Limitations and outstanding questions

In order to study recognition from motion, it is necessary to isolate motion cues from the identity cues conveyed by form. This necessity for avatar or point-light methods inevitably makes stimulus generation a lengthy and computationally demanding process. The practical constraints imposed by stimulus generation had a number of consequences for the present study. First, the small sample sizes (6 friend pairs; 12 participants) used in Experiment 1 reduced the power of the statistical analyses. Second, ideally Experiment 2 would have replicated the self-recognition advantage for inverted facial motion, with a novel sample, and thereby demonstrate that the effects observed extend beyond the particular friend pairings used. This would have also excluded the possibility that completing the testing for Experiment 1 somehow influenced performance on the testing for Experiment 2. Third, the production of avatar stimuli for six friend pairs meant that only these 12 individuals could participate in the experiments. Due to the limited availability of these participants it was not possible to run every condition that might have been desirable. For example, the manipulations reported in Experiment 2 were tested in the inverted orientations only. It remains to be seen whether anti-framing would disrupt friend recognition in the upright orientation.

The self-recognition advantage is puzzling because people have relatively few opportunities to observe their own perceptually opaque movements and thereby to acquire knowledge about the topographic features of their own actions. It has been suggested that the findings described in Chapter 2 solve this puzzle by showing that, in both upright and inverted conditions, people use temporal rather than topographic cues for self-recognition. In principle, the results of Experiments 1 and 2 are also consistent with an alternative interpretation – that participants typically use configural topographic cues to recognise themselves in the upright orientation, but then resort to a local temporal strategy when forced to do so by stimulus inversion. This is a coherent interpretation, but it lacks theoretical and empirical motivation. At the theoretical level, this account leaves the correspondence problem unresolved for facial actions viewed upright – it fails to address how participants acquire the necessary topographic knowledge. As described above, this would require an unspecified endogenous process. Moreover, it is unclear why we would use hard-to-derive topographic knowledge, when readily available temporal cues permit self-recognition in both orientations. At the empirical level, there is no evidence that topographic rather than temporal cues mediate self-recognition in either orientation.

Reports of superior self-recognition are conceptually similar to evidence that non-visual motor training, during which the participant is blindfolded, improves action recognition (Casile & Giese, 2006; Hecht et al., 2001). Both effects have been cited as evidence that the motor system may transfer topographic information to the visual system in the absence of correlated sensorimotor experience. The finding that self-recognition is dependent on temporal cues therefore raises the possibility that the effects of non-visual motor training may also be mediated by temporal cues. In the absence of empirical evidence, this remains a tentative speculation. Nevertheless, this is an intriguing possibility, and one of the most important outstanding questions to be addressed by future research on motor contributions to perception.

6.3 Contingent visual feedback is necessary to refine visuomotor correspondence knowledge

6.3.1 Summary and interpretation

The aim of the experiments reported in Chapter 3 was to identify the conditions necessary for the refinement of visuomotor correspondence knowledge. According to associative accounts, the only way to improve the accuracy of visuomotor correspondence knowledge is through the provision of accurate third-person visual feedback. In contrast, AIM predicts that participants should continue to hone the accuracy of their visuomotor matching, even in the absence of visual feedback, via a hardwired proprioceptive feedback loop. Two experiments were conducted to test these alternative models using a new technology which allowed the automated objective quantification of imitative accuracy. Euclidean distances, measured in image-derived principal component space, were used for the first time to quantify the accuracy of adult participants' attempts to replicate their own facial expressions, before, during and after training.

Experiment 3 compared the improvement of a visual feedback group with that of a no feedback group. During training, participants in the visual feedback group were presented with third-person visual feedback contingent on the current attempt, while the no visual feedback group were reliant solely on proprioception. Whereas the no visual feedback group got progressively worse at self-imitation, the visual feedback group showed progressive improvement. Using a similar design, Experiment 4 compared the relative improvement of a second visual feedback group with a non-contingent visual feedback group who also received visual feedback, from an earlier attempt, not their most recent attempt. Whereas the participants given contingent visual feedback again showed evidence of incremental improvement, the non-contingent feedback group demonstrated no improvement.

The present findings indicate that adults find it extremely difficult to use proprioception to inform imitative error correction. The absence of improvement in the no visual feedback group (Experiment 3) could conceivably be attributed to poor perceptual representation due to limited opportunity for perceptual learning. However, such an account cannot explain the differential learning effects observed in Experiment 4, as participants in the visual feedback and non-contingent feedback groups observed the same number of target-attempt displays. According to AIM, humans are born with an innate ability to use proprioception in order to achieve accurate visuomotor correspondences. However, that adults cannot use proprioception to guide facial imitation clearly argues against this view. Instead these findings suggest that adults' ability to achieve accurate visuomotor matching depends on the nature of the sensorimotor experience they have been exposed to, consistent with associative solutions to the correspondence problem.

6.3.2 Limitations and outstanding questions

The present findings indicate that the mature visuomotor links mediating adult imitation cannot be improved without visual feedback, challenging the view that supramodal representations allow correspondence knowledge to be inferred (Meltzoff & Decety, 2003; Meltzoff & Moore, 1997). However, the original improvement data cited as support for AIM came from infant participants (Meltzoff & Moore, 1994; Nagy et al., 2005; Soussignan et al., 2011). It therefore remains possible that adult imitation is mediated by visuomotor links established through correlated sensorimotor experience, but that infant imitation is mediated by a separate innate mechanism. For example, infants may be endowed with a mechanism, such as that encapsulated within the AIM model, which allows corresponding visual and motor representations to be identified in the absence of correlated sensorimotor experience. After a critical period has elapsed, the infant may subsequently become reliant on visuomotor links acquired associatively. Nevertheless, while this account cannot be excluded, it should be noted that the original neonatal improvement data do not provide compelling support for the existence of a supramodal mechanism in infants (section 1.3.2.1.3).

As described in the General Discussion of Chapter 3 (section 3.4), the non-contingent manipulation used in Experiment 4 degraded, not only the attempt-feedback contingency, but also the attempt-feedback contiguity. The performance of the contingent and contiguous visual feedback group was therefore compared with a non-contingent, non-contiguous condition. This makes it impossible to identify which aspect – contingency, contiguity or both, is crucial for imitative improvement. While the experiments in this chapter meet their principle objective – to test the opposing predictions of the AIM and ASL models – they do not distinguish between the rival associative models based on contiguity (Hebbian Hypothesis) and contiguity and contingency (ASL). This issue was addressed explicitly in Chapter 4.

6.4 The acquisition of visuomotor links is sensitive to contingency

6.4.1 Summary and interpretation

In order to distinguish between the rival associative accounts - ASL and the Hebbian theory of MN origins - Experiments 5 and 6 sought to determine whether the acquisition of 'counter-mirror' visuomotor links is sensitive to contingency. Through its appeal to the principles of associative learning established through the study of conditioning, ASL predicts that the acquisition of visuomotor links is sensitive, not only to temporal contiguity, but also to sensorimotor contingency (section 1.2.1.1). In contrast, through its appeal to the learning principles documented by Hebb (1949), the Hebbian account of MNs predicts that correspondence acquisition is sensitive only to temporal contiguity (section 1.2.1.3).

In Experiment 5, residual automatic imitation was measured following counter-mirror training in which the action stimulus was a perfect predictor of the response (contingent) or not at all predictive of the response (non-contingent). A contingency effect was observed: there was less automatic imitation, indicative of more learning, in the contingent group. Experiment 6 replicated this contingency effect and showed that, as predicted by associative learning theory, it can be abolished by signalling trials in which the response occurs in the absence of an action stimulus. Because the number of temporally contiguous pairings between action observation and action execution were held constant throughout, these findings are challenging for the Hebbian account of visuomotor link acquisition.

A Hebbian account can be sustained if it assumed that the differential learning exhibited by the contingent and non-contingent groups was due to response habituation. Because the non-contingent group executed twice as many responses as the contingent group, the resulting habituation of the motor representation may have prevented the acquisition of visuomotor links. However, the additional signalling effect observed in Experiment 6 argues against this view; participants in the signalled group made the same number of responses, but exhibited learning comparable with the contingent group. Rather, the signalling effect suggests that the contingency effect may result from contextual overshadowing; competition for associative strength between the action stimuli and the experimental context.

6.4.2 Limitations and outstanding questions

Counter-mirror training has been shown to attenuate automatic imitation (Heyes et al., 2005); modulate the BOLD response of the human MNS (Catmur et al., 2008); and reverse the characteristic mirror pattern of MEPs seen during action observation (Catmur et al., 2007). In each case the findings have been interpreted as evidence that training causes the formation of counter-mirror associations conceptually identical to the mirror associations connecting the corresponding visual and motor representations. However, this inference is complicated by the fact that the first-learned associations, acquired with a novel stimulus, may not behave in the same way as second-learned associations acquired with a familiar stimulus (see section 6.5). Proponents of Hebbian MN theory might therefore argue that, while the acquisition of original mirror links is sensitive only to contiguity, subsequent revisions might depend on both contiguity and contingency. While this speculation lacks empirical support, it remains a logical possibility that cannot be excluded. This question could be addressed explicitly in the future by testing whether the contingency manipulations employed in Chapter 4 also modulate the acquisition of first-learned visuomotor associations, for example between robotic actions and matching responses (Press et al., 2007).

It might be argued that it is impossible to make inferences about the human MNS from behavioural data on automatic imitation. In the absence of an accompanying neurophysiological method, one cannot be certain of the neural substrate of this behavioural effect. However, evidence that automatic imitation is both ‘automatic’ – that is minimally dependent on the actor’s intentions - and ‘imitation’ - dependent on the topographic qualities of observed actions – has led to a widely held consensus that this effect is a product of the human MNS (Ferrari et al., 2009; Heyes, 2011; Longo, Kosobud, & Bertenthal, 2008). This interpretation accords with evidence that the application of disruptive TMS to inferior frontal gyrus (IFG) abolishes the automatic imitation effect (Catmur, Walsh, & Heyes, 2009; Heiser, Iacoboni, Maeda, Marcus, & Mazziotta, 2003; Newman-Norlund, Ondobaka, van Schie, van Elswijk, & Bekkering, 2010). Moreover, evidence that counter-mirror training reverses the mirror pattern of MEPs seen during action observation (Catmur et al., 2007), and modulates the BOLD response of the human MNS (Catmur et al., 2008) provides convergent evidence that the present effects reflect learning within this system.

6.5 Second-learned visuomotor links are modulated by context

6.5.1 Summary and interpretation

Experiments 7 and 8 tested whether the learning observed during the counter-mirror training paradigm is sensitive to context. The associative literature makes a number of testable predictions regarding the acquisition of visuomotor links that might serve to distinguish the ASL and Hebbian accounts. However, the context-specificity of learning was chosen for this purpose, in order to better understand previous reports of MNS modulation by action context (Bonini et al., 2010; Fogassi et al., 2005; Iacoboni et al., 2005). Studies of counter-conditioning and renewal have found that ‘first-learned’ associations formed with a novel stimulus tend to generalise well to different contexts, while second learned associations which contradict the original learning, are typically subject to contextual control. If ASL is correct that the acquisition of visuomotor links is mediated by the same phylogenetically ancient mechanisms of associative learning that mediate conditioning, second learned visuomotor links, such as those established during counter-mirror learning, ought to show context specificity.

Experiment 7 tested this prediction. Two separate pre-tests were first conducted to establish participants’ baseline automatic imitation effects in distinctive red and blue contexts. During the second and third sessions, participants received counter-mirror training in either the red or the blue contexts. In the final session two separate post-tests were conducted to establish participants’ residual automatic imitation effects in both the red and blue contexts. Context-specificity was indexed by comparing the magnitude of the pre- to post-test reduction in automatic imitation when participants were tested in the context in which they received counter-mirror training and in their untrained context. The results confirmed the prediction of the ASL model: Less residual automatic imitation was observed when participants were tested in their counter-mirror training context, relative to their untrained context.

The ASL model argues that the context-specificity observed in Experiment 1 was due to conflict between the second-learned ‘counter-mirror’ associations established by experimental training, and first-learned ‘mirror’ associations acquired by the participants before the experiment began, during normal ontogeny. However, the conflict-generating learning could have taken place, not during normal development, but during the pre-test. During the pre-test there was no contingency between the action stimuli observed and the responses executed by participants. The onset of the action stimulus served simply as a go-signal. However, during counter-mirror training the onset of the action stimulus not only served as a go-signal but also indicated which response participants should execute. It could thus be argued that the ‘meaning’ of the action stimuli differed between the pre-test and training procedures (Bouton, 1994), and that it was this ambiguity, rather than the conflict between the mirror and counter-mirror responses, that caused the counter-mirror learning to become subject to contextual control.

To demonstrate that context-specificity was not an artefact of the pre-test procedure, Experiment 8 used a four session design, identical to that employed in Experiment 1, to train arbitrary S-R mappings with abstract geometric shapes. It was predicted that novel S-R learning should generalise well to untrained contexts because there was no conflict with pre-experimental learning to render the stimuli ambiguous. Crucially, the only possible source of ambiguity was the transfer from the simple RT pre-test to the choice RT training procedure; it could not have come from pre-experimental learning. Consequently, any evidence of context-specificity would confirm that the use of the pre-test procedure was sufficient to render learning context-specific.

Participants again made open- and close-hand responses, but this time to the onsets of geometric trapezoid forms. In the first session they completed simple RT pre-test procedures in the red and blue contexts to confirm the absence of any pre-existing compatibility effects. They then completed two training sessions with a choice RT task, either in the red or blue contexts, where they learned to make open and close responses to the onset of top- and bottom-heavy trapezoids, respectively. In the final session they were tested again in both the red and blue contexts to determine the magnitude of their newly acquired SRC effect. In contrast to the results of Experiment 7, learning was found to be insensitive to the context manipulation.

These results accord with the view of the ASL model. Insofar as contextual control of second learned associations is a prediction derived from conditioning experiments, this effect indicates that the acquisition of visuomotor links is mediated by the same psychological processes as conditioning – by phylogenetically ancient mechanisms of associative learning. Associative accounts of contextual control argue that counter-mirror learning is modulated by context in order to resolve ambiguity. When action stimuli become associated with two responses, the stimulus is rendered ambiguous. However, the modulation of link excitability by contextual cues ensures that different responses can be prepotent in different situations.

The results reported in Chapter 5 are important because they confirm that a coherent associative account can be advanced to interpret neurophysiological (Bonini et al., 2010; Fogassi et al., 2005) and neuroimaging (Iacoboni et al., 2005) evidence that MN responses may be modulated by context. In each case where MNS modulation has been reported, salient contextual cues have been present which might be expected to exert contextual control over second-learned visuomotor links. Consequently, reports of contextual modulation do not necessarily indicate that MNs are encoding the ‘goal’ of an action (Rizzolatti & Craighero, 2004; Rizzolatti et al., 1996), but may instead reflect their associative origins. In contrast, the simple contiguity-based learning mechanism advanced by Hebbian MN theory offers no account of the contextual modulation observed in Experiment 7 or in the MN literature.

6.5.2 Limitations and outstanding questions

An alternative design to that employed in Experiment 8 would have first trained arbitrary S-R mappings to novel stimuli in an initial training phase, and then reversed the mapping in a second counter-conditioning phase. Associative accounts of contextual control predict that the original learning should generalise to novel contexts, while the counter-conditioning ought to show context specificity. This design was rejected on pragmatic grounds. In order to be strictly comparable with Experiment 7 participants would have needed to complete a seven day procedure (day-1 pre-test; days 2/3 training with the first S-R mapping; day-4 first post-test; days 5/6 training with reversed mapping; day-7 second-post-test). Such a lengthy procedure would have made recruiting participants difficult and likely led to high attrition rates. This procedure could be abbreviated for use in future studies by using only a single training session for each mapping, or relaxing the requirement to complete consecutive sessions 24-hours apart. It remains a question for future research whether human S-R learning yields the predicted pattern of results when tested using a full counter-conditioning procedure.

6.6 Conclusion

This thesis sought to identify how humans solve the visuomotor correspondence problem. At the outset, several theories were described which offer accounts of the acquisition of visuomotor correspondence knowledge. Two associative theories were reviewed - ASL and Hebbian MN theory - which posit that correlated sensorimotor is necessary to link visual and motor representations. Both of these associative theories were supported by evidence that sensorimotor expertise modulated the response of the MNS (section 1.3.1.1); by evidence that markers of MNS functioning were modulated by sensorimotor training (section 1.3.1.2); and by neurophysiological reports of tool-use and audiovisual MNs in macaques (section 1.3.1.3). In contrast, two nativist solutions were also described: innate MN theory which posits the existence of innate connections between corresponding visual and motor representations; and the AIM hypothesis which proposes an innate mechanism for matching observed actions to their corresponding motor programs. These nativist theories were supported by evidence of neonatal imitation and imitative improvement in the absence of visual feedback (section 1.3.2.1); by reports of MNS responses to visual stimuli which have not been contingently paired with action performance (section 1.3.2.2); and by evidence of motor-to-visual transfer in the absence of correlated sensorimotor experience (section 1.3.2.3).

The first two empirical chapters addressed previous reports that appeared to challenge associative accounts of the acquisition of visuomotor correspondence knowledge: motor-to-visual transfer in the absence of correlated sensorimotor experience and imitative improvement in the absence of visual feedback. These were regarded as the most challenging lines of evidence against associative solutions, given that reports of MNS responses to stimuli that have not been paired with action execution can be accounted for through stimulus generalisation or excitation of complementary responses (see section 1.3.2.2).

The finding described in Chapter 2 – that self-recognition is mediated by temporal cues – suggests that the evidence for motor-to-visual transfer provided by self-recognition effects can be readily accommodated within an associative framework. That the crucial information playing the facilitating role is temporal and not topographic means that a motor-to-visual transfer may be mediated by links acquired associatively during correlated sensorimotor experience. Moreover, this finding raises the possibility that other reports of motor-to-visual transfer in the absence of correlated sensorimotor experience – for example, improved visual discrimination following non-visual motor training – may also be mediated by temporal cues.

Chapter 2 demonstrated that one of the most important findings from the neonatal imitation literature – improvement in the absence of visual feedback - cannot be replicated when an objective scoring procedure is used to assess adult facial self-imitation. Rather, adults require visual feedback contingent on gesture execution to improve the accuracy of their facial self-imitation. This finding suggests that reports of imitative improvement in the absence of visual feedback in infants may reflect the particular methodologies and subjective scoring procedures used. Despite ongoing methodological concerns, the neonatal improvement data remain significant as they represent the best evidence that the matching of tongue-protrusion by neonates reflects innate correspondence knowledge. That these findings cannot be replicated with an objective scoring procedure adds weight to the view that the matching of tongue protrusion may instead be an artefact of arousal (Jones, 1996, 2006, 2009) or reflect an IRM (Jacobson, 1979).

The first four experiments reported in this thesis addressed two objections to associative solutions to the correspondence acquisition, i.e. the transfer of topographic information from the motor to visual system in the absence of correlated sensorimotor experience; and imitative improvement in the absence of correlated sensorimotor experience. Together with previous findings from neuroimaging studies of sensorimotor expertise; sensorimotor training effects; and evidence from single-cell recording, these results form a compelling body of evidence that argues against the existence of i) hardwired visuomotor links (innate MN theory); or ii) an innate mechanism for identifying corresponding visual and motor representations (AIM). The foregoing evidence therefore supports the view that the visuomotor correspondence problem is solved by direct visuomotor links, acquired associatively during correlated sensorimotor experience.

However these findings are consistent with both of the rival associative accounts: The ASL model - which argues that the acquisition of visuomotor links is mediated by the same phylogenetically ancient mechanisms of associative learning responsible for conditioning - and Hebbian MN theory – which argues that visuomotor link acquisition is mediated by a simple contiguity-based learning mechanism. The aim of the experiments reported in Chapters 4 and 5 was therefore to determine which of these associative accounts best describes the acquisition of the visuomotor links which appear to solve the correspondence problem.

Consistent with the prediction of the ASL model, the experiments reported in Chapter 4 found that the acquisition of counter-mirror visuomotor links was sensitive to contingency. This finding is challenging for a Hebbian account of visuomotor link acquisition because the number of contiguous pairings was held constant. Similarly, in Chapter 5 second-learned counter-mirror links were found to be modulated by context, indicating that counter-mirror learning behaves like a counter-conditioning process. While this finding is predicted by the ASL framework, contextual modulation is beyond the scope of the simple contiguity-based learning mechanism described by the Hebbian Hypothesis in its present form.

Together with the existing body of evidence, the experiments contained within this thesis therefore argue that the visuomotor correspondence problem is solved by associations, acquired during correlated sensorimotor experience, which connect directly the visual and motor descriptions of actions. Consistent with the view encapsulated within the ASL model, the acquisition of these visuomotor links seems to conform to the principles of associative learning established through the study of conditioning in humans and animals.

Critics might argue that, as a theory, ASL is in some respects under-specified and potentially hard to falsify. Because its central claim is that corresponding visual and motor representations are linked through domain-general mechanisms of associative learning, it appeals to an entire literature - a substantial body of theoretical models and conditioning effects. However, this literature is often equivocal; it contains numerous unresolved questions and rival theoretical accounts. Where such controversies exist, they are problematic for ASL because the position of the model is ambiguous. For example, there remains widespread disagreement as to whether backward conditioning is a robust phenomenon. Consequently, it is unclear whether the sensorimotor associations described by ASL should be exclusively unidirectional, or whether they can, under some circumstances, be bidirectional (see section 1.2.1.1). Similarly, whereas some models of associative learning appeal to attentional processes to explain various conditioning effects (Mackintosh, 1975; Pearce & Hall, 1980), other models are framed in terms of direct associative competition (Rescorla & Wagner, 1972). Such fundamental differences in these accounts of associative processes prevent the ASL model from specifying in more detail the precise learning mechanisms. Furthermore, its appeal to an entire literature also means that the predictions of the ASL model may change over time. Where new theoretical or empirical advances are reported in the associative learning literature, these are automatically incorporated into the predictions of ASL. That the model is never ‘set in stone’ makes it hard to falsify. For these reasons, critics may characterise ASL as a philosophical or theoretical approach to the visuomotor correspondence problem, rather than a true theoretical model.

However, the value of a theory lies is its ability to make testable predictions which differ from those of rival models. While there are areas of controversy within the associative literature there are also many areas of robust consensus which may be used to derive testable, unambiguous predictions. For example, while the mechanism of the effect may be unresolved, it is widely accepted that the acquisition of associations is sensitive to the degree of contingency. The same is true of contextual modulation: While the mechanisms are not yet fully understood, empirical effects such as renewal and context-specific counter-conditioning are widely accepted. Areas of consensus, such as these, allow ASL to generate predictions that either directly conflict with, or are beyond the scope of rival models. That ASL has generated, and continues to generate novel predictions, confirms its value as a theoretical model. Moreover, ASL is falsifiable. For example, if it were shown unequivocally that newborn infants were endowed with correspondence knowledge, in the absence of correlated sensorimotor experience, this would falsify the fundamental claim of the model.

Many of the criticisms that may be levelled at the ASL model - that it is under specified and subject to the prevailing consensus in the associative literature - may be levelled at any theory which seeks to invoke domain-general processes (e.g. associative learning; perceptual learning; attention) to explain particular cognitive functions. For example, if one were to invoke perceptual learning to explain the mechanisms of face perception, this position would necessarily appeal to the ‘equivocal’ and ‘ever-changing’ literature on perceptual learning. Nevertheless, this theoretical position has proved a fruitful one within the face perception literature: It has allowed testable predictions to be derived which distinguish a domain-general view from rival domain-specific models (Diamond & Carey, 1986; Gauthier & Tarr, 1997). The ability to specify the precise parameters of domain-general accounts is inevitably constrained by our understanding of the particular process in question. In this respect, it may well be easier to propose highly-specified domain-specific models because they are not subject to this constraint. However, it should be noted that AIM - the leading domain-specific solution to the correspondence problem – is also underspecified in several respects. Despite the inherent challenges associated with domain-general accounts, it would be regressive to reject on this basis, the very concept of domain-general processes.

It is the conclusion of this thesis that the sensorimotor links that solve the correspondence problem are conceptually equivalent to those which mediate other types of generic sensorimotor relationship. According to this view imitative and complimentary (or counter-mirror) responses are mediated by identical sensorimotor processes. However, a further parallel may also be drawn between imitative responses and object affordances. In much the same way that contingent sensorimotor experience causes the sight of an action to excite corresponding motor representations, contingent sensorimotor experience may also cause the sight of an object to excite the motor programs associated with its use. Interestingly, canonical neurons – the sensorimotor units thought to mediate object affordances – and MNs are known to be found in similar, and in some cases overlapping, areas of inferior parietal and premotor cortex (Grezes, Armony, Rowe, & Passingham, 2003; Rizzolatti & Luppino, 2001). Equally, a number of parallels exist between echopraxia and utilisation behaviour – two dysexecutive syndromes associated with involuntary imitation and object use, respectively (Lhermitte, 1983).

Nevertheless, the conclusion that domain-general mechanisms mediate imitation should not detract from the view that imitation is a crucial form of behaviour. Studies of social learning suggest that the ability to acquire novel behaviours through observation conveys a substantial advantage on groups and individuals (Heyes, in press; Heyes & Galef, 1996). Moreover, we are only beginning to understand the profound consequences of being imitated, on mood, inter-personal attributions, and co-operation (van Baaren, Janssen, Chartrand, & Dijksterhuis, 2009). For example, being imitated appears to cause interactants to like each other more (Chartrand & Bargh, 1999; Lakin & Chartrand, 2003); and increases the likelihood they will help one another and donate money to charity (van Baaren, Holland, Kawakami, & van Knippenberg, 2004). While the core mechanism at the heart of imitation may not be an evolutionary adaptation, it remains likely that imitation has played an important role in the development and evolution of human society.

References:

Abravanel, E., & Deyong, N. G. (1991). Does object modeling elicit imitative-like gestures from young infants. *Journal of Experimental Child Psychology, 52*(1), 22-40.

Abravanel, E., & Sigafoos, A. D. (1984). Exploring the Presence of Imitation during Early Infancy. *Child Development, 55*(2), 381-392.

Anisfeld, M. (1991). Neonatal imitation. *Developmental Review, 11*(1), 60-97.

Anisfeld, M. (1996). Only tongue protrusion modeling is matched by neonates. *Developmental Review, 16*(2), 149-161.

Ayres, J. J., Haddad, C., & Albert, M. (1987). One-trial excitatory backwards conditioning as assessed by conditioned supression of licking in rats: Concurrent observations of lick suppression and defensive behaviors *Animal Learning and Behavior 15*, 212-217.

Beardsworth, T., & Buckner, T. (1981). The ability to recognize oneself from a video recording of ones movements without seeing ones body. *Bulletin of the Psychonomic Society, 18*(1), 19-22.

Berisha, F., Johnston, A., & McOwan, P. W. (2010). Identifying regions that carry the best information about global facial configurations. *Journal of Vision, 10*(11), 1-8.

Blake, R., & Shiffrar, M. (2007). Perception of human motion. *Annual Review of Psychology, 58*, 47-73.

Blakemore, S. J., & Frith, C. (2003). Self-awareness and action. *Current Opinion in Neurobiology, 13*(2), 219-224.

Blakemore, S. J., & Frith, C. (2005). The role of motor contagion in the prediction of action. *Neuropsychologia, 43*(2), 260-267.

Bonardi, C. (1998). Conditional learning: an associative analysis. In P. C. Holland & N. A. Schmajuk (Eds.), *Occasion Setting: Associative Learning and Cognition in Animals* (pp. 37-67). Washington DC: American Psychological Association.

Bonini, L., Rozzi, S., Serventi, F. U., Simone, L., Ferrari, P. F., & Fogassi, L. (2010). Ventral premotor and inferior parietal cortices make distinct contribution to action organization and intention understanding. *Cerebral Cortex, 20*(6), 1372-1385.

Borroni, P., & Baldissera, F. (2008). Activation of motor pathways during observation and execution of hand movements. *Social Neuroscience, 3*(3-4), 276-288.

Bouton, M. E. (1993). Context, time, and memory retrieval in the interference paradigms of Pavlovian learning. *Psychological Bulletin, 114*(1), 80-99.

Bouton, M. E. (1994). Context, ambiguity, and classical-conditioning. *Current Directions in Psychological Science, 3*(2), 49-53.

Bouton, M. E., & King, D. A. (1983). Contextual control of the extinction of conditioned fear: tests for the associative value of the context. *Journal of Experimental Psychology-Animal Behavior Processes, 9*(3), 248-265.

Bouton, M. E., & Peck, C. A. (1992). Spontaneous-recovery in cross-motivational transfer (Counterconditioning). *Animal Learning & Behavior, 20*(4), 313-321.

Brandon, S. E., & Wagner, A. R. (1998). Occasion setting: Influences of conditioned emotional responses and configural cues. In P. C. Holland & N. A. Schmajuk (Eds.), *Occasion setting: Associative learning and cognition in animals* (pp. 343-382). Washington DC: American Psychological Association.

Brass, M., Bekkering, H., & Prinz, W. (2001). Movement observation affects movement execution in a simple response task. *Acta Psychologica, 106*(1-2), 3-22.

Brass, M., & Heyes, C. (2005). Imitation: is cognitive neuroscience solving the correspondence problem? *Trends in Cognitive Sciences, 9*(10), 489-495.

Brooks, D. C., Hale, B., Nelson, J. B., & Bouton, M. E. (1995). Reinstatement after counterconditioning. *Animal Learning & Behavior, 23*(4), 383-390.

Buccino, G., Binkofski, F., Fink, G. R., Fadiga, L., Fogassi, L., Gallese, V., et al. (2001). Action observation activates premotor and parietal areas in a somatotopic manner: an fMRI study. *European Journal of Neuroscience, 13*(2), 400-404.

Buccino, G., Lui, F., Canessa, N., Patteri, I., Lagravinese, G., Benuzzi, F., et al. (2004). Neural circuits involved in the recognition of actions performed by nonconspecifics: an FMRI study. *Journal of Cognitive Neuroscience, 16*(1), 114-126.

Bulthoff, I., Bulthoff, H., & Sinha, P. (1998). Top-down influences on stereoscopic depth-perception. *Nature Neuroscience, 1*(3), 254-257.

Calvo-Merino, B., Glaser, D. E., Grezes, J., Passingham, R. E., & Haggard, P. (2005). Action observation and acquired motor skills: An fMRI study with expert dancers. *Cerebral Cortex, 15*(8), 1243-1249.

Calvo-Merino, B., Grezes, J., Glaser, D. E., Passingham, R. E., & Haggard, P. (2006). Seeing or doing? Influence of visual and motor familiarity in action observation. *Current Biology, 16*(19), 1905-1910.

Casile, A., Caggiano, V., & Ferrari, P. F. (2011). The mirror neuron system: a fresh view. *Neuroscientist, 17*(5), 524-538.

Casile, A., & Giese, M. A. (2006). Nonvisual motor training influences biological motion perception. *Current Biology, 16*(1), 69-74.

Catmur, C., Gillmeister, H., Bird, G., Liepelt, R., Brass, M., & Heyes, C. (2008). Through the looking glass: counter-mirror activation following incompatible sensorimotor learning. *European Journal of Neuroscience, 28*(6), 1208-1215.

Catmur, C., & Heyes, C. (2010). Time course analyses confirm independence of imitative and spatial compatibility. *Journal of Experimental Psychology: Human Perception and Performance*.

Catmur, C., Mars, R. B., Rushworth, M. F., & Heyes, C. (2010). Making mirrors: Premotor cortex stimulation enhances mirror and counter-mirror motor facilitation. *Journal of Cognitive Neuroscience*.

Catmur, C., Walsh, V., & Heyes, C. (2007). Sensorimotor learning configures the human mirror system. *Current Biology, 17*(17), 1527-1531.

Catmur, C., Walsh, V., & Heyes, C. (2009). Associative sequence learning: the role of experience in the development of imitation and the mirror system. *Philosophical Transactions of the Royal Society of London Series B: Biological Sciences 364*(1528), 2369-2380.

Chartrand, T. L., & Bargh, J. A. (1999). The chameleon effect: the perception-behavior link and social interaction. *Journal of Personality and Social Psychology, 76*(6), 893-910.

Cook, R., Bird, G., Lunser, G., Huck, S., & Heyes, C. (2011). Automatic imitation in a strategic context: players of rock-paper-scissors imitate opponents' gestures. *Proceedings of the Royal Society B-Biological Sciences*.

Cowe, G. (2003). *Example-based computer generated facial mimicry.* University of London, London.

Craighero, L., Bello, A., Fadiga, L., & Rizzolatti, G. (2002). Hand action preparation influences the responses to hand pictures. *Neuropsychologia, 40*(5), 492-502.

Cross, E. S., Hamilton, A. F., & Grafton, S. T. (2006). Building a motor simulation de novo: observation of dance by dancers. *Neuroimage, 31*(3), 1257-1267.

Cross, E. S., Hamilton, A. F., Kraemer, D. J., Kelley, W. M., & Grafton, S. T. (2009). Dissociable substrates for body motion and physical experience in the human action observation network. *European Journal of Neuroscience, 30*(7), 1383-1392.

Cutting, J. E., & Kozlowski, L. T. (1977). Recognizing friends by their walk: Gait perception without familiarity cues. *Bulletin of the Psychonomic Society, 9*(5), 353-356.

Daprati, E., Wriessnegger, S., & Lacquaniti, F. (2007). Kinematic cues and recognition of self-generated actions. *Experimental Brain Research, 177*(1), 31-44.

Del Giudice, M., Manera, V., & Keysers, C. (2009). Programmed to learn? The ontogeny of mirror neurons. *Developmental Science, 12*(2), 350-363.

di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., & Rizzolatti, G. (1992). Understanding motor events: a neurophysiological study. *Experimental Brain Research, 91*(1), 176-180.

Diamond, R., & Carey, S. (1986). Why faces are and are not special: an effect of expertise. *Journal of Experimental Psychology: General, 115*(2), 107-117.

Dickinson, A., & Charnock, D. J. (1985). Contingency effects with maintained instrumental reinforcement. *Quarterly Journal of Experimental Psychology, 37B*, 197-416.

Dickinson, A., Watt, A., & Griffiths, W. J. H. (1992). Free-operant acquisition with delayed reinforcement. *Quarterly Journal of Experimental Psychology, 45B*(3), 241-258.

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

Elsner, B., & Hommel, B. (2004). Contiguity and contingency in action-effect learning. *Psychological Research, 68*(2-3), 138-154.

Enticott, P. G., Johnston, P. J., Herring, S. E., Hoy, K. E., & Fitzgerald, P. B. (2008). Mirror neuron activation is associated with facial emotion processing. *Neuropsychologia, 46*(11), 2851-2854.

Fadiga, L., Fogassi, L., Pavesi, G., & Rizzolatti, G. (1995). Motor facilitation during action observation: a magnetic stimulation study. *Journal of Neurophysiology, 73*(6), 2608-2611.

Ferrari, P. F., Bonini, L., & Fogassi, L. (2009). From monkey mirror neurons to primate behaviours: possible 'direct' and 'indirect' pathways. *Philosophical Transactions of the Royal Society B-Biological Sciences*, *364*(1528), 2311-2323.

Ferrari, P. F., Rozzi, S., & Fogassi, L. (2005). Mirror neurons responding to observation of actions made with tools in monkey ventral premotor cortex. *Journal of Cognitive Neuroscience, 17*(2), 212-226.

Field, T. M., Woodson, R., Cohen, D., Greenberg, R., Garcia, R., & Collins, K. (1983). Discrimination and imitation of facial expressions by term and preterm neonates. *Infant Behavior & Development, 6*(4), 485-489.

Field, T. M., Woodson, R., Greenberg, R., & Cohen, D. (1982). Discrimination and imitation of facial expressions by neonates. *Science, 218*(4568), 179-181.

Flach, R., Knoblich, G., & Prinz, W. (2004). Recognizing one's own clapping: the role of temporal cues. *Psychological Research, 69*(1-2), 147-156.

Fogassi, L., Ferrari, P. F., Gesierich, B., Rozzi, S., Chersi, F., & Rizzolatti, G. (2005). Parietal lobe: from action organization to intention understanding. *Science, 308*(5722), 662-667.

Fontaine, R. (1984). Imitative skills between birth and 6 months. *Infant Behavior & Development, 7*(3), 323-333.

Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain, 119 (Pt 2)*, 593-609.

Gallese, V., Gernsbacher, M., Hickok, G., Heyes, C., & Iacoboni, M. (2011). Mirror neuron forum *Perspectives on Psychological Science, 6*(4), 369-407.

Gallese, V., & Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends in Cognitive Sciences, 2*(12), 493-501.

Gauthier, I., & Tarr, M. J. (1997). Becoming a "greeble" expert: Exploring mechanisms for face recognition. *Vision Research, 37*(12), 1673-1682.

Gazzola, V., Rizzolatti, G., Wicker, B., & Keysers, C. (2007). The anthropomorphic brain: the mirror neuron system responds to human and robotic actions. *Neuroimage, 35*(4), 1674-1684.

Giese, M. A., & Poggio, T. (2003). Neural mechanisms for the recognition of biological movements. *Nature Reviews Neuroscience, 4*(3), 179-192.

Gillmeister, H., Catmur, C., Liepelt, R., Brass, M., & Heyes, C. (2008). Experience-based priming of body parts: A study of action imitation. *Brain Research, 1217*, 157-170.

Grezes, J., Armony, J. L., Rowe, J., & Passingham, R. E. (2003). Activations related to "mirror" and "canonical" neurones in the human brain: an fMRI study. *Neuroimage, 18*(4), 928-937.

Grill-Spector, K., Henson, R., & Martin, A. (2006). Repetition and the brain: neural models of stimulus-specific effects. *Trends in Cognitive Sciences, 10*(1), 14-23.

Guttman, N., & Kalish, H. I. (1956). Discriminability and stimulus generalization. *Journal of Experimental Psychology, 51*(1), 79-88.

Haruno, M., Wolpert, D. M., & Kawato, M. (2001). Mosaic model for sensorimotor learning and control. *Neural Computation, 13*(10), 2201-2220.

Haslinger, B., Erhard, P., Altenmuller, E., Schroeder, U., Boecker, H., & Ceballos-Baumann, A. O. (2005). Transmodal sensorimotor networks during action observation in professional pianists. *Journal of Cognitive Neuroscience, 17*(2), 282-293.

Hebb, D. (1949). *The Organization of Behaviour*. New York: Wiley.

Hecht, H., Vogt, S., & Prinz, W. (2001). Motor learning enhances perceptual judgment: a case for action-perception transfer. *Psychological Research, 65*(1), 3-14.

Heiser, M., Iacoboni, M., Maeda, F., Marcus, J., & Mazziotta, J. C. (2003). The essential role of Broca's area in imitation. *European Journal of Neuroscience, 17*(5), 1123-1128.

Heyes, C. (2010). Where do mirror neurons come from? *Neuroscience and Biobehavioral Reviews, 34*(4), 575-583.

Heyes, C. (2011). Automatic imitation. *Psychological Bulletin, 137*(3), 463-483.

Heyes, C. (in press). What can imitation do for cooperation? In B. Calcott, R. Joyce & K. Sterelny (Eds.), *Signalling, Commitment and Emotion*: MIT Press.

Heyes, C., Bird, G., Johnson, H., & Haggard, P. (2005). Experience modulates automatic imitation. *Cognitive Brain Research, 22*(2), 233-240.

Heyes, C., & Galef, B. G. (Eds.). (1996). *Social Learning in Animals: The Roots of Culture*. San Diego: Academic Press.

Heyes, C., & Ray, E. D. (2000). What is the significance of imitation in animals? *Advances in the Study of Behavior, 29*, 215-245.

Hickok, G., & Hauser, M. (2010). (Mis)understanding mirror neurons. *Current Biology, 20*(14), R593-594.

Hill, H., & Johnston, A. (2001). Categorizing sex and identity from the biological motion of faces. *Current Biology, 11*(11), 880-885.

Holland, P. C. (1983). Occasion-setting in Pavlovian feature positive discriminations. In M. L. Commons, R. J. Herrnstein & A. R. Wagner (Eds.), *Quantitative Analysis of Behavior: Discrimination Processes* (Vol. 4). New York: Ballinger.

Hommel, B., Musseler, J., Aschersleben, G., & Prinz, W. (2001). The theory of event coding (TEC): a framework for perception and action planning. *Behavioral Brain Sciences, 24*(5), 849-878.

Iacoboni, M. (2009). Imitation, empathy, and mirror neurons. *Annual Review of Psychology, 60*, 653-670.

Iacoboni, M., Molnar-Szakacs, I., Gallese, V., Buccino, G., Mazziotta, J. C., & Rizzolatti, G. (2005). Grasping the intentions of others with one's own mirror neuron system. *Plos Biology, 3*(3), e79.

Iacoboni, M., Woods, R. P., Brass, M., Bekkering, H., Mazziotta, J. C., & Rizzolatti, G. (1999). Cortical mechanisms of human imitation. *Science, 286*(5449), 2526-2528.

Jackson, P. L., Meltzoff, A. N., & Decety, J. (2006). Neural circuits involved in imitation and perspective-taking. *Neuroimage, 31*(1), 429-439.

Jacobson, S. W. (1979). Matching behavior in the young infant. *Child Development, 50*(2), 425-430.

James, W. (1890). *The principles of psychology*. New York,: H. Holt and company.

Johansson, G. (1973). Visual-perception of biological motion and a model for its analysis. *Perception & Psychophysics, 14*(2), 201-211.

Johnston, A., McOwan, P. W., & Benton, C. P. (1999). Robust velocity computation from a biologically motivated model of motion perception. *Proceedings of the Royal Society B-Biological Sciences, 266*(1418), 509-518.

Johnston, A., McOwan, P. W., & Buxton, H. (1992). A computational model of the analysis of some first-order and second-order motion patterns by simple and complex cells. *Proceedings of the Royal Society B-Biological Sciences, 250*(1329), 297-306.

Jokisch, D., Daum, I., & Troje, N. F. (2006). Self recognition versus recognition of others by biological motion: Viewpoint-dependent effects. *Perception, 35*(7), 911-920.

Jones, S. S. (1996). Imitation or exploration? Young infants' matching of adults' oral gestures. *Child Development, 67*(5), 1952-1969.

Jones, S. S. (2006). Exploration or imitation? The effect of music on 4-week-old infants' tongue protrusions. *Infant Behavior & Development, 29*(1), 126-130.

Jones, S. S. (2009). The development of imitation in infancy. *Philosophical Transactions of the Royal Society B-Biological Sciences, 364*(1528), 2325-2335.

Kaitz, M., Meschulachsarfaty, O., Auerbach, J., & Eidelman, A. (1988). A reexamination of newborns ability to imitate facial expressions. *Developmental Psychology, 24*(1), 3-7.

Keysers, C., & Gazzola, V. (2006). Towards a unifying neural theory of social cognition. *Progress in Brain Research 156*, 379-401.

Keysers, C., Kohler, E., Umilta, M. A., Nanetti, L., Fogassi, L., & Gallese, V. (2003). Audiovisual mirror neurons and action recognition. *Experimental Brain Research, 153*(4), 628-636.

Keysers, C., & Perrett, D. I. (2004). Demystifying social cognition: a Hebbian perspective. *Trends in Cognitive Sciences, 8*(11), 501-507.

Kilner, J. M., Friston, K. J., & Frith, C. D. (2007a). The mirror-neuron system: a Bayesian perspective. *Neuroreport, 18*(6), 619-623.

Kilner, J. M., Friston, K. J., & Frith, C. D. (2007b). Predictive coding: an account of the mirror neuron system. *Cognitive Processing, 8*(3), 159-166.

Kilner, J. M., Neal, A., Weiskopf, N., Friston, K. J., & Frith, C. D. (2009). Evidence of mirror neurons in human inferior frontal gyrus. *Journal of Neuroscience, 29*(32), 10153-10159.

Knight, B., & Johnston, A. (1997). The role of movement in face recognition. *Visual Cognition, 4*(3), 265-273.

Knoblich, G., & Flach, R. (2003). Action identity: evidence from self-recognition, prediction, and coordination. *Consciousness and Cognition, 12*(4), 620-632.

Knoblich, G., & Prinz, W. (2001). Recognition of self-generated actions from kinematic displays of drawing. *Journal of Experimental Psychology: Human Perception and Performance, 27*(2), 456-465.

Kohler, E., Keysers, C., Umilta, M. A., Fogassi, L., Gallese, V., & Rizzolatti, G. (2002). Hearing sounds, understanding actions: action representation in mirror neurons. *Science, 297*(5582), 846-848.

Lakin, J. L., & Chartrand, T. L. (2003). Using nonconscious behavioral mimicry to create affiliation and rapport. *Psychological Science, 14*(4), 334-339.

Larsson, J., & Smith, A. T. (2011). fMRI repetition suppression: neuronal adaptation or stimulus expectation? *Cerebral Cortex*.

Legerstee, M. (1991). The role of person and object in eliciting early imitation. *Journal of Experimental Child Psychology, 51*(3), 423-433.

Leighton, J., & Heyes, C. (2010). Hand to mouth: automatic imitation across effector systems. *Journal of Experimental Psychology-Human Perception and Performance, 36*(5), 1174-1183.

Lepage, J. F., & Theoret, H. (2007). The mirror neuron system: grasping others' actions from birth? *Developmental Science, 10*(5), 513-523.

Lhermitte, F. (1983). Utilization behavior and its relation to lesions of the frontal lobes. . *Brain, 106*, 237-255.

Locke, J. (1694). *An essay concerning human understanding* (Second edition ed.). London: Eliz. Holt for Thomas Basset.

Longo, M. R., Kosobud, A., & Bertenthal, B. I. (2008). Automatic imitation of biomechanically possible and impossible actions: effects of priming movements versus goals. *Journal of Experimental Psychology: Human Perception and Performance, 34*(2), 489-501.

Lotze, H. (1852). *Medicinische psychologie oder physiologie der seele*. Leipzig: Weidmann.

Loula, F., Prasad, S., Harber, K., & Shiffrar, M. (2005). Recognizing people from their movement. *Journal of Experimental Psychology: Human Perception and Performance, 31*(1), 210-220.

Mackintosh, N. J. (1975). A theory of attention: Variations in the associability of stimuli with reinforcement. *Psychological Review, 82*, 276-298.

Macmillan, N. A., & Creelman, C. D. (1991). *Detection theory: a user's guide*. New York: Cambridge University Press.

Maeda, F., Kleiner-Fisman, G., & Pascual-Leone, A. (2002). Motor facilitation while observing hand actions: Specificity of the effect and role of observer's orientation. *Journal of Neurophysiology, 87*(3), 1329-1335.

Malatesta, C. Z., & Haviland, J. M. (1982). Learning display rules: The socialization of emotion expression in infancy. *Child Development, 53*(4), 991-1003.

Margulis, E. H., Mlsna, L. M., Uppunda, A. K., Parrish, T. B., & Wong, P. C. (2009). Selective neurophysiologic responses to music in instrumentalists with different listening biographies. *Human Brain Mapping, 30*(1), 267-275.

Maurer, D., Le Grand, R., & Mondloch, C. J. (2002). The many faces of configural processing. *Trends in Cognitive Sciences, 6*(6), 255-260.

Meltzoff, A. N., & Decety, J. (2003). What imitation tells us about social cognition: a rapprochement between developmental psychology and cognitive neuroscience. *Philosophical Transactions of the Royal Society B-Biological Sciences, 358*(1431), 491-500.

Meltzoff, A. N., & Moore, M. K. (1977). Imitation of facial and manual gestures by human neonates. *Science, 198*(4312), 74-78.

Meltzoff, A. N., & Moore, M. K. (1983). Newborn-infants imitate adult facial gestures. *Child Development, 54*(3), 702-709.

Meltzoff, A. N., & Moore, M. K. (1989). Imitation in newborn-infants - exploring the range of gestures imitated and the underlying mechanisms. *Developmental Psychology, 25*(6), 954-962.

Meltzoff, A. N., & Moore, M. K. (1994). Imitation, memory, and the representation of persons. *Infant Behavior & Development, 17*(1), 83-99.

Meltzoff, A. N., & Moore, M. K. (1997). Explaining facial imitation: A theoretical model. *Early Development & Parenting, 6*(3-4), 179-192.

Mukamel, R., Ekstrom, A. D., Kaplan, J., Iacoboni, M., & Fried, I. (2010). Single-neuron responses in humans during execution and observation of actions. *Current Biology, 20*(8), 750-756.

Nagy, E., Compagne, H., Orvos, H., Pal, A., Molnar, P., Janszky, I., et al. (2005). Index finger movement imitation by human neonates: motivation, learning, and left-hand preference. *Pediatric Research, 58*(4), 749-753.

Nelissen, K., Luppino, G., Vanduffel, W., Rizzolatti, G., & Orban, G. A. (2005). Observing others: multiple action representation in the frontal lobe. *Science, 310*(5746), 332-336.

Nelson, J. B. (2002). Context specificity of excitation and inhibition in ambiguous stimuli. *Learning and Motivation, 33*(2), 284-310.

Nelson, J. B., Sanjuan Mdel, C., Vadillo-Ruiz, S., Perez, J., & Leon, S. P. (2011). Experimental renewal in human participants. *Journal of Experimental Psychology-Animal Behavior Processes, 37*(1), 58-70.

Newman-Norlund, R. D., Ondobaka, S., van Schie, H. T., van Elswijk, G., & Bekkering, H. (2010). Virtual lesions of the IFG abolish response facilitation for biological and non-biological cues. *Frontiers in Behavioral Neuroscience 4*, 5.

O'Toole, A. J., Roark, D. A., & Abdi, H. (2002). Recognizing moving faces: a psychological and neural synthesis. *Trends in Cognitive Sciences, 6*(6), 261-266.

Patuzzo, S., Fiaschi, A., & Manganotti, P. (2003). Modulation of motor cortex excitability in the left hemisphere during action observation: a single- and paired-pulse transcranial magnetic stimulation study of self- and non-self-action observation. *Neuropsychologia, 41*(9), 1272-1278.

Pawlby, S. J. (1977). Imitative interaction. In H. Schaffer (Ed.), *Studies in mother-infant interaction* (pp. 203-224). New York: Academic Press.

Pearce, J. M., & Hall, G. (1980). A model for Pavlovian learning: Variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychological Review, 87*, 532-552.

Peck, C. A., & Bouton, M. E. (1990). Context and performance in aversive-to-appetitive and appetitive-to-Aversive Transfer. *Learning and Motivation, 21*(1), 1-31.

Prasad, S., & Shiffrar, M. (2009). Viewpoint and the recognition of people from their movements. *Journal of Experimental Psychology: Human Perception and Performance, 35*(1), 39-49.

Press, C., Bird, G., Flach, R., & Heyes, C. (2005). Robotic movement elicits automatic imitation. *Cognitive Brain Research, 25*(3), 632-640.

Press, C., Bird, G., Walsh, E., & Heyes, C. (2008). Automatic imitation of intransitive actions. *Brain and Cognition, 67*(1), 44-50.

Press, C., Catmur, C., Widman, H., Heyes, C., & Bird, G. (under review). fMRI adaptation reveals geometric shape 'mirror neurons'.

Press, C., Gillmeister, H., & Heyes, C. (2007). Sensorimotor experience enhances automatic imitation of robotic action. *Proceedings of the Royal Society B-Biological Sciences, 274*(1625), 2509-2514.

Press, C., Heyes, C., & Kilner, J. M. (2011). Learning to understand others' actions. *Biology Letters, 7*(3), 457-460.

Prinz, W. (1997). Perception and action planning. *European Journal of Cognitive Psychology, 9*(2), 129-154.

Ray, E., & Heyes, C. (2011). Imitation in infancy: the wealth of the stimulus. *Developmental Science, 14*(1), 92-105.

Reissland, N. (1988). Neonatal imitation in the first hour of Life: observations in rural Nepal. *Developmental Psychology, 24*(4), 464-469.

Rescorla, R. A. (1968). Probability of shock in the presence and absence of CS in fear conditioning. *Journal of Comparative and Physiological Psychology, 66*(1), 1-5.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and non-reinforcement In A. H. Black & W. F. Prokasy (Eds.), *Classical Conditioning II: Current Research and theory* (pp. 64-99). New York: Appleton-Century-Crofts.

Rizzolatti, G., & Arbib, M. A. (1998). Language within our grasp. *Trends in Neurosciences, 21*(5), 188-194.

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

Rizzolatti, G., Fadiga, L., Gallese, V., & Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Cognitive Brain Research, 3*(2), 131-141.

Rizzolatti, G., & Luppino, G. (2001). The cortical motor system. *Neuron, 31*(6), 889-901.

Schutz-Bosbach, S., & Prinz, W. (2007). Perceptual resonance: action-induced modulation of perception. *Trends in Cognitive Sciences, 11*(8), 349-355.

Sevdalis, V., & Keller, P. E. (2010). Cues for self-recognition in point-light displays of actions performed in synchrony with music. *Consciousness and Cognition, 19*(2), 617-626.

Shanks, D. R. (1989). Selectional processes in causality judgment. *Memory and Cognition, 17*(1), 27-34.

Shanks, D. R., & Dickinson, A. (1991). Instrumental judgment and performance under variations in action-outcome contingency and contiguity. *Memory and Cognition, 19*(4), 353-360.

Soussignan, R., Courtial, A., Canet, P., Danon-Apter, G., & Nadel, J. (2011). Human newborns match tongue protrusion of disembodied human and robotic mouths. *Developmental Science, 14*(2), 385-394.

Strafella, A. P., & Paus, T. (2000). Modulation of cortical excitability during action observation: a transcranial magnetic stimulation study. *Neuroreport, 11*(10), 2289-2292.

Tanaka, J. W., & Farah, M. J. (1993). Parts and wholes in face recognition. *Quarterly Journal of Experimental Psychology, 46*(2), 225-245.

Tinbergen, N. (1953). *The herring gulls' world; a study of the social behaviour of birds*. London: Collins.

Tomasello, M. (1996). Do apes ape? In C. M. Heyes & B. J. G. Jr (Eds.), *Social learning in animals: The roots of culture*. San Diego: Academic Press.

van Baaren, R., Holland, R. W., Kawakami, K., & van Knippenberg, A. (2004). Mimicry and pro-social behaviour. *Psychological Science, 15*, 71-74.

van Baaren, R., Janssen, L., Chartrand, T., & Dijksterhuis, A. (2009). Where is the love? The social aspects of mimicry. *Philosophical Transactions of the Royal Society B-Biological Sciences, 364*, 2381-2389.

Van Gucht, D., Vansteenwegen, D., Beckers, T., & Van den Bergh, O. (2008). Return of experimentally induced chocolate craving after extinction in a different context: divergence between craving for and expecting to eat chocolate. *Behaviour Research and Therapy, 46*(3), 375-391.

Vansteenwegen, D., Hermans, D., Vervliet, B., Francken, G., Beckers, T., Baeyens, F., et al. (2005). Return of fear in a human differential conditioning paradigm caused by a return to the original acquistion context. *Behaviour Research and Therapy, 43*(3), 323-336.

Vinter, A. (1986). The role of movement in eliciting early imitations. *Child Development, 57*(1), 66-71.

Vogt, S., Buccino, G., Wohlschlager, A. M., Canessa, N., Shah, N. J., Zilles, K., et al. (2007). Prefrontal involvement in imitation learning of hand actions: effects of practice and expertise. *Neuroimage, 37*(4), 1371-1383.

Voss, M., Ingram, J. N., Haggard, P., & Wolpert, D. M. (2006). Sensorimotor attenuation by central motor command signals in the absence of movement. *Nature Neuroscience, 9*(1), 26-27.

Voss, M., Ingram, J. N., Wolpert, D. M., & Haggard, P. (2008). Mere expectation to move causes attenuation of sensory signals. *PLoS One, 3*(8), e2866.

Wagner, A. R. (1981). SOP: A model of automatic memory processing in animal behavior. In N. E. Spear & R. R. Miller (Eds.), *Information Processing in Animals: Memory Mechanisms* (pp. 5-47). Hillsdale, NJ: Erlbaum.

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

Wilson, P. N., & Pearce, J. M. (1992). A configural analysis for feature-negative discrimination learning. *Journal of Experimental Psychology: Animal Behavior Processes, 18*, 265-277.

Wolpert, D. M., Doya, K., & Kawato, M. (2003). A unifying computational framework for motor control and social interaction. *Philosophical Transactions of the Royal Society B-Biological Sciences, 358*(1431), 593-602.

Yin, R. K. (1969). Looking at upside-down faces. *Journal of Experimental Psychology, 81*(1), 141-145.

Young, A. W., Hellawell, D., & Hay, D. C. (1987). Configurational information in face perception. *Perception, 16*(6), 747-759.