CrossMark

Association but not Recognition: an Alternative Model for Differential Imitation from 0 to 2 Months

Stefano Vincini[1] · Yuna Jhang[2]

Abstract Skepticism toward the existence of neonatal differential imitation is fostered by views that assign it an excessive significance, making it foundational for social cognition. Moreover, a misleading theoretical framework may generate unwarranted expectations about the kinds of findings experimentalists are supposed to look for. Hence we propose a theoretical analysis that may help experimentalists address the empirical question of whether early differential imitation really exists. We distinguish three models of early imitation. The first posits automatic visuo-motor links evolved for sociocognitive functions and we call it Genetically Programmed Direct Matching (GPDM). The second is Meltzoff and Moore's Active Intermodal Matching (AIM), which postulates a comparison between the acts of self and other. The third is the alternative we propose and we call it "Association by Similarity Theory" (AST), as it relies on the tacit functioning of this domain-general process. AST describes early imitation merely as the differential induction or elicitation of behaviors that already tend to occur spontaneously. We focus on the contrast between AIM and AST, and argue that AST is preferable to AIM for two reasons. First, AST is more parsimonious and more plausible, especially because it does not require infants to be able to recognize self-other similarities. Second, whereas the extant findings tend to disqualify AIM, AST can account for them adequately. Furthermore, we suggest that AST has the potential to give new impulse to empirical research because it discriminates promising lines of inquiry from unproductive ones.

✉ Stefano Vincini
   stefano.vincini@gmail.com

[1] Present address: Instituto de Investigaciones Filosóficas, Universidad Nacional Autónoma de México, Circuito Mtro. Mario de la Cueva, Ciudad de la Investigación en Humanidades, Zona Cultural, Ciudad Universitaria, Deleg. Coyoacán C.P. 04510 Cd.Mx., Mexico

[2] Department of Speech Language Pathology and Audiology, Chung Shan Medical University, Taichung, Taiwan

🖄 Springer

1 Introduction

After forty years, the debate on "neonatal imitation" (NI) is still open. The very question of whether neonates imitate at all is far from being settled. Although NI was taken to be proven beyond doubt (Heimann 2002; Meltzoff 2005; Nagy and Molnar 2004; Trevarthen and Aitken 2001), in recent years more and more studies have come to deny its existence (Anisfeld 2005; Keven and Akins 2016; Oostenbroek et al. 2016; Ray and Heyes 2011). In addition, as various authors have lamented (Jones 2009; Lodder et al. 2014; Oostenbroek et al. 2013), the equivocality of the findings goes together with a lack of clarity at the level of explanations. Indeed, the main consideration motivating the present paper is that conceptual confusions in the theoretical domain hinder not just a realistic understanding of the phenomenon, but also the progress of empirical research.

There is a tendency to lump together different theoretical alternatives, which look similar only at first glance. Unwarranted theoretical frameworks lead to inappropriate empirical methodologies and generate wrong expectations about the kind of findings one is supposed to look for. Views conferring an exaggerated significance on NI create skepticism toward the reality of the phenomenon, and skepticism can dissuade experimentalists from devoting their resources to seek a solid answer to primary empirical questions. All this indicates that there is much conceptual work that needs to be done. Thus the purpose of the present paper is primarily theoretical: to distinguish three different theoretical accounts of NI and indicate what we believe to be the most credible one. We can expect empirical research to find out whether NI exists only under the guidance of a plausible theory of what NI might be.

The first theoretical model we identify can be designated as Genetically Programmed Direct Matching (GPDM). The second was articulated by Meltzoff and Moore's (1997) under the name of Active Intermodal Matching model (AIM). The third is the one we propose and we call it Association by Similarity Theory (AST). We focus on the contrast between AIM and AST, and argue that AST is preferable to AIM for two reasons. First, as a theoretical model, AST has merits that AIM lacks (especially parsimony and plausibility). Second, AST better accounts for the extant findings. In addition, we suggest that AST has the potential to give new impulse to empirical research on NI.

To begin, in section 2, we make preliminary remarks that are necessary to clarify the empirical and theoretical issues at stake. After recalling the operational definition of NI, we specify the time frame targeted by the three models we consider (2.1); we examine how to formulate the problem of explaining NI (2.2), and we motivate why we do not address certain questions in this paper (2.3).

2 Preliminary Clarifications

2.1 Operational Definition and Time Frame

In studies of NI, imitation was usually operationalized as follows: infants produce specific gestures more frequently on statistical terms when the corresponding gestures are presented than when other gestures are presented (e.g. infants produce more mouth openings when the mouth opening model is presented than when the

tongue protrusion or lip protrusion model is presented). The operational definition entailed an essential reference to a plurality of gestures exhibiting the comparative increase just described. In their inaugural study, Meltzoff and Moore (1977) were well aware that, if only one gesture was matched, this matching behavior would have been more parsimoniously interpreted as an arousal response than as an imitative act. If, however, more than one gesture is solicited more by the corresponding model than by other models, this differential behavior cannot be explained by mere arousal (cf. Anisfeld et al. 2001, p. 113). Therefore, NI was operationalized as "differential imitation" (Meltzoff and Moore 1977, p. 76).

This operational definition has an important feature that may be easily overlooked. One way to make it explicit is to refer to two key experiments in the NI literature—Experiment 2 in Meltzoff and Moore (1977) and Meltzoff and Moore's (1994), which investigated imitation of tongue protrusion (TP) and mouth opening (MO). In both experiments, infants produced more TP than MO in response to the MO model. Meltzoff and Moore (1994) even found that, when MO was presented, the sum of mean frequencies (31.90) of infant TP was well over twice the sum of mean frequencies of MO (13.70). Considering these results somewhat naively, we would say that they do not count as MO imitation because, according to our ordinary notion of imitation, infants would be imitating MO only if they were producing more MO than TP in response to the MO model. However, Meltzoff and Moore's (1977; 1994) results *do* count as MO imitation because they meet the requirements of the operational definition. These experiments ascertained differential imitation for both MO and TP. Specifically, in Meltzoff and Moore's (1977), infants produced more TP in response to TP than in response to MO or during baseline; infants also produced more MO in response to MO than in response to TP or during baseline (see Meltzoff and Moore 1977, p. 77, Fig. 4). Meltzoff and Moore (1994) found the same pattern of results, although differential imitation did not reach statistical significance for the frequency of occurrence of MO. Differential imitation for MO reached statistical significance with respect to MO duration, as infants produced longer MO in response to MO than in response to TP or during baseline.

In short, the lesson that should be learned from the operational definition is that NI may not match our ordinary notion of imitation. Even if infants produce more TP than MO in response to MO, their behavior may still be described as imitation because what counts in the operational definition is not the comparison between gestures (what gesture infants produce more within a condition), but the comparison between conditions (whether a gesture is produced more in response to the corresponding model than in other stimulus conditions).

As many studies indicate, "early" or "neonatal" imitation designates a phenomenon concerning infants in the first two month of postnatal development (Meltzoff and Moore 1997). The term "neonatal" is often used in a wide sense to indicate this time frame (Heimann et al. 1989; Heimann 2002; Nagy et al. 2013; Oostenbroek et al. 2016; Ullstadius 1998). Indeed, Meltzoff and Moore (1997) are explicit in specifying that their model targets imitation in the first two months. The other two models we discuss in this paper (GPDM and AST) also apply to imitation in the same time frame, and so are directly comparable with Meltzoff and Moore's. Accordingly, in the present paper, we use the acronym "NI" as shorthand for "differential imitation in the first eight weeks of postnatal development."

2.2 How to Formulate the Problem of Explaining Imitation

Various authors identify a minimal requirement for imitation in the causal relationship between a visual perception and the activation of a corresponding "action plan" (Prinz et al. 2009, p. 48) or "motor programme" (Jones 2009, p. 2329). As Meltzoff and Moore (1994, p. 83) put it, "in order to imitate, the child must see the adult's actions, use this visual perception as a basis for an action plan, and execute the motor output." This minimal requirement is captured by the operational definition, as this guarantees that a specific visual perception tends to activate the corresponding action plan more than other visual perceptions. Explanations of imitation differ precisely on how and why a visual representation tends to activate the corresponding action representation (e.g. how the visual representation of MO tends to activate the action representation of MO). In formulating the problem of the possible explanations for imitation, one has to make different explanations comparable and, at the same time, seek not to exclude a priori reasonable accounts of the relationship between visual and action representations. To this end, we adopt the following terminology.

A *visual representation* of a modeled gesture (e.g. mouth opening) denotes the processing of visual inputs relative to the gesture in abstraction from the integration with other modalities. In other words, a visual representation refers to the features of the stimulus that are tracked by processing the visual input: shades of color, spatial configurations, motion kinematics, etc. We stress that by visual representation we mean what is also designated as "mere visual description" (Rizzolatti and Sinigaglia 2010) and as "visual analysis of the observed action" (Giese and Rizzolatti 2015). That is: a visual representation can certainly be combined with other processes to form more complex representations; however, a visual representation per se does not contain information that derives from other senses but is not tracked by vision. Accordingly, when a subject sees an action, we say that the subject has a (active) visual representation of the action. Equivalently, we may say that the subject has a (present) "visual experience" of the action. This experience includes, for example, the spatial features of the effector involved and its kinematics. Lastly, a visual representation can be stored in memory and is potentially reactivated by new cognitive processes such as visual imagination or a new perception of the same gesture.

An *action representation* is the counterpart of a visual representation in the domain of action execution. An action representation comprises (some of) the motor codes originating the action and the information that is derived from the proprioceptive experience of the action (in abstraction from possible integration with data from other modalities).[1] For example, an action representation can refer to the motion kinematics of the action, to the spatial configurations assumed by bodily parts, and to which movement reaches the action's end state. When a subject executes an action, we say

---

[1] Our definition is deliberately inclusive. Primarily, we want the definition to capture the representational basis of an *impulse* to a specific action. Therefore, we say that an action representation is activated even if not all the motor codes of the action are activated. Indeed, an awakened impulse may not be complied with because of antagonistic conditions and, in this case, it is reasonable to assume that not all the motor codes of the action have been activated (in this regard, see also the use of the expression "action planning" in Zoia et al. 2007). However, we also want the expression to be able to designate the processing relative to a full-fledged action execution. Here the impulse to act is complied with, all motor codes originating the action are (presumably) activated, and the proprioceptive feedback (normally) conforms to what could be expected from the action.

that it has an (active) action representation. Interchangeably, we may say that it has a (present) "action experience." Importantly, an action representation can be stored in memory and reactivated in later processing such as mere action planning or a new full-fledged action execution (cf. Prinz et al. 2009, p. 45). In this way, our definition is compatible with the one provided by Hommel and Elsner (2009, p. 382), for whom a minimal action representation comprises "sensorimotor associations between the per-ceptual [e.g. proprioceptive] codes of a particular action and the motor program realizing them." Including a reference to learning in the definition of an action representation is particularly appropriate when dealing with NI because all relevant studies investigate gestures spontaneously produced by infants.[2]

The advantage of defining visual and action representations in these terms is that they are neutral with respect to three general approaches to the relationships between visual perception and action. The first is the "separate coding" approach, which takes visual and action representation to belong to different domains so that none of the codes or representational components employed in visual representations are also employed in action representations. The second is the "common coding" approach, in which the same resources that are used to represent specific action features in vision are used to represent those features when the action is proprioceptively experienced. Hence, in this approach, visual and action representation share some codes in common (Prinz 1990, 1997). The third is a hybrid approach. Here perception and action are separate domains and no visual code is identical with an action code, yet perception and action are made commensurable by a third domain evolved specifically to compare visual and action codes while at the same time maintaining the distinction between them.

We formulate the problem of explaining imitation as that of charactherizing how a visual representation contributes to activating a corresponding action representation. This formulation does not discriminate between different approaches to the relation-ships between visual and action representations. Indeed, it can be equally applied to three models of NI that correspond to the separate coding, common coding, and hybrid approches. These three models (presented below) are GPDM, AST, and AIM, respectively.

2.3 Motivating Exclusions from our Discussion

Notice that we do not examine the arousal hypothesis for the NI findings (Anisfeld 2005; Jones 2009; Ray and Heyes 2011). In another paper (Vincini et al. 2017), we argue that this hypothesis is a viable account of the findings, but we do not discuss it here because it relies on the empirical claim that infants reliably match only one gesture. Arousal is not an explanation of differential imitation. Furthermore, apart from some critical remark, we do not consider the relationship between NI and social cognition because we discuss it in another paper (Vincini et al. 2017). Finally, we exclude a thorough examination of topics such as the neural substrate of NI and its presence in non-human species for no other reasons than space limitations.

---

[2] Even Meltzoff and Moore (1994), who investigated the "unusual" tongue-protrusion-to-the-side behavior, found that, in a 90-s test period, such behavior was spontaneously produced by 6 infants out of 30 who had never seen the tongue-protrusion-to-the-side model.

3 Genetically Programmed Direct Matching (GPDM)

We now start our analysis of explanations of differential imitation.[3] Meltzoff and Decety (2003) and Jones (2009) present a possible explanation of NI in association with the classical genetic account of mirror neurons (Cook et al. 2014; Tramacere et al. 2015) and distinguish it from AIM—we call that explanation GPDM. In order to explain how visual representations are linked to the corresponding action representations, GPDM postulates that these links are coded by the genes. In our evolutionary past, natural selection established links between certain visual representations and their action counterparts, so that, when one of these visual processes occurs, the corresponding action representation is automatically activated. For an infant who visually perceives a modeled gesture, the activation of the corresponding action processing would constitute a motivation to act. But imitation is never compulsory (Meltzoff and Moore 1997); so imitation occurs only if the motivation is complied with. Figure 1 shows the basic conceptual structure of GPDM.

Because our primary focus in this paper is the contrast between AIM and AST, we examine GPDM only insofar as one cannot understand AIM and AST if it is not clear how they differ from GPDM. Hence, in this subsection, we anticipate the main differences that differentiate AIM and AST from GPDM.

Both Meltzoff and Decety (2003) and Jones (2009) make it clear that a model like GPDM requires infants to have significantly fewer cognitive abilities than what AIM assumes. In particular, Meltzoff and Decety (p. 494) note that, in contrast to a mirror neuron based account, AIM posits "an active comparison and lack of confusion between self and other." The result of the comparison is the recognition of "both the similarity and the distinction between actions of the self and other" (Meltzoff and Decety 2003, p. 494). Specifically, recognition of similarity has the form "Something familiar! That seen event is like this felt event" or "Here is something like me" (Meltzoff 2002, 2007a, b, 2005, 2010; Meltzoff and Decety 2003). For Meltzoff and Decety, this recognition of similarities and differences cannot be implemented by mirror neurons alone. Rather, it likely requires a specific activation of the inferior parietal lobule. Accordingly, we can express the main difference between AIM and GPDM by saying that in GPDM visual representations directly activate the corresponding action representations without entailing any process of comparison. Consequently, in GPDM a baby who imitates does not need to recognize that the acts of the other are like the acts of the self. The baby merely has an impulse to act in a particular way and this impulse has emerged because of a genetically wired up connection; if the baby complies with the impulse, it executes an act we designate as imitative.[4]

In opposition to AST, GPDM rejects that the connection between visual and action representations occurs in virtue of a domain-general process of association. In particular,

---

[3] All three models can share the assumption that infants prefer social stimuli such as faces, eyes, or traits of the human voice to other non-social stimuli. This common assumption allows all of them to explain why it is not easy to solicit imitation with non-social stimuli (Legerstee 1991)—e.g. by stating that the non-social stimuli do not present interesting features that attract the infant's attention. Indeed, the non-social stimuli presented by Legerstee (1991) do not present the kinds of perceptual features that are attractive to newborns (Simion et al. 2011).

[4] The distinction between AIM and a mirror neuron based model is reiterated in Meltzoff (2009, p. 38) where it is argued that the latter would not account for presumed features of NI such as "response correction" and "the imitation of novel acts."
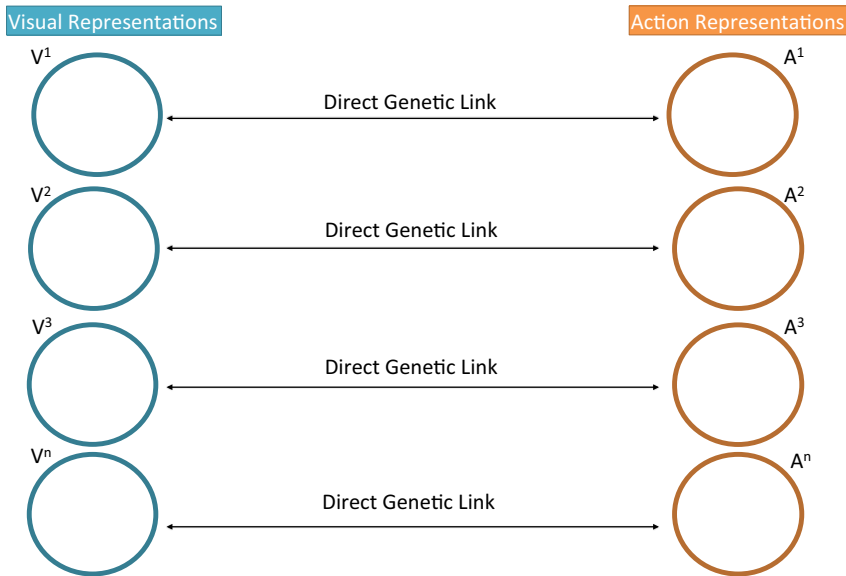
Genetically Programmed Direct Matching (GPDM)

Fig. 1 Conceptual schematic of the GPDM model. Visual representations (Vn) are connected with corresponding action representations (An) through automatic connections encoded by the genes. These genetic links are not instantiations of a domain-general process of association; rather, they were specifically selected for social or socio-cognitive functions. This schematic is comparable in format with those of AIM and AST in Fig. 2. The expressions "visual representation" and "action representation" are defined in subsection 2.2. (From Vincini et al. 2017)

GPDM denies a role for association by similarity, i.e. it denies that visual representations activate corresponding action representations because they share representational resources in common. GPDM is a separate coding approach (2.2). In GPDM, there is no functional role for a representational overlap between visual and action representations. Quite the opposite, GPDM states that the function of activating corresponding action representations is fulfilled by specialized mechanisms evolved for social or socio-cognitive functions such as imitation and action understanding. This points to a pregnant notion of nativism that can be used to distinguish between different models of NI.[5] Both GPDM and AIM are nativist explanations precisely in the sense that the mechanisms for matching visual and action representations evolved specifically for social or socio-cognitive functions. Yet AST is a non-nativist explanation because it does not presuppose a mechanism evolved for a specific domain, but rather relies on a general mechanism of association.

Furthermore, in GPDM proprioceptive experience does not have an indispensable role as in AST (and AIM—see section 5.3). GPDM can posit that the relevant genetic links connect visual representations just with the motor codes of the corresponding action

---

[5] Notoriously, terms like "nativism" or "innate" have many different meanings (Griffiths 2002; Maclaurin 2002). For example, by "innate" one might mean "existing from birth" or "something existing prior to birth" (Gallagher 2005, p. 73). This usage of the term does not distinguish between models of NI because all models must posit that the capacity to activate corresponding action representations exists at birth. We submit that notion of nativism we adopt in this paper is a pregnant one because it helps distinguish different models of NI.

representations, not with their proprioceptive components. Indeed, GPDM is compatible with the claim that infant proprioception is too vague and does not discriminate the morphokinetic features of the different actions. In contrary to AST (and AIM), the specialized mechanisms postulated by GPDM could in principle activate the corresponding action representations independently of any role played by proprioceptive experience.

4 Active Intermodal Matching (AIM)

4.1 A Comparison Computation Foundational for Social Cognition

For many years, Meltzoff and Moore's AIM has been the dominant model for explaining NI. In our exposition, we refer primarily to Meltzoff and Moore (1997) because in subsequent years Meltzoff repeatedly mentions this article as the most detailed presentation of his model (Meltzoff and Decety 2003; Meltzoff 2009, 2013). It is critical to examine the model in full detail and consider what specific predictions derive from it.

The first assumption Meltzoff and Moore (1997) make is that NI requires "organ identification." Meltzoff and Moore claim that, when presented with a particular action, most infants respond at first with a partial activation of the corresponding body part (e.g. they slightly elevate the tongue in the oral cavity). They suppose that this partial response indicates the presence of a first cognitive step. The step is to identify the body part that corresponds to the visually perceived action: "infants isolate *what* part of their body to move before *how* to move it" (p. 183).

Another assumption in Meltzoff and Moore (1997) is that infants learn to map specific configural relations between organs (e.g. "tongue-beyond-lips") to the specific movements that achieve them. This learning process occurs by means of self-generated activity ("body babbling") and begins before birth. As Meltzoff and Moore put it, through proprioceptive monitoring of their own movements infants build up a "directory" where each entry connects a particular movement to the final bodily state attained by it. To use the terminology introduced in 2.2, we would say that infants acquire a set of action representations through the experience of their own actions. Indeed, we said that an action representation can encode information about which movement reaches an action's end state. Thus, an entry in Meltzoff and Moore's directory is equivalent to what we called an action representation.

Hence we come to the central supposition of the model, "the crux of the AIM hypothesis" (Meltzoff 1999, p. 254).[6] NI is a goal-directed behavior where the goal constitutes the criterion for success; successes or failures are ratified by a computational process of comparison. The goal of the infant is to achieve a "match" between features of the visually perceived target and features of the infant's own bodily state. In the first two months of life, the features in question include: (a) the configuration of body parts achieved by an action and (b) some of the actions' dynamic properties such as "speed, duration, and manner" (Meltzoff and Moore 1997, p. 189). In our terminology, the goal is to achieve a "match" between features of the present visual representation and features of a present action representation.

---

[6] Cf. Meltzoff (2002, p. 24; 2005, p. 72) and Meltzoff and Moore (1997, p. 180 and 182; 1999, p. 52).

Whether the goal is achieved or not is established by a computation requiring two inputs. One input is the targeted features of the visual representation (e.g. the configuration of the modeler's tongue and lips). The other input is the corresponding features of the action representation (e.g. the configuration of tongue and lips currently achieved by the infant). The computation compares the two inputs: if both inputs present the same information (e.g. they both present "tongue-beyond-lips"), the output is a "match." In this case, the infant has detected the similarity between the actions of self and other. If the information presented by the two inputs differs, the output is a "mismatch." In this case, the system specifies a new bodily state that the infant has to achieve in a new imitative attempt. The process of correction of imitative attempts continues until the computation gives a "match" result. Since the goal-directedness of the "active comparison" is the crux of the AIM hypothesis, the plausibility of AIM depends to a large extent on whether there is evidence for such a process of response correction.

In Meltzoff and Moore's (1997) schematic of AIM (p. 186) the two inputs entering the computation are kept clearly separate. The information coming from the visual system stands to the left of the "equivalence detector;" the information coming from the proprioceptive system stands to the right. We have replicated this structural characteristic of the model in our schematic of AIM in Fig. 2. In various passages, Meltzoff and Moore (1997) emphasize that the visual representations of the modeled actions must be "independent" or "separate" from the corresponding proprioceptive representations: the "representation of the other's body is separate from [the] representation of the infant's body" (p. 188; cf. 185). The reason for this emphasis is that without such a separation the idea of a computational process that guides the correction of the imitative attempts would not make sense. If there are not two distinct inputs entering the comparison, the hypothesis of a comparison for similarity detection falls apart.

Meltzoff and Moore (1997) also posit that there are imitative responses that occur without a process of correction, "on first try." How can infants know, without trials, what gesture in their motor repertoire matches the one they see? Meltzoff and Moore (1997) claim infants can know this by "looking up" or "reading out" the directory (p. 185). Admittedly, Meltzoff and Moore do not say much about this "looking up" process, so we must interpret them in a way that is coherent with the rest of the model. Given the passages cited above, one thing is sure: on one side, the infant has a visual representation specifying the bodily state to be matched; on the other side, it has a "directory" of stored action representations that connect movements to specific bodily states. Thus, the "looking up" process seems to be a process of searching out in the directory the entry that presents the same bodily state specified by the currently active visual representation. In order to give a coherent interpretation of Meltzoff and Moore, it is fair to think that the mechanism that identifies the correct "entry" for imitation on first try is the same mechanism that identifies a "match" result in a series of imitative attempts. In cases of imitation on first try, the equivalence detector would consult the directory until it finds the entry that matches the perceived model; at that point, the correct entry can be activated.

In Meltzoff and Moore's (1997, p. 180 and p. 186) diagrams of AIM, perception and action are separate systems. Indeed, Meltzoff and Moore assume that the comparability of perception and action required the evolution of a mediatory "supramodal representational system." Therefore, in Meltzoff and Moore's diagrams, this system stands in between the perception and action systems and appears to be sensitive to selected information deriving from those systems. The "special neural-cognitive machinery"
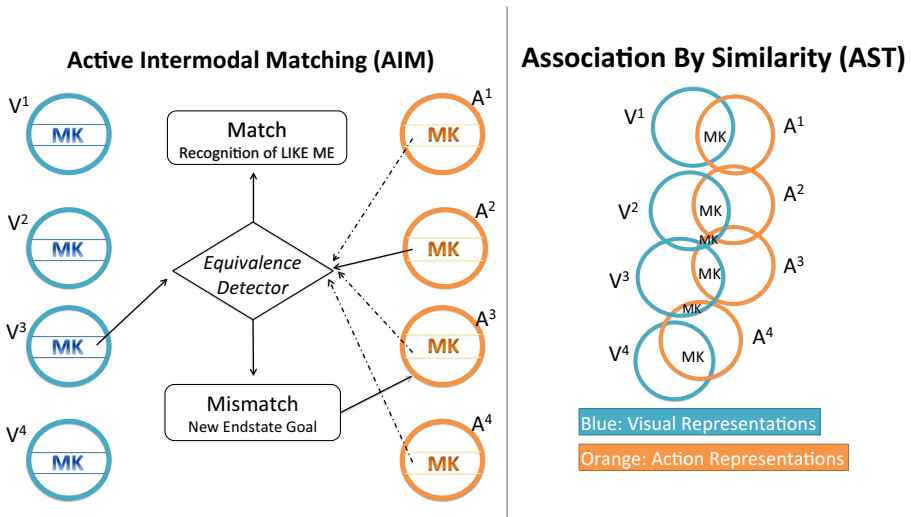
The contrast between AIM and AST

Fig. 2 The contrast between AIM and AST. AIM postulates a comparison between visual and action representations. If the comparison gives a "mismatch" output (e.g. A2 does not match with V3, judged by the Equivalence Detector), a new action representation is activated for a novel imitative attempt. If the comparison gives the "match" output, the infant has recognized a self-other similarity. In AST, there is no such comparison. Each visual representation overlaps the most with the corresponding action representation (i.e. V1 overlaps the most with A1 and so on); thus, when a visual representation is activated, the corresponding action representation tends to be activated too. The action's "morphokinesis" (MK) designates the set of action features experienced both in visual perception and action execution; the set includes the action's kinetic features and the peculiar configurations of body parts achieved through the action. In AIM, the peculiar MK of an action constitutes two distinct inputs that enter the comparison computation. In AST it simply indicates the information that visual and corresponding action representations share in common. (From Vincini et al. 2017)

(Meltzoff 2002, p. 9) implemented by the supramodal system evolved through "Darwinian means" (Meltzoff 2005, p. 55) for social and socio-cognitive functions. Thus, AIM is a nativist model in the same sense GPDM is (section 3). For AIM, "nature designed a baby with an imitative brain" (Meltzoff 2005, p. 77) and with an "innate propensity to imitate" (Meltzoff 2002) in order to ground its socio-cognitive development (Meltzoff 2007a, 2013), in particular its theory of mind abilities (Meltzoff and Decety 2003) and the capacity of testing other people's identity (Meltzoff and Moore 1992). It has been added that the NI module is inherited to promote parental attachment and later language development (Heimann 2002; Strid et al. 2006).

AIM is not a separate coding approach because it makes perception and action commensurable. However, it should neither be considered as a common coding approach, but rather as a "hybrid" (cf. Prinz 2012, p. 66). This can be evinced in at least two aspects of the model. First, in contrary to the common coding approach (Prinz 1990, 1997), in AIM perception and action do not have an intrinsic relationship due to the fact that they can represent identical characteristics of the spatiotemporal world (instantiated by something external and one's own body respectively). Rather, as we just noted, AIM postulates that evolution had to engineer a third supramodal system in addition to perception and action precisely to overcome the fundamental

incommensurability between them. Second, while a consistent common coding approach insists that "*identical* representational structures are involved in the perceiving and the performing of actions" (Prinz 2002, p. 153; our emphasis), AIM stresses the separation and independence of the representations of seen and executed actions. As Meltzoff and Moore's (1997, p. 186) schematic shows, these different kinds of representations must stand to the left and the right of the comparison process and do not constitute a representational overlap.

4.2 What are Characteristic Predictions of AIM and are they Confirmed by the Data?

In addition to predicting the existence of a process of response correction and organ identification, there are characteristic predictions that can be derived from AIM with respect to what the current state of the empirical literature should be, the range of imitated actions, and the settings of imitation. Moreover, the core assumptions of AIM do not point to a differentiation between the actions infants imitate. We examine whether the data support AIM in these respects.

4.2.1 Response Correction and Organ Identification

First of all, is there evidence for "the crux of the AIM hypothesis"? The presumed evidence Meltzoff and Moore (1997) refer to is their 1994 study and some previous non-systematic observation (Meltzoff and Moore 1983). Meltzoff and Moore (1994) found that infant production progressively matched the tongue-protrusion-to-the-side model over successive trials. However, as many critics have emphasized (Anisfeld 2005; Jones 2009; Ray and Heyes 2011), there is little in the findings that points to an interpretation in terms of response correction. We cannot expect an immediate full-fledged tongue-protrusion-to-the-side from infants, especially because this action is relatively infrequent in spontaneous behavior. At first, an infant may have an impulse to move in a particular way, but this impulse may be somewhat vague. It is precisely partial responses that provide the momentum for a complete response. So progressive match between model and infant production occurs not because infants gradually correct their responses, but because the vigor and amplitude of their responses increase.

This more parsimonious interpretation is supported by two considerations. First, Anisfeld et al. (2001) found that response rates increased over the course of their experiment independently of other factors and this is consistent with the Piagetian idea that action production calls for, or encourages, its repetition. Second, as Anisfeld (2005) and Ray and Heyes (2011) noted, there is an aspect of Meltzoff and Moore's (1994) own analysis of the sequence of responses that suggests increase in vigor rather than increase in the fidelity of the matching. Meltzoff and Moore assumed that large midline tongue protrusion is a closer match to the model than "small non-midline tongue protrusion" and "small tongue protrusion to the side," but this assumption is questionable. Yet it is clear that large midline tongue protrusion is more vigorous than the other two, so if it tends to appear later in the experiment, it is probably because it is more vigorous.

Furthermore, given the immaturity of the visual system, we cannot expect infants to immediately perceive the features of the models relevant to imitation (Jones 2009). Infant may need time to acquire a distinct perception of the modeled action

characteristics. Therefore, if infants match the model more in later phases of the experiment, it can be in part because they perceive the relevant action characteristics relatively late. Overall, then, we agree with Ray & Heyes (2011, p. 96) that a progressive match between action production and a modeled gesture could be interpreted as "increase in vigour [and amplitude] with response repetition, or […] perceptual learning—[…] the formation of a better perceptual representation of the modelled movement with repeated exposures."

For the same reasons, we doubt that there is any clear evidence for organ identification as a cognitive step prior to imitation. Stimulus presentation is unlikely to leave the infants' body unaffected. So infants' first reaction may just be a bodily repercussion of model presentation, or, alternatively, may constitute the initial preparation of the matching response, i.e. the mere energization of the body part that is to execute the response (cf. Heimann 2000 for a discussion of how slow infants are in producing a full-fledged response).

4.2.2 Overall State of the Current Evidence for NI

If infants have an "innate propensity to imitate," if NI is foundational for crucial aspects of social cognition, if evolution provided the baby with an "imitative brain" precisely to fulfill socio-cognitive and social functions—in short, if NI has the meaning AIM attributes to it—then one would expect NI to be a substantial phenomenon, i.e. one would expect NI to occur often and to be commonly detectable. One would expect that, after about forty years of numerous experimental efforts (Oostenbroek et al. 2013), there would be solid evidence for the existence of NI. As it is well known, this is not the case. Results of NI studies are often negative or ambiguous. This has led influential reviewers of the empirical literature to deny or call into question the existence of NI and, consequently, of all the cognitive operations postulated by AIM (Anisfeld 2005; Jones 2009; Lodder et al. 2014; Ray and Heyes 2011). Facing this skepticism, one could respond that the extant findings point to a nucleus of differential imitation whose existence will be confirmed by more appropriately designed future studies. In any case, the point we would like to suggest here is that AIM is difficult to reconcile with the relatively weak state of the evidence for NI. Indeed, AIM proponents do not seem to consider it something to be accounted for (Meltzoff and Moore 1997; Meltzoff 2013). A theory that denies a major role of NI in (socio-cognitive) development and interprets NI as a subtler phenomenon, as having a more episodic nature, accounts better for the limitations of the positive evidence. In later sections, we suggest that AST is such a theory.

4.2.3 Range

Meltzoff and Moore (1997) emphasize that their model relies on the empirical claim that infants imitate a wide range of gestures—they provide a list of seven or eight different kinds of imitative responses. Indeed, the idea of a comparison seems particularly appropriate if there are many gestures that potentially count as "matches" in response to their respective models. A sophisticated mechanism of comparison is needed to discriminate the correct matching response from the large set of responses that do not match the current model. However, the claim that a wide range of gestures are imitated is undermined by the reviews cited above, which suggest that the evidence

for the imitation of many gestures discussed by Meltzoff and Moore is unreliable. Hence it seems more prudent to hypothesize that if differential imitation exists, it is circumscribed to a few gestures (perhaps two or three).

4.2.4 Imitation Settings

What AIM predicts with respect to the settings of imitation is well expressed by Oostenbroek et al. (2013, p. 337), who also evaluate the evidence in favor of AIM's prediction: "There is evidence […] to suggest that newborns will imitate within these extremely controlled [laboratory] settings (e.g. Meltzoff and Moore 1977; Nagy et al. 2007), but not at home observations (e.g. Heimann and Schaller 1985). This really calls into question the underlying purpose of newborn imitation, because if the function of imitation in the newborn period is to facilitate social interaction as Meltzoff and Moore (1977) posit, then one would expect infants to imitate their mothers in a natural setting, rather than imitating an unfamiliar experimenter in a highly controlled and intimidating setting." Furthermore, Oostenbroek et al. (2016) tested AIM's prediction that imitation should occur at the infants' home and found a negative result.[7]

4.2.5 Differentiation Between Gestures

Unless auxiliary assumptions are added to the model, AIM seems to suggest there should be no significant differentiation between gestures in the frequencies of imitative responses. In fact, if the goal of the infant is to match features of the modeled action, there is no reason why infants should seek to match some actions more than others. However, considering the extant empirical literature, two gestures from Meltzoff and Moore's (1997) list of imitative responses, i.e. tongue protrusion and mouth opening, are imitated more often than the other gestures in the list, and tongue protrusion is imitated more than mouth opening (Coulon et al. 2013; Heimann et al. 1989; Meltzoff and Moore 1992; Nagy et al. 2013; Ullstadius 1998). AIM cannot explain this fact without resorting to auxiliary assumptions.

4.2.6 Recap

Our evaluation of the evidence relative to the characteristic predictions of AIM is consistent with the one proposed by Ray and Heyes (2011). None of the predictions peculiar to AIM is clearly supported by the findings. Therefore, we tend to side with Oostenbroek et al. (2016, p. 3) in claiming that AIM is "not empirically supported and should be modified or abandoned altogether."

5 The Association by Similarity Hypothesis (AST)

Basic idea of AST is that differential imitation can be explained through a domain-general process of association. In order to understand how AST is an alternative to GPDM and AIM, it is necessary to appreciate the scope and the fundamental character

---

[7] See Kugiumutzakis (1999) for an exception.

of association by similarity. Thus, before presenting AST in subsection 5.2, in subsection 5.1 we define association by similarity in general terms, recall some examples of its functioning, and specify how ideomotor theory resorts to this associative process to describe the phenomenon of action modulation through perception.

5.1 The Domain-General Process of Association by Similarity and Ideomotor Theory

According to a traditional classification from British associationism, there are three principles of association: similarity, contiguity, and cause and effect (Hume 2000). Modern scientific psychology has appropriated these principles. Classical conditioning is the heir of the principle of contiguity (the bell sound is contiguous to the food, so it becomes associated with it), whereas operant conditioning can be seen as an application of cause and effect (behavior is associated with its positive or negative effect). Although one could argue that it has been unjustly neglected (Allen 2012), similarity has also been recognized as a fundamental psychological process (Shepard 1987; Vigo and Allen 2009) and has been studied in sophisticated ways (Nosofsky 1992; Tversky 1988; Vigo 2009).

Association by similarity has been called the "factotum" of cognition because it plays a central role in a number of psychological phenomena such as stimulus generalization, categorization, recognition, memory retrieval, gestalt organization, analogical and inductive reasoning, problem solving and decision (Larkey and Markman 2005). Furthermore, considering that practically any organism capable of learning must be able to determine its behavior in the face of a new situation on the basis of the experience of similar situations in the past, it is reasonable to suppose that association by similarity must be functioning from a very early stage of evolution (Shepard 1987).[8]

It is possible to define association by similarity on two levels—the phenomenological level of the regulation of lived experience and the cognitive level of information processing. From a phenomenological perspective, association by similarity designates the associative process by which a present experience tends to activate content of past similar experience(s). For example, as we get out of a building and there is a car parked on the street, we see a car at first glance. Although only a side of the car is actually given from where we stay, we see a "car," i.e. an object with four sides, a specific practical meaning, and a specific set of features we could experience in the future. The present visual appearance (the side actually given) activates content (the general meaning of a car) from past similar experience (experiences in which an object with a similar side was given). We do not mistake what is given for a dog or a tree because it shares characteristic components with the experience of cars, not with the experience of dogs or trees (Husserl 1999; Merleau-Ponty 1964a, b, 2012).[9]

From a cognitivist perspective, association by similarity indicates the process by which the activation of bits of information tends to activate wholes in which they are normally integrated. For instance, in Hebbian learning models of perception, bits of information that are activated together become associated to constitute a complex object representation. Thus, a novel activation of an information bit due to sensory input facilitates the activation

---

[8] "Because any object or situation experienced by an individual is unlikely to recur in exactly the same form and context, psychology's first general law should, I suggest, be a law of generalization" (Shepard 1987, p. 1317).

[9] The functioning of association by similarity can be highlighted in a classic example of Piagetian assimilation, the child who, looking at a zebra, says "that's a horse."

of the associated bits that complete the representation of the object, making object recognition possible (Mongillo 2012). Indeed, a great number of theories in philosophy and cognitive science acknowledge that present sensory stimuli are apprehended in light of past perceptual experiences that presented similarities with the current stimuli (Barsalou 2008; Clark 2013; Meyer and Damasio 2009; Vetter and Newen 2014).[10]

Importantly, similarity is context-dependent. The features of the stimulus that initiate the association are those that are relevant to the practical experience of the subject (Decock and Douven 2011). In the case of perceiving a car, seeing an elongated shape and two wheels is a practically relevant experience associated with our global experience of cars. Combining phenomenology and connectionism, Dreyfus and Dreyfus (1999) provided a description of how association by similarity is involved in contexts of habitual actions. If a subject is used to responding to specific stimuli in a specific way, when it encounters a new similar stimulus (a stimulus that shares some characteristic features with those specific stimuli), a specific action response is facilitated.

The reference to phenomenological analyses and Hebbian learning models allow us to stress an important point about association by similarity. This basic process of association—as it can account for perception and action-related phenomena—should not be conflated with the recognition of similarity between compared objects or events. Recognition of similarity is the product of a comparison that finds a common feature between two (or more) distinct *relata*. The content of such recognition has the form: "This object is like that object (or those objects)." Thus, recognition of similarity is a psychological process that targets precisely the relation of similarity between objects; it has this relation (a is like b) as its own content or "object." One could claim that this kind of recognition is involved in perception. The idea would be that, in order to perceive a newly encountered object as a car, I have to recall past objects, compare the present object with them, recognize that "this objects is like those objects," acknowledge that those objects are cars, and infer that this object is a car. One problem with this account is how the subject can gather the objects for the comparison; it may have to recall a great (indefinite) number of objects experienced in the past until it finds objects with some relevant similarities. A second problem is that it is not clear that in order to perceive a car one really needs a recognition of the form "this object is like that (or those objects)."

Phenomenologists (e.g. Husserl 1999; Merleau-Ponty 1964a, b, 2012) would insist that normally, when we come across a car, we do not have to recall objects experienced in the past, nor do we have to compare the present object with past objects; rather, we just perceive the newly encountered object as a car. Analogously, in a Hebbian learning model of perception (Mongillo 2012), there is no computation that detects similarities and dissimilarities between the present stimulus and past objects. The mere activation of information that has been processed in the past—this *is* the similarity—activates a whole of associated information, i.e. a complex object representation. Thus, the content of my perception is simply: "a car."[11]

---

[10] Perhaps James provided us with the model of any neural formulation of association by similarity when he hypothesized that the evocation of an object occurs when a stimulus is "due to a brain-process some of whose elements awaken through habit some of the elements of the brain-process of the object which comes to view" (1981, p. 556).

[11] Obviously, it is possible to compare the present object with another object. For example, after I see a car parked on the street, when I walk pass by it I can think, "This car is like my friend's car because it has just two doors." However, this kind of comparison is not necessary to see the car as a car in the first place.

In short, for both phenomenological and Hebbian models of perception, association by similarity is not the recognition of a particular relationship between objects. Rather, it is merely the process that regulates the activation of meanings that constitute the content of perception. In other words, in order to perceive a stimulus as an object of a particular kind, one does not have to compare (an indefinite number of) objects; instead, it suffices to activate the complex representation that is most strongly associated with the features presented by the stimulus.

If association by similarity is understood as we have defined it above, ideomotor theory can be interpreted as a theory of the functioning of similarity in the domain of action modulation through perception. The ideomotor principle states that "every representation of action awakens in some degree the actual movement which is its object" (James cited by Prinz 2005, p. 143). This principle can account for action modulation through perception by relying on the hypothesis of common coding between perception and action (Prinz 2005). Indeed, if perception and action share identical resources to represent features that are common to both actions and perceptual stimuli, then the activation of those resources in perception will tend to activate the global action representations in which those resources are habitually integrated; consequently, the action representation awakened by perception will modulate action production. This account implies the functioning of association by similarity. It posits that the activation of representational resources in perception tends to activate wholes in which they are normally integrated, i.e. action representations.

Prinz (1997) reviews evidence deriving from induction and interference paradigms. In both paradigms, "the events presented as stimuli share certain properties with the actions to be performed in response to them. […] Induction paradigms study the way in which perceptual events induce, or enhance, actions by virtue of similarity. […] Interference paradigms study how perceptual events and actions interfere with each other by virtue of similarity" (Prinz 1997, p. 133). Prinz depicts similarity as a "graphic overlap" between perceptual and action representations. Because specific perceptions employ the same resources as specific actions, they can induce or interfere with those actions. Notably, the perceptual stimuli in the evidence reviewed by Prinz do not belong to the social domain.

Hence we can understand why, in ideomotor/common coding theory, imitation is just one of the ways in which perception modulates action. Imitation can be explained by the general notion that "the perception of an event that shares features with an event that one has learned accompanies or follows from one's own action will tend to induce that action" (Prinz 2005, p. 144). In 4.1, we observed that AIM is not compatible with a consistent common coding approach because it relies on the separation of the representations that must be compared. Here we can make explicit other features of the ideomotor/common coding theory of imitation that contradict assumptions made in AIM.

1) In Prinz's texts (Massen and Prinz 2009; Prinz 1990, 1997, 2002, 2005; Prinz et al. 2009), perceptions "induce," "modulate," "suggest," "facilitate," "awaken," "elicit," or "prime," corresponding actions. The perceptual system may even be said to "seduce" the action system. All these expressions denote the passivity by which action production is affected by perception. Thus, while the crux of the AIM hypothesis is "the active nature of the matching process" (Meltzoff and Moore 1997, p. 182), ideomotor theory proposes *passive similarity-based induction*.

2) For Prinz, there is no role for the recognition of similarity proposed by Meltzoff, which has the form "That seen event is like this felt event." The "functional role of similarity" (Prinz 2002, p. 160) is not confused with a comparison computation between distinct representations. Rather, association by similarity operates in that a perceptual representation directly activates the action representation with which it shares characteristic representational resources. For this reason, Prinz (2002, 160) states: "action imitation is […] a natural by-product of action perception."

3) Whereas AIM posits that the imitation module evolved for socio-cognitive or social functions, ideomotor theory suggests that "imitation is not based upon special purpose mechanisms, but, rather, relies on the general organization of learning and action control" (Massen and Prinz 2009, p. 2357).

Returning to the examination of young infants, we recall that the habituation procedure, a methodology often used by experimentalists in developmental studies (Sommerville et al. 2005; Van Heteren et al. 2000), is based precisely on the association by similarity between present and past (innocuous) experiences. Moreover, that association by similarity is operative in newborns is clear from studies of perceptual discrimination. Neonates can discriminate experiences had before birth in the domains of audition, taste, and smell (Hepper 2015). The case of the mother's voice is particularly suggestive because, although somewhat distorted in the intrauterine environment, the mother's voice before birth presents enough similarities to the voice heard after birth such that the latter can be discriminated. These kinds of studies show that similarity is in place in newborns for reasons other than vision-action translation. Even Meltzoff and Moore (1997, p. 181) assume that, when a newborn recognizes a still face as a face that has produced a specific gesture 24 h before, it is coding a visual stimulus F^1 in terms of a past similar visual stimulus F.

5.2 Differential Induction of Spontaneous Behavior through Similarity

AST can be considered as an application of Prinz's (1990, 1997, 2005) ideomotor theory and constitutes a consistent common coding approach to infant imitation from 0 to 2 months. It posits that NI consists in differential induction of spontaneous behavior through the similarity that each visual model entertains with the corresponding action experience. AST stresses that the actions that infants imitate are habitual and spontaneous actions of the infants' repertoire. Model presentation tends to awaken the corresponding action representation; in this way, it increases the probability that the corresponding action is executed (or reinforces aspects of such an execution, such as the duration of a more pronounced MO).

Figure 2 shows the way AST describes the functioning of association by similarity. The key point is that visual processes relative to model presentation overlap with specific action processes; these areas of overlap are the areas that track the contents experienced both in visual perception and proprioception. In the previous subsection we noted that similarity is context-dependent; in the particular case of NI, the features of the stimulus that initiate the association are those that are habitually instantiated in a spontaneous action of the infant.

For AST, infant imitation presupposes spontaneous, habitual action execution (but see footnote 14 for an important qualification). Thanks to proprioception, spontaneous habitual execution constitutes a learning process. Infants learn what movements instantiate specific morphokinetic features experienced proprioceptively. This learning process coincides with

the acquisition of global action representations: motor codes originating a specific action become associated with the proprioceptive experience of characteristic morphokinetic features of the action. In other words, in action execution specific motor codes wire together with specific proprioceptive codes so as to form global action representations. This is the kind of learning that AST requires; it does not require infants to learn associations via contiguity between visually perceived and executed actions.

AST hypothesizes that, when the infant sees the modeled action, the representational resources used in vision to represent the morphokinetic features of the action are the same resources that have been used to represent those morphokinetic features in proprioception. In other words, AST postulates a representational overlap. In this way, AST posits that the visual representation of an action involves representational resources that have been wired up with the motor components of a global action representation in spontaneous execution. Because of this prior association, the activation of the overlap area in visual processing will tend to activate the other areas with which it was habitually linked in action processing. In this way, a habitual action possibility is reawakened, and, if the infant does not have stronger impulses that lead it to behave otherwise, it will adhere to this action possibility, i.e. it will execute the act that we designate as "imitative." There is no comparison and no recognition that "That seen event is like this felt event." There is simply first a perception and then an impulse to act in a certain way; association by similarity regulates which perception activates which action tendency.

We will present other characteristic elements of AST in section 6 as we highlight the advantages of AST over AIM.

5.3 Two Elements AIM and AST have in Common

Before we explore the contrast between AIM and AST, in this subsection we need to identify two assumptions these models have in common.

First, both AIM and AST entail that NI relies on the existence of a body schema in the infant—Meltzoff and Moore (1997) describe it as the "movement-end state directory." According to both AIM and AST, the body schema starts developing prenatally; therefore it can support imitation right after birth. The idea of a body schema acquisition through spontaneous prenatal motility is confirmed by a number of authors in philosophy and science (Gallagher 2005; Hepper 2015; Piontelli 2015; Sheets-Johnstone 2011; Van Heteren et al. 2000). Specifically, the examination of prenatal behavior reveals that all the actions that newborns (are claimed to) imitate after birth have already been regularly executed before birth. For each neonatal imitative response, Table 1 provides at least two studies that prove the existence of the corresponding prenatal motor habit. These studies show that the frequency of the actions in question before birth (in particular during the third trimester of pregnancy) is comparable to their frequency after birth.[12]

Second, AIM and AST posit that there are features of an action that are experienced both in the visual perception of the model and in the proprioceptive experience of the corresponding action execution. In AIM, this assumption is taken to mean that there are two distinct inputs of a comparison computation that present the same information (the same information is represented twice). In AST, the same assumption means that visual and action representations share information in common (the common information is represented only once).

---

[12] Keven and Akins 2016, sect. 7, para. 2

Table 1 Prenatal motor habits corresponding to actions imitated after birth

Imitative responses after birth	Studies attesting corresponding motor habits before birth
Mouth opening	Roodenburg et al. 1991, D'Elia et al. 2001
Hand movements	Katz et al. 2007, Kurjak et al. 2008
Head movements	Roodenburg et al. 1991, Andonotopo and Kurjak 2006
Lip movements	Hata et al. 2005, Reissland et al. 2011, Reissland et al. 2012
Eye blinking	Kurjak et al. 2004, Yigiter and Kavak 2006
Emotional expressions	Kurjak et al. 2003, Kurjak et al. 2004
Tongue protrusion	D'Elia et al. 2001, Kurjak et al. 2004

For each neonatal imitative response, the table provides at least two studies that document the existence of the corresponding prenatal motor habit (e.g. for mouth opening imitation, two studies attesting to prenatal spontaneus mouth opening). The studies show that the frequency of the actions in question before birth (in particular during the third trimester of pregnancy) is comparable to their frequency after birth

Nonetheless, both AIM and AST postulate that each action is characterized by a peculiar set of features and this set is experienced both in visual perception and action execution.

Recall that for Meltzoff and Moore (1997) each entry in the "directory" is characterized by the final configuration of body parts it achieves. They also claim that, in the first two months of postnatal development, infants match the "speed, duration, and manner" of the other's actions (Meltzoff and Moore 1997, p. 189). Meltzoff (2013) leaves aside the emphasis on spatial configurations and stresses movement patterns, i.e. "kinetic signatures," as the content that is experienced in both visual perception and action execution. We agree with this shift in emphasis since, at a basic level, a human subject is more a moving organism than a passive observer of spatial relations (Gallagher 2005; Sheets-Johnstone 2011). However, we also maintain the reference to spatial configurations because we believe that the description of the features experienced in both visual and proprioceptive modalities must be as inclusive as possible. We propose the expression "morphokinetic features" to indicate the set of features that each action presents in both modalities.[13]

To recap, AIM and AST rely on the existence of the body schema and postulate that there are action features experienced both in visual perception and action execution. If one rejects AST because of one of these two assumptions, he or she has to reject AIM for the same reason and vice versa.

6 Advantages of AST over AIM

6.1 AST is more Parsimonious

The main theoretical difference between AST and AIM becomes apparent at first glance by taking a look at Fig. 2. In AIM there is a comparison between visual and action representation that eventually lead to a "recognition experience" (Meltzoff 2002,

---

[13] The tactile mode can also contribute to the infant's registration of the relative positions of organs in the body and of the usual configurations they assume. For example, the fetus can touch its face, mouth, nose and eyes and experience the disposition of these organs (Kurjak et al. 2004); such a "tactile image" may be integrated in the experience of spatial positions acquired through proprioception.

2007a; Meltzoff and Decety 2003). Precisely because this recognition experience is the product of a comparison that finds a common feature between two distinct *relata*, Meltzoff proposes that it has a content of the form "a is like b:" "That seen event is like this felt event" or "Here is something like me" (Meltzoff 2002, 2007a, 2005, 2010; Meltzoff and Decety 2003). In contrast, in AST there is no comparison between visual and action representations, and, therefore, there is no recognition of similarity having the content Meltzoff proposes. Accordingly, in AST there is no need to postulate that a specialized module for the comparison of perceived and executed actions was selected through evolution. Rather than a comparison, a visual representation tends to activate the corresponding action representation because it includes elements that are habitually associated with the other components of that action representation. Rather than a specialized module, a mechanism that is available to the newborn for more fundamental functions (e.g. stimulus generalization) takes the role of mediating a new kind of behavior (i.e. imitation) under a very specific environmental condition (the repeated presentation of modeled acts). That is: association by similarity ends up motivating newborn imitative responses even if that is not the function for which it evolved.

Consequently, AST offers a more parsimonious account of the progressive match between models and infant action production over the course of the experiment. Infants cannot be expected to give a full-fledged action response immediately, nor can they be expected to perceive the model distinctly from the beginning (e.g. initially infants may not perceive TP-to-the-side as distinct from midline-TP). Repetitive exposure to the model over trials increases the probability to induce the corresponding actions. Moreover, an induced action tendency may be vague at first or only imperfectly realized; it is through partial responses that infants may acquire momentum (in line with the Piagetian idea that action calls for its repetition) and increase the vigor and amplitude of their responses. This explanation is consistent with what various authors observed about the limitations of the capabilities of young infants, in particular the slowness and graduality of their responses (Anisfeld 2005; Heimann 2002; Jones 2009; Ray and Heyes 2011).[14]

AIM posits a recognition experience in which self-executed action is on equal footing with other-executed action (two compared inputs). On the contrary, for AST the baby simply experiences a visual perception, then, after the appropriate amount of time and action preparation, it releases a specific action. This interpretation is more parsimonious and fits well with observations that model perception "absorbs" the infant and inhibits spontaneous movements (Anisfeld 1991); then, as the modeled action stops, action execution is released in part according to the action tendencies evoked by the model.

Postulating a comparison computation between visual and action representation creates a further problem with AIM, which also seeks to account for imitation "on first

---

[14] In footnote 2, we noted that Meltzoff and Moore (1994) found that infants produce TP-to-the-side spontaneously. Keven and Akins (2016) suggest that TP-to-the-side belongs to the action repertoire of the fetus eight weeks before birth. Nonetheless, we should not exclude a priori that an infant might imitate TP-to-the-side before having executed it spontaneously. AST would explain this circumstance as follows. It may be that the perceived action presents features that are experienced to different degrees in different actions already accomplished by the subject, but none of these actions instantiated all of them. This may lead to the activation of a bodily part that is more strongly associated with some features presented by the model than the bodily part involved in the corresponding action (the laterality of TP-to-the-side activates head lateral movement—cf. Meltzoff and Moore 1997, p. 182). However, with time, the fact that action features are awakened that do not correspond to any particular spontaneous action may lead to a new configuration in action planning (a new Gestalt: TP-to-the-side), so that a new action is extrapolated and induced starting from an action repertoire.

try" (5.2). The problem is how the infant selects the action in its repertoire that matches the target. Because, according to AIM, the only means to detect "matches" is a comparison, the equivalence detector may have to examine all action representations in the infant repertoire before it finds the matching one. A costly task![15] This problem, however, disappears in AST. Differential imitation is explained by the mere supposition that each visual representation shares most information, or most characteristic information, with the corresponding action representation (Fig. 2). Thus, each visual representation tends to activate the corresponding action representation as a direct consequence of the overlap between them.

Unlike AIM, AST does not posit that the infant intends to match the behavior of others (the goal-directed character of imitation) and wants to test other people's identity (the motivation for imitation). How, then, is the infant motivated to produce the matching response? In line with the ideomotor approach, AST states that the response is "induced" or "suggested" by the presentation of the model. The modeled act evokes a motor habit that can be implemented. *The mere evocation of an action possibility is a motivation, or "enticer," to fulfill it*, when stronger motivations are not conditioning the newborn otherwise. In other words, once an action possibility has been awakened, this being-awakened makes that action more prominent in the range of action possibilities that constitute the background; thus, *other things being equal* (i.e. if stronger, unpredictable impulses do not favor other responses over the imitative response), the infant will be more likely to enact the action possibility that has come to stand out.

No costly comparison, no recognition, no specialized module, no intermediary step of identification, no intention to match or test other people's identity, just a basic process of association and the resulting solicitation. Even if AIM could be as empirically accurate as AST is, AST would still be preferable for reasons of parsimony.[16]

6.2 AST Fits the Extant Empirical Findings

6.2.1 Overall State of the Current Evidence for NI

Meltzoff (2010, p. 16) claims that imitation is a behavior characteristic of "typical newborns." Indeed, in AIM there are two reasons for which one should expect NI to occur often and in a relatively large variety of circumstances. First, infants have an innate propensity to imitate and this propensity fulfills indispensable socio-cognitive or social functions. Second, infants have the goal of matching what others do and this intention is a relatively strong factor in determining infant behavior. Obviously, AIM

---

[15] We thank an anonymous reviewer for raising this point.

[16] As a consequence of being more parsimonious from a cognitive-psychological point of view, AST is also more biologically plausible. Both AST and AIM require that brain processes encoding a specific visual model activate brain processes encoding the corresponding action. Otherwise imitation could not occur. In addition to this, AIM postulates an "equivalence detector" that produces "match" or "mismatch" outputs. Where is the equivalence detector, where are the comparison outputs located in the brain? These postulates look particularly cumbersome from a biological point of view. Moreover, AIM suggests that the information characterizing each action is encoded twice, i.e. as input from vision *and* as input from proprioception. In contrast, AST merely entails that visual processes are associated with corresponding motor process so that the former can activate the latter. Hence the action-characterizing information can be encoded only once, in the areas where the two kinds of processes overlap. For these reasons, AST can be more easily substantiated by neuroscientific research than AIM.

accepts that if the baby is uncomfortable, sleepy, hungry, interested in something else, etc., it will not imitate. However, even when the baby is judged to be in "quiet alert state," there are always uncontrollable variations of those affects that condition the infant, as attested by the fact that NI sessions are interrupted innumerable times for the emerging unavailability of the infant at various points of the experiment (e.g. Anisfeld et al. 2001; Oostenbroek et al. 2016). Given Meltzoff's emphasis on NI as a "typical," widespread behavior, one is lead to think that the imitative intention has a relatively strong capacity to compete with moderate antagonistic affects. For example, when busy with the task of matching the actions of others, infants can keep other moderate impulses or action tendencies at bay, at least to some extent and until they achieve the goal. As argued in 4.2, assumptions like these two make AIM difficult to reconcile with the fact that after about forty years of experimental efforts, the evidence for the existence of differential imitation is still weak and ambiguous.

Conversely, in AST one does not expect infants to imitate often and in a variety of circumstances for the two opposite reasons. First, newborns have no innate propensity to imitate. NI has more to do with the scientific interest of the experimenter of testing aspects of newborns' visuomotor ability, than with the baby's fulfilling indispensable socio-cognitive or social functions. NI is the increase in the frequency of some gestures that can only be detected if researchers compare that frequency across different response conditions.[17] Second, in AST infants do not have the intention to match what others do. Infants merely follow their own action tendencies and model presentation can only promote one tendency over others. Precisely because there is no intention to imitate, slight differences in infants' affective states can become more preponderant and condition infants more and in unpredictable ways. Indeed, the reawakening of an action possibility through association by similarity is a weaker motivation than the intention to match others' actions. Thus, for AST it may well be that infants often do not react to modeled acts or react in unpredictable ways even when they are judged to be in quiet alert state.[18]

In short, because AST makes it intelligible why infants often do not imitate (lack of developmentally-crucial propensity and intention to imitate), AST is not undermined by the current limitations of the positive evidence. In opposition to experimental practices inspired by AIM, AST indicates specific conditions in which differential induction can be maximized and more easily detected. We briefly examine these conditions in 6.2.

---

[17] The idea that NI does not fulfill any decisive socio-cognitive or social function is supported by the fact that, contrary to expectations, Coulon et al. (2013) found that newborns did not imitate more when the modeled gesture was accompanied by the "congruent" auditory social stimulus (e.g. mouth opening with sound [a]). However, as anticipated in footnote 3, it is clear that AST does not exclude the existence of innate social propensities of other kinds. For example, it is compatible with Coulon et al.'s (2013) finding that newborns looked significantly longer at "audiovisual congruent" stimuli and with the hypothesis that newborns may have an innate preference for these kinds of stimuli.

[18] It is not so much that newborns are split into "imitators" or "non-imitators" (Simpson et al. 2014), although it is true that some infants may have greater motor and visual capacities and may have a greater inclination for social stimuli than others. Rather, there are many subtle factors that affect the behavior of infants and these factors may impede the imitative response. Along these lines, AST can elegantly accommodate Jones' (2009) observations concerning tongue protrusion as arousal response. In AST arousal can still function as a reinforcing motivation for tongue-protrusion execution, which would help explain the predominance of tongue-protrusion in NI studies.

6.2.2 Range

AIM relies on the claim that a large variety of gestures are imitated, but many reviewers question the tenability of this claim (4.2). If only two or three gestures are imitated, a sophisticated comparison mechanism seems unnecessary. On the contrary, AST would be supported even if empirical research proved that only two or three gestures are imitated. It is perfectly compatible with AST that only actions that are most habitual and most differentiated in proprioceptive experience can be induced by a visual model.

6.2.3 Imitation Settings

Unlike AIM, since it does not as ascribe any indispensable socio-cognitive function to NI, AST accounts for the fact that differential imitation is detectable in extremely controlled laboratory settings, but practically not detectable in natural settings (cf. 4.2).

6.2.4 Differentiation Between Gestures

A core prediction of AST is that there will be differentiation between imitated gestures: the more frequent the gesture in spontaneous execution, the easier it will be to induce it. This prediction is intrinsic to AST for at least two reasons. First, if a gesture is frequently produced in spontaneous behavior, it means there is a strong action tendency for that gesture. Thus, it will be easier for the model to bring that action tendency above threshold. Remember that in AST the infant has no innate intention to match the model. So an attempt at differential solicitation will tend to be more effective to the extent that it can take advantage of a spontaneous inclination toward executing the action in question.[19] Second, the more habitual the action, the stronger will be the connections constituting its action representation. In other words, in frequent action execution, the motor codes originating the action will have more opportunity to strengthen their association with the representation of characteristic morphokinetic features activated by proprioception. Then, when the representation of those morphokinetic features is activated by vision, this activation will more easily propagate to the motor components of the action representation in virtue of the stronger prior association. Hence the global action representation corresponding to the visual model will be more easily awakened.

AST's prediction is confirmed. In 4.2 we noted that tongue protrusion is the most imitated gesture and mouth opening is the second most imitated. Here we add that

---

[19] We take it that it is difficult to elicit in an infant from 0 to 8 weeks an action that it is very little disposed to execute. In this case, awakening the representations of certain morphokinetic events would constitute too weak of a motivation, almost negligible, compared to other tendencies that may condition the infant at the given moment. The activation of those representations is much more likely to be effective in bringing about the action if it is combined with a certain easiness of activation of the motor components of the action, i.e. if it is combined with a spontaneous action inclination. Note that by itself a spontaneous action inclination may often not be enough to provoke the relevant action because an infant may be conditioned by antagonistic inclinations, or by a variety of internal and external stimuli. In NI the activation of components of an action representation through vision will combine with the easiness of activation of the motor components of that action representation so as to modulate action execution. Specifically, the representation of specific morphokinetic features will more effectively activate the motor components with which it has been associated in spontaneous behavior because those motor components are already characterized by a certain easiness of activation.

tongue protrusion is the gesture that occurs most frequently in spontaneous behavior and mouth opening is the second most frequent (their frequencies of spontaneous execution are approximately 1.85 and 1.15/per minute respectively).[20] AIM cannot explain this correlation without resorting to auxiliary assumptions (4.2).

6.2.5 AST Fits the Operational Definition Better

As discussed in section 2, differential imitation does not change the relative frequencies of two gestures. For example, in MO imitation MO does not become more frequent than TP (Meltzoff and Moore 1977; Meltzoff and Moore 1994). Rather, there is an increase in frequency or duration of MO with respect to other control conditions (e.g. presentation of TP or of a passive face). AST seems to fit this operational definition. Indeed, AST consists in the hypothesis that a specific model presentation can facilitate or enhance a specific action experience because it shares more characteristic features with that action experience than the presentation of other models (Fig. 2).

In contrast, the operational definition seems to pose some challenges to AIM. If an infant imitating MO produces more TP than MO can we still claim that it has the goal to match MO? Moreover, consider again the evidence for MO imitation in Meltzoff and Moore (1994). Infants produce well over two TP for every MO they produce. In this situation, the equivalence detector would have to recognize that self-produced actions are much more frequently *unlike* other-produced actions than they are similar to them. Is an experience of this kind capable of grounding "infants' apprehension that the other is, in some primitive sense, 'like me'" (Meltzoff and Moore 1997, p. 185)?

6.2.6 AST Better Explains the "Drop Out"

Meltzoff and Moore (1992) explain the decrease in imitation after the second month by positing that infants become more interested in other forms of interaction at that point. This explanation has left many critics unsatisfied (Keven and Akins 2016). It is true that the end of the second month sanctions the beginning of a more active engagement with the social environment (Rochat and Striano 1999), but, if infants actively match the behavior of others and test others' identity through imitation, it is unclear why they should stop being interested in doing that after the second month. AST sheds more light on the effect that a transition to a more active interaction may have on imitation. According to AST, the imitative response is passively induced; therefore, when infants become more active in social interaction, they will be less disposed to let the "choice" of their behavior be determined by a passive stimulus. Rather, infants will behave more according to a self-determined stance at that point.

Furthermore, AST understands imitation to be the differential solicitation of actions that already tend to occur spontaneously. If the actions in question stop being

---

[20] The number for tongue protrusion was retrieved and averaged across the following studies: Heimann and Schaller 1985; Heimann et al. 1989; Meltzoff and Moore 1989; Nagy et al. 2013; Ullstadius 1998; that for mouth opening across the following: Coulon et al. 2013; Heimann and Schaller 1985; Heimann et al. 1989. Compare these frequencies with that of, for example, spontaneous lip movements, i.e. 0.14 per minute (Ekman and Rosenberg 1997). Frequencies for the other gestures can be found in other studies and are significantly lower than those for tongue protrusion and mouth opening (Meltzoff and Moore 1989; Kurjak et al. 2004; Oostenbroek et al. 2016).

spontaneously executed, it will be more difficult if not impossible to solicit them. This prediction is verified by the existence of a correlation between decrease in imitation and decrease in spontaneous execution (Ray and Heyes 2011; Keven and Akins 2016).

7 AST Gives New Impulse to Empirical Research

AST indicates specific conditions in which differential induction can be maximized and detected.

We discuss these conditions more extensively and propose an experimental design conforming to them in another paper (Vincini et al. 2017). Some of the procedures we propose have already been implemented to different extents (Heimann and Schaller 1985; Kugiumutzakis 1999; Meltzoff and Moore 1992; 1994). However, AIM does not allow discriminating these experimental procedures from opposite ones. Indeed, a consistent application of AIM seems to lead precisely to the lines of research that have been revealed to be dead ends by recent research. A symptomatic example is the study by Oostenbroek et al. (2016), which set out to test AIM's assumptions and concluded both that AIM was falsified and that NI did not exist. Here we consider this study in some detail to highlight how AST and AIM can lead to opposite experimental procedures.

7.1 Imitation Settings

We already noted that Oostenbroek et al. (2016) tested AIM's assumption that imitation is a social behavior of typical newborns in their domestic environments (4.2). However, from the point of view of AST, domestic environments are rich in potential distractors. Therefore, unlike AIM, AST specifies that sessions should occur in silent laboratory settings. Temperature and lighting should be adjustable, and the visual background should be a uniform soothing color in order to promote a calm affective state in the infant and guarantee as much as possible that the only variable that changes across sessions is the modeled gesture (internal validity). The modeler's face must be spotlighted and its luminance regulated in order to increase the probability that infants focus on the features of the stimulus that can awaken the corresponding action.

7.2 Sample Size

AIM's assumption that NI is foundational for social cognition in the typical newborn must be tested in studies involving a large number of infants. Otherwise, results cannot be generalized to the typical newborn. For this reasons, Oostenbroek et al. tested 106 infants. An experiment of this kind aims at "external validity" but tends to have poor "internal validity," i.e. little experimental control on determining which variables affect the outcome (Campbell and Stanley 1966; Kratochwill 1992; Kennedy 2005).[21] For example, examining so many infants makes it impractical for experimenters to focus on the optimal conditions for inducing specific actions. Indeed, Oostenbroek et al. say little

---

[21] More precisely, "internal validity" refers to the degree of certainty that manipulation of the independent variable is responsible for observed changes in the dependent variable.

or nothing about how experimenters captured the infant's optimal alert state and facilitated attention to the model. On the contrary, experimenters should monitor the infant beginning 10–15 min after feeding and wait until the optimal alert state is coming. Moreover, there should be a preliminary phase in which the infant acclimatizes to the test setting, the experimenter seeks its optimal posture and attracts attention to the modeler's face (e.g. producing sounds without opening the mouth). In a nutshell, AST recommends constrained sample size for maximal internal validity (around 30 infants—cf. Simpson et al. 2014).

7.3 Method of Analysis

Oostenbroek et al. averaged data across all infants and analyzed these averages. Again, this procedure is a test for AIM because if NI is a behavior of the typical newborn it should be detectable through measures of their typical behavior. In contrast, AST suggests that averaging data across infants tends to "iron out" genuine episodes of imitation. Each infant has its own action tendencies and habits, so its behavior should be analyzed separately. AST proposes to use each infant as its own control, i.e. compare responses to the target model with those to other models for each infant. This method offers the possibility of doing a statistical analysis of the proportion of infants who exhibit an increase in action production in response to the corresponding models (e.g. Meltzoff and Moore 1992).

7.4 Number and Variety of Models

If the infant has the goal to match what others do (as AIM posits), it will seek to produce a specific action when the corresponding model is presented, but will not have particular motivations to produce the action in response to other disparate stimuli. Thus, the increase of the corresponding action should be detectable in comparison to a relatively large variety of stimuli. Accordingly, Oostenbroek et al. presented infants with 11 models of different kinds: 4 facial, 2 non-social, 3 vocal, and 2 hand models. However, different kinds of stimuli may provoke different affective reactions, e.g. they may arouse infants to different extents. Thus, increasing the number and variety of stimuli can be a confounding factor. AST suggests using a limited number of models of the same kind, for example no more than 5 facial models: TP, MO, lip protrusion, head rotation, and passive face (baseline). If two models of this set induce the corresponding actions more then the other models, there is evidence for AST. That is, it is enough to support the idea that a specific model presentation can facilitate a specific action experience because it share more characteristic features with that action experience than other model presentations.

7.5 Number of Models in a Session

If infants have the goal to match what others do, it is possible that they will change their behavior to match the changes they see in the modeler's behavior. Accordingly, Oostenbroek et al. (2016) presented their 11 models in a row, for a total of 11 min including model presentation phases and response phases. For AST, however, induced responses can be rather slow. Indeed, Heimann (2000) states that it can take more than

60-s for the corresponding action to emerge. Thus, it is possible that, in Oostenbroek et al.'s design, imitative responses ended up occurring when subsequent models were presented and so counted against differential imitation. AST favors a different experimental design. In one session, only one model should be tested, possibly preceded by the measurement of baseline (responses to passive face), in order to avoid carryover effects. Sessions should be distributed across days or well separated in the same day. Presentation phases should be longer (for a total of 60-s presentation at least for each model) to increase the probability of inducing an action through repeated model presentation. Response phases should be comprehensive enough to detect slow responses (including at least 75-s after the last model presentation). Overall, in this kind of design, test sessions are shorter. This diminishes the problem of interrupting the experiment because of the baby's unavailability, which leads to having to start again with the baby in a somewhat altered state.

7.6 Differentiation Between Gestures at the General and Individual Level

AST encourages empirical inquiry into the differentiation between imitated gestures, a line of inquiry that has been neglected in Oostenbroek et al., but more broadly under AIM (4.1). AST predicts that, in general, gestures that are more frequently executed in spontaneous behavior are more easily induced (6.2). This correlation can also be investigated at the individual level. An infant who spontaneously produces a gesture at a particularly frequent rate (especially compared to other gestures of its own) is expected to produce more imitative responses for that gesture than for others.

7.7 Recap

Guided by assumptions peculiar to AIM, Oostenbroek et al. were led to an empirical study characterized by a domestic setting, large sample size, calculation of averages across infants, large number and variety of control models, more models per session, and no particular attention to the correlation between spontaneous behavior and matching responses.[22] In order to maximize and detect differential induction, AST leads into opposite directions: a controlled laboratory setting, constrained sample size, taking each infant as its own control, constrained number of comparison models, one modeled gesture per session, and inquiry into the correlation between spontaneous behavior and matching responses at the general and individual levels.

8 Note on Subsequent Imitation Development

The field of imitation development has to take different approaches into consideration (Anisfeld 2005; Froese and Leavens 2014; Jones 2009; Piaget 1962; Ray and Heyes 2011; Subiaul 2010). AST is an application of the ideomotor/common coding approach (Prinz 2005) to NI. After our discussion, it is easy to see how the ideomotor/common coding

---

[22] One could argue that there can be better implementations of AIM than Oostenbroek et al.' study. However, while this study can be considered as a legitimate implementation of AIM, it cannot be taken to be a legitimate implementation of AST.

approach could apply to Piaget's (1962) observations on infant imitation of adult vocalizations from the second month of life. Piaget emphasized that infants imitated vocalizations similar to those they had already produced and experienced auditorily. Hence one can suppose that hearing an adult-produced vocalization awakens infants' tendency to vocalize in a similar manner. This may occur in virtue of the overlap between perceptual and action representations (where action representations are constituted by the associations between movements of the vocal apparatus and their auditory effects). Nonetheless, it is beyond the scope of this paper to investigate how ideomotor/common coding theory accounts for imitation at different stages of development.

For our purposes, it is more opportune to make a few remarks about the stage of imitation that, in Meltzoff and Moore (1997, pp. 189–190), follows the stage characteristic of NI. This stage would occur at about 1 year of age and would be characterized by (a) a more abstract notion of the matching relationship and (b) sense-specific information enrichment. The example Meltzoff and Moore provide for (a) is the finding that if experimenters imitate 1-year-olds in everything they do when they play with toys, these infants change their behavior abruptly to test whether the experimenter is doing the same. In this regard, we emphasize that AST does not deny that an infant can recognize self-other similarities at later stages of development. Specifically, 1-year-olds have gone (or are going) through the so-called "9-month revolution" (Tomasello 1999). At this age, infants have entered the stage of "secondary intersubjectivity" in which they "monitor others in relation to objects" and exhibit "gestural communication, […] pointing, joint attention, gaze following, and social referencing" (Rochat and Striano 1999). Infants have learned a lot about what adults typically do with them, and, as Meltzoff and Moore observe, have developed their understanding of themselves. In other words, at this age, infants are significantly different from newborns. They have the cognitive abilities and a sufficient experience of self and others to be surprised or amused by an adult who faithfully copies their actions with toys; and so they can notice that, in that situation, "the other is doing the same as I do."

However, the example of 1-year-olds testing adult imitative behavior is not strictly relevant to our topic of the psychological mechanism of infant imitation because the imitative behavior in question belongs to the adult, not the infant. Similar problems concern the examples Meltzoff and Moore provide for (b). Nonetheless, it is instructive to consider one of them, namely "Infants at this age [1 year] also tactually compare the unseen parts of their bodies with those of adults, feeling the adult's mouth before reaching to their own" (p. 190). According to Meltzoff and Moore, here we have evidence for an "active comparison" that achieves a recognition of the form "that is like this." The evidence could consist in the fact that infants successively direct their hands to others' body parts and theirs, and perhaps linger over this tactile experience. Importantly, any evidence of this kind is absent in newborns. Since NI is not accompanied by any behavior of the kind that really attests an active comparison, it appears more parsimonious to assume that it is mediated by a simpler mechanism.

Overall, the only clear examples that Meltzoff and Moore provide for the recognition that "something about the other (actions or body parts) is like something about the self" seem to reinforce the supposition that this kind of cognition does not occur in newborns, but rather occurs in older infants who have developed their sense of themselves and others and a number of cognitive capacities that are more complex than those of newborns.

9 Conclusion

In this paper, we identified three possible explanations for early differential imitation: GPDM (section 3), AIM (section 4), and AST (section 5). We did not engage in a critical examination of GPDM. We presented GPDM only to stress its difference from both AIM and AST. GPDM rejects that differential imitation entails a mechanism for detecting the similarities between the acts of self and other. GPDM also denies the functional role of the domain-general process of association by similarity. Quite the contrary, it posits that vision-action connections are automatic specialized adaptations that were selected for socio-cognitive or social functions.

We focused on the contrast between the currently dominant model (AIM) and the alternative hypothesis we propose (AST). The latter is preferable to the former for two reasons. First, even if AIM could be as empirically accurate as AST, AST would still be preferable for reasons of parsimony and developmental plausibility (6.1). AST relies on association by similarity (5.1). It posits that NI is nothing else than differential induction of behaviors that already tend to occur spontaneously (5.2). Consequently, it does not have to endorse a number of assumptions about specialized modules, recognition acts, intentions to match and identify others, etc. Second, AST better accounts for the extant findings (6.2). Indeed, whereas the extant findings undermine AIM (4.2), AST can explain the limitations of the positive evidence for NI, the narrow range of and the differentiation between imitated gestures, the efficacy of laboratory settings as opposed to domestic ones, and the drop of imitation at 2–3 months. AST also better fits the operational definition of differential imitation, i.e. what is really measured in NI studies: a mere increase of specific actions in response to the corresponding models compared to other models, not a response that makes the corresponding action more frequent than other spontaneous actions. Furthermore, we suggested that AST can give new impulse to empirical research (7). AST clarifies why lines of research inspired by AIM are destined to remain unproductive and specifies conditions to enhance and detect differential induction.

We noted that skepticism toward the existence of NI is increasing (1). This skepticism may be fostered by inflationary interpretations of NI. For example, the claim that a newborn less than an hour old is capable of recognizing that the acts of self are like the acts of others (Meltzoff 2002, 2005, 2007a, 2013; Meltzoff and Decety 2003) strikes us as implausible. This problematic claim may originate in a confusion at the root of the AIM model. AIM conflates a principle of cognition with what is cognized in cognition. It seems to assume that, because a relation of similarity is part of what explains imitative behavior, it must also become the content, or "object," of a cognitive act that targets precisely that particular relation between relata. This is the recognition experience "a is like b." However, in its most primitive and fundamental functioning, the relation of similarity is operative in an associative process. Association by similarity regulates the activation of cognitive processes (e.g., in perception, the activation of a complex object representation given the processing of characteristic features of a stimulus), but there is no comparison that detects the relation of similarity between cognitive processes or between objects represented by cognitive processes. In the ideomotor/common coding theory of imitation and in AST, association by similarity is merely the process that regulates how visual representations tend to activate action plans.

A developmental psychologist who considers recent reviews of the findings and recent empirical studies may conclude that NI does not exist and that further research

on this topic would be a waste of time and resources. If NI is what the dominant model posits, i.e. a goal-directed behavior entailing an act of recognition that is foundational for the development of social cognition, then the findings suggest there is no such thing. Yet, thanks to AST, NI may appear once again within the reach of empirical validation. Developmental psychologists may more easily adopt AST as working hypothesis because it requires not much else than background assumptions they already accept. They may be newly intrigued by the remarkable response that a simple but carefully thought-out experimental design may evoke in infants. Hence, AST may contribute to solving the question of whether early differential imitation really exists.

Author Contributions SV is lead author and wrote the final version. YJ made contributions to a previous version that were integrated in the final version. YJ reviewed and accepted the final version.

References

Allen, C. 2012. Private codes and public structures. In *The complex mind: An interdisciplinary approach*, ed. D. McFarland, K. Stenning, and M. McGonigle-Chalmers, 223–242. Houndmills: Palgrave-Macmillan.

Andonotopo, W., and A. Kurjak. 2006. The assessment of fetal behavior of growth restricted fetuses by 4D sonography. *Journal of Perinatal Medicine* 34: 471–478.

Anisfeld, M. 1991. Neonatal imitation: Review. *Developmental Review* 11: 60–97.

Anisfeld, M. 2005. No compelling evidence to dispute Piaget's timetable of the development of representational imitation in infancy. *Perspectives on Imitation: From Neuroscience to Social Science* 2: 107–131.

Anisfeld, M., G. Turkewitz, S.A. Rose, F.R. Rosenberg, F.J. Sheiber, D.A. Couturier-Fagan, and I. Sommer. 2001. No compelling evidence that newborns imitate oral gestures. *Infancy* 2 (1): 111–122. https://doi.org/10.1207/S15327078IN0201_7.

Barsalou, L.W. 2008. Grounded cognition. *Annual Review of Psychology* 59: 617–645.

Campbell, D.T., and J.C. Stanley. 1966. *Experimental and quasi-experimental designs for research*. Chicago: Rand-McNally.

Clark, A. 2013. Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences* 36 (03): 181–204.

Cook, R., G. Bird, C. Catmur, C. Press, and C. Heyes. 2014. Mirror neurons: From origin to function. *Behavioral and Brain Sciences* 37 (2): 177–192.

Coulon, M., C. Hemimou, and A. Streri. 2013. Effects of seeing and hearing vowels on neonatal facial imitation. *Infancy* 18 (5): 782–796.

D'Elia, A., M. Pighetti, G. Moccia, and N. Santangelo. 2001. Spontaneous motor activity in normal fetuses. *Early Human Development* 65: 139–147.

Decock, L., and I. Douven. 2011. Similarity after Goodman. *Review of Philosophy and Psychology* 2: 61–75.

Dreyfus, H., and S. Dreyfus. 1999. The challenge of Merleau-Ponty's phenomenology of embodiment for cognitive science. In *Perspectives on embodiment: The intersection of nature and culture*, ed. G. Weiss and H.F. Haber, 103–120. New York: Routledge.

Ekman, P., and E.L. Rosenberg. 1997. *What the face reveals: Basic and applied studies of spontaneous expression using the facial action coding system (FACS)*. Oxford: Oxford University Press.

Froese, T., and D.A. Leavens. 2014. The direct perception hypothesis: Perceiving the intention of another's action hinders its precise imitation. *Frontiers in Psychology* 5: 65. https://doi.org/10.3389/fpsyg.2014.00065.

Gallagher, S. 2005. *How the body shapes the mind*. Oxford: Oxford University Press.

Giese, M.A., and G. Rizzolatti. 2015. Neural and computational mechanisms of action processing: Interaction between visual and motor representations. *Neuron* 88 (1): 167–180.

Griffiths, P. 2002. What is innateness. *The Monist* 85 (1): 70–85.

Hata, T., K. Kanenishi, M. Akiyama, H. Tanaka, and K. Kimura. 2005. Real-time 3-D sonographic observation of fetal facial expression. *Journal of Obstetrics and Gynaecology Research* 3: 337–340.

Heimann, M. 2000. Imitation – a "fuzzy" phenomenon? In *Emerging cognitive abilities in early infancy*, ed. F. Lacerda, C.V. Hofsten, and M. Heimann, 231–246. Mahwah: Erlbaum.

Heimann, M. 2002. Notes on individual differences and the assumed elusiveness of neonatal imitation. In *The imitative mind: Development, evolution, and brain bases*, 74–84.

Heimann, M., and J. Schaller. 1985. Imitative reactions among 14-21 days old infants. *Infant Mental Health Journal* 6 (1): 31–39.

Heimann, M., K.E. Nelson, and J. Schaller. 1989. Neonatal imitation of tongue protrusion and mouth opening: Methodological aspects and evidence of early individual differences. *Scandinavian Journal of Psychology* 30 (2): 90–101.

Hepper, P.G. 2015. Behavior during the prenatal period: Adaptive for development and survival. *Child Development Perspectives* 9 (1): 38–43.

Hommel, B., and B. Elsner. 2009. Acquisition, representation, and control of action. In *Oxford handbook of human action*, eds. E. Morsella, J. A. Bargh, and P. M. Gollwitzer, 371–398. New York: Oxford University Press.

Hume, D. 2000. *A treatise of human nature*. Oxford: Oxford University Press.

Husserl, E. (1999). *Cartesian meditations: An introduction to phenomenology* (D. Cairns, Trans.). The Hague: Martinus Nijhoff. (Original work published 1931).

James, W. 1981. *Principles of psychology*. Cambridge: Harvard University Press.

Jones, S. 2009. The development of imitation in infancy. *Philosophical Transactions of the Royal Society, B: Biological Sciences* 364: 2325–2335.

Katz, K., R. Mashiach, and I. Meizner. 2007. Normal range of fetal finger movements. *Journal of Pediatric Orthopaedics. Part B* 16: 252–255.

Kennedy, C. 2005. *Single-case designs for educational research*. Boston: Pearson.

Keven, N., and K.A. Akins. 2016. Neonatal imitation in context: sensorymotor development in the perinatal period. *Behavioral and Brain Sciences Sci* 1–107. https://doi.org/10.1017/S0140525X16000911.

Kratochwill, T.R. 1992. Single-case research design and analysis: An overview. In *Single-case research design and analysis: New directions for psychology and education*, ed. T.R. Kratochwill and J.R. Levin, 1–14. Hillsdale: Lawrence Erlbaum Associates.

Kugiumutzakis, G. 1999. Genesis and development of early infant mimesis to facial and vocal models. In *Imitation in infancy*, ed. J. Nadel and G. Butterworth, 36–59. New York: Cambridge University Press.

Kurjak, A., A. Tikvica, M. Stanojevic, B. Miskovic, B. Ahmed, G. Azumendi, and G.C. Di Renzo. 2008. The assessment of fetal neurobehavior by three-dimensional and four-dimensional ultrasound. *The Journal of Maternal-Fetal and Neonatal Medicine* 21: 675–684.

Kurjak, A., G. Azumendi, N. Vecek, S. Kupesic, M. Solak, D. Varga, and F. Chervenak. 2003. Fetal hand movements and facial expression in normal pregnancy studied by four-dimensional sonography. *Journal of Perinatal Medicine* 3: 496–508.

Kurjak, A., M. Stanojevic, W. Andonotopo, A. Salihagic-Kadic, J.M. Carrera, and G. Azumendi. 2004. Behavioral pattern continuity from prenatal to postnatal life a study by four-dimensional (4D) ultrasonography. *Journal of Perinatal Medicine* 32: 346–353.

Larkey, L.B., and A.B. Markman. 2005. Processes of similarity judgment. *Cognitive Science* 29: 1061–1076.

Legerstee, M. 1991. The role of person and object in eliciting early imitation. *Journal of Experimental Child Psychology* 51: 423–433.

Lodder, P., M. Rotteveel, and M. van Elk. 2014. Enactivism and neonatal imitation: Conceptual and empirical considerations and clarifications. *Frontiers in Psychology* 5: 967.

Maclaurin, J. 2002. The resurrection of innateness. *The Monist* 85: 105–130.

Massen, C., and W. Prinz. 2009. Movements, actions and tool-use actions: An ideomotor approach to imitation. *Philosophical Transactions of the Royal Society B* 364: 2349–2358.

Meltzoff, A.N. 1999. Origins of theory of mind, cognition and communication. *Journal of Communication Disorders* 32: 251–269.

Meltzoff, A.N. 2002. Imitation as a mechanism of social cognition: Origins of empathy, theory of mind, and the representation of action. In *Blackwell handbook of childhood cognitive development*, ed. U. Goswami, 6–25. Oxford: Blackwell Publishers.

Meltzoff, A. N. (2005). Imitation and other minds: The "like me" hypothesis. From: Hurley, S. & Chater, N. (Eds.), Perspectives on imitation: From neuroscience to social science. Vol. 2, pp. 55–77. Cambridge: MIT Press.

Meltzoff, A.N. 2007a. 'Like me': A foundation for social cognition. *Developmental Science* 10 (1): 126–134.

Meltzoff, A.N. 2007b. The 'like me' framework for recognizing and becoming an intentional agent. *Acta Psychologica* 124: 26–43.

Meltzoff, A.N. 2009. Roots of social cognition: The like-me framework. In *Minnesota symposia on child psychology: Meeting the challenge of translational research in child psychology*, vol. 35, 29–58. Hoboken: Wiley.

Meltzoff, A. N. 2010. Bridging between action representation and infant theory of mind. In *Cognition and Neuropsychology: International Perspectives on Psychological Science*, 15–33. Psychology Press. https://doi.org/10.4324/9780203845820.

Meltzoff, A.N. 2013. Origins of social cognition: Bidirectional self-other mapping and the "like-me" hypothesis. In *Navigating the social world: What infants, children, and other species can teach us*, ed. M. Banaji and S. Gelman, 139–144. New York: Oxford University Press.

Meltzoff, A. N., and M.K Moore. 1983. Newborn infants imitate adult facial gestures. *Child Development* 54: 702–709.

Meltzoff, A.N., and J. Decety. 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., and M.K. Moore. 1977. Imitation of facial and manual gestures by human neonates. *Science* 198 (4312): 75–78.

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

Meltzoff, A.N., and M.K. Moore. 1992. Early imitation within a functional framework: The importance of person identity, movement, and development. *Infant Behavior & Development* 15: 479–505.

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

Meltzoff, A.N., and M.K. Moore. 1997. Explaining facial imitation: A theoretical model. *Early Development & Parenting* 6: 179–192.

Merleau-Ponty, M. 1964a. *The primacy of perception: And other essays on phenomenological psychology, the philosophy of art, history, and politics*. Evanston: Northwestern University Press.

Merleau-Ponty, M. 1964b. *The primacy of perception*. Trans. W. Cobb. Evanston: Northwestern University Press.

Merleau-Ponty, M. (2012). *Phenomenology of perception*. (D. A. Landes, Trans.). London: Routledge.

Meyer, K., and A. Damasio. 2009. Convergence and divergence in a neural architecture for recognition and memory. *Trends in Neurosciences* 32: 376–382.

Mongillo, G. 2012. Hebbian learning. In *Encyclopedia of the Sciences of Learning*, ed. N. Seel. Springer Science+Business Media, LLC. https://doi.org/10.1007/978-1-4419-1428-6.

Nagy, E., and P. Molnar. 2004. Homo imitans or homo provocans? Human imprinting model of neonatal imitation. *Infant Behavior & Development* 27 (1): 54–63.

Nagy, E., H. Kompagne, H. Orvos, and A. Pal. 2007. Gender-related differences in neonatal imitation. *Infant and Child Development* 16: 267–276. https://doi.org/10.1002/icd.497.

Nagy, E., K. Pilling, H. Orvos, and P. Molnar. 2013. Imitation of tongue protrusion in human neonates: Specificity of the response in a large sample. *Developmental Psychology* 49 (9): 1628.

Nosofsky, R.M. 1992. Similarity scaling and cognitive process models. *Annual Review of Psychology* 43 (1): 25–53.

Oostenbroek, J., V. Slaughter, M. Nielsen, and T. Suddendorf. 2013. Why the confusion around neonatal imitation? *A review Journal of Reproductive and Infant Psychology* 31 (4): 328–341.

Oostenbroek, J., T. Suddendorf, M. Nielsen, J. Redshaw, S. Kennedy-Costantini, J. Davis, et al. 2016. Comprehensive longitudinal study challenges the existence of neonatal imitation in humans. *Current Biology* 26 (10): 1334–1338.

Piaget, J. (1962). *Play, dreams and imitation in childhood* (C. Attegno & F.M. Hodgson, Trans.). New York: Norton. (Original work published 1951.)

Piontelli, A. 2015. *Development of normal fetal movements: The last 15 weeks of gestation*. Milan: Springer-Verlag.

Prinz, W. 1990. A common coding approach to perception and action. In *Relationships between perception and action: Current approaches*, ed. O. Neumann and W. Prinz, 167–201. Berlin: Springer.

Prinz, W. 1997. Perception and action planning. *European Journal of Cognitive Psychology* 9: 129–154.

Prinz, W. 2002. Experimental approaches to imitation. In *The imitative mind: Development, evolution, and brain basis*, ed. A.N. Meltzoff and W. Prinz, 143–162. Cambridge: Cambridge University Press.

Prinz, W. 2005. An ideomotor approach to imitation. In *Perspectives on imitation*, ed. S. Hurley and N. Chater, vol. 1, 141–156. Cambridge: MIT Press.

Prinz, W. 2012. *Open minds: The social making of agency and intentionality.* Cambridge: MIT Press.

Prinz, W., G. Aschersleben, and I. Koch. 2009. Cognition and action. In *Oxford handbook of human action,* ed. E. Morsella, J.A. Bargh, and P.M. Gollwitzer, vol. 2, 35–71. Oxford: Oxford University Press.

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

Reissland, N., B. Francis, J.M. Mason, and K. Lincoln. 2011. Do facial expressions develop before birth? *PLoS One* 6: 24081.

Reissland, N., C. Mason, B. Schaal and K. Lincoln. 2012. Prenatal mouth movements: Can we identify co-ordinated fetal mouth and LIP actions necessary for feeding? International Journal of Pediatrics, 848596, 5 pages.

Rizzolatti, G., and C. Sinigaglia. 2010. The functional role of the parieto-frontal mirror circuit: Interpretations and misinterpretations. *Nature Reviews Neuroscience* 11 (4): 264–274.

Rochat, P., and T. Striano. 1999. Social cognitive development in the first year. In *Early social cognition: Understanding others in the first months of life,* ed. P. Rochat, 257–280. Mahwah: Lawrence Erlbaum Associates.

Roodenburg, P.J., J.W. Wladimiroff, and H.F. Prechtl. 1991. Classification and quantitative aspects of fetal movements during the second half of normal pregnancy. *Early Human Development* 25 (1): 19–35.

Sheets-Johnstone, M. 2011. *The primacy of movement.* 2nd ed. New York: John Benjamins.

Shepard, R.N. 1987. Toward a universal law of generalization for psychological science. *Science* 237: 1317–1323.

Simion, F., E. Di Giorgio, I. Leo, and L. Bardi. 2011. The processing of social stimuli in early infancy: From faces to biological motion perception. *Progress in Brain Research* 189: 173–193.

Simpson, E.A., L. Murray, A. Paukner, and P.F. Ferrari. 2014. The mirror neuron system as revealed through neonatal imitation: Presence from birth, predictive power and evidence of plasticity. *Philosophical Transactions of the Royal Society B* 369 (1644): 20130289.

Sommerville, J.A., A.L. Woodward, and A. Needham. 2005. Action experience alters 3-month-old perception of other's actions. *Cognition* 96 (1): 1–11.

Strid, K., T. Tjus, L. Smith, A.N. Meltzoff, and M. Heimann. 2006. Infant recall memory and communication predicts later cognitive development. *Infant Behavior & Development* 29 (4): 545–553. https://doi.org/10.1016/j.infbeh.2006.07.002.

Subiaul, F. 2010. Dissecting the imitation faculty: The multiple imitation mechanisms hypothesis. *Behavioral Processes* 83 (2): 222–234.

Tomasello, M. 1999. *The cultural origins of human cognition.* Cambridge: Harvard University Press.

Tramacere, A., P.F. Ferrari, and A. Iriki. 2015. Epigenetic regulation of mirror neuron development, and related evolutionary hypotheses. In *New frontiers in mirror neurons research,* ed. P.F. Ferrari and G. Rizzolatti, 222–244. Oxford: Oxford University Press.

Trevarthen, C., and K.J. Aitken. 2001. Infant intersubjectivity: Research, theory, and clinical applications. *Journal of Child Psychology and Psychiatry* 42 (1): 3–48.

Tversky, A. 1988. Features of similarity. In *Readings in cognitive science,* ed. A. Collins and E. Smith, 290–302. San Mateo: Morgan Kaufmann.

Ullstadius, E. 1998. Neonatal imitation in a mother–infant setting. *Early Development and Parenting* 7 (1): 1–8.

Van Heteren, C.F., P.F. Boekkooi, H.W. Jongsma, and J.G. Nijhuis. 2000. Fetal learning and memory. *Lancet* 356: 1169–1170.

Vetter, P., and A. Newen. 2014. Varities of cognitive penetration in visual perception. *Consciousness and Cognition* 27: 62–75.

Vigo, R. 2009. Modal Similarity. *Journal of Experimental & Theoretical Artificial Intelligence* 21 (3): 181–196.

Vigo, R., and C. Allen. 2009. How to reason without words: Inference as categorization. *Cognitive Processing* 10 (1): 77–88.

Vincini, S., Y. Jhang, E.H. Buder, and S. Gallagher. 2017. Neonatal imitation: Theory, experimental design, and significance for the field of social cognition. *Frontiers in Psychology* 8: 1323. https://doi.org/10.3389/fpsyg.2017.01323.

Yigiter, A.B., and Z.N. Kavak. 2006. Normal standards of fetal behavior assessed by four-dimensional sonography. *Journal of Maternal-Fetal and Neonatal Medicine* 19: 707–721.

Zoia, S., L. Blason, G. D'ottavio, M. Bulgheroni, E. Pezzetta, A. Scabar, and U. Castiello. 2007. Evidence of early development of action planning in the human fetus: A kinematic study. *Experimental Brain Research* 176: 217–226.