

A Metastable Attractor Model of Self-Other Integration (MEAMSO) in Rhythmic Synchronization

Ole Adrian Heggli^{1,*}, Ivana Konvalinka², Morten L. Kringelbach^{1,3}, Peter Vuust¹

1. Center for Music in the Brain, Department of Clinical Medicine, Aarhus University and the Royal Academy of Music Aarhus/Aalborg, Aarhus, Denmark.

2. SIne Lab, Section for Cognitive Systems, DTU Compute, Technical University of Denmark, Kongens Lyngby, Denmark.

3. Center for Eudaimonia and Human Flourishing, Department of Psychiatry, University of Oxford, Oxford, UK.

* Corresponding author at ole.heggli@clin.au.dk

Abstract

Human interaction is often accompanied by synchronized bodily rhythms. Such synchronization may emerge spontaneously as when a crowd's applause turns into a steady beat, be encouraged as in nursery rhymes, or be intentional as in the case of playing music together. The latter has been extensively studied using joint finger tapping paradigms as a simplified version of rhythmic interpersonal synchronization. A key finding is that synchronization in such cases is multifaceted, with synchronized behaviour resting upon different synchronization strategies such as mutual adaptation, leading-following, and leading-leading. However, there are multiple open questions regarding the mechanism behind these strategies and how they develop dynamically over time. Here, we propose a metastable attractor model of self-other integration (MEAMSO). This model conceptualizes dyadic rhythmic interpersonal synchronization as a process of integrating and segregating signals of self and other. Perceived sounds are continuously evaluated as either being attributed to *self*-produced or *other*-produced actions. The model entails a metastable system with two particular attractor states: one where an individual maintains two separate predictive models for *self*- and *other*-produced actions, and the other where these two predictive models integrate into one. The MEAMSO explains the three known synchronization strategies, and makes testable predictions about the dynamics of interpersonal synchronization both in behaviour and the brain.

Introduction

The ability to coordinate one's movements with those of others is a critical aspect of daily life. An everyday action such as shaking hands in greeting is a complex feat of reciprocally anticipating and adjusting movements, yet we perform it with ease and without apparent conscious effort [1]. One of the most fascinating aspects of such adaptation is interpersonal synchronization, wherein periodic movements exhibit phase and frequency locking between persons [2]. Such synchronization often emerges spontaneously, as in the tendency towards synchronized walking, or in a crowd's applause converging on a steady rhythm [3, 4]. However, in many cases interpersonal synchronization is necessary and intentional, for instance in human bucket brigades [5]. The most salient example is perhaps found in musical performance.

When musicians sing and play music together, each individual musician coordinates their actions towards producing sounds which in concert with other musicians creates the cohesive auditory patterns we identify as music [6]. A critical observation here is that a reactive approach is insufficient. To produce a synchronized rhythm with other musicians, any action producing a sound needs to be initiated well before the actual sound is produced [7]. Hence, intentional rhythmic synchronization must rely on processes of anticipation and prediction [8, 9].

An often-overlooked aspect of such predictive processes is the selection of *which* perceptual signal one should adapt to. Any decent musician in an ensemble is capable of recognizing whether another ensemble member is deviating from the ensemble's shared musical goal, and then swiftly perform an on-the-fly judgment whether the deviating member should be ignored, or adapted to. To do so, musicians continuously monitor how the actions of others relate to self-produced actions, as well as the combined action outcomes [10, 11]. In laboratory settings, this type of interpersonal synchronization has predominantly been studied using joint finger tapping paradigms [10, 12-14].

In joint finger tapping two participants are asked to perform simple isochronous rhythms together, often initiated by a few beats of a metronome, with instructions to maintain tempo and synchronize. A key finding from the field is that rhythmic interpersonal synchronization may be more than just a linear transition from unsynchronized to synchronized actions. Rather, different interpersonal strategies underlie synchronization, such as mutual adaptation and leading-following [10, 12-14]. Note that here we use the term strategy loosely, as the emergence and dynamics of dyad synchronization behaviour is still a topic under active research. Mutual adaptation occurs when bidirectionally or omnidirectionally coupled dyad members continuously and reciprocally adapt to each other tap-by-tap, and has been observed in both musicians and non-musicians [10, 12]. Leading-following can occur either due to experimental restraints or emerge from the interaction. For instance, in cases of

unidirectional coupling wherein dyad member 1 hears themselves, and dyad member 2 hears dyad member 1, the dyad has to adapt a leader-follower strategy to maintain synchronization [12]. However, naturally emerging leader-follower behaviour has also been observed in bidirectionally coupled interactions between non-musicians [13], and are routinely found in musical interactions [15]. Common for all of these strategies is that they are not necessarily differentiable based on synchronization measures alone, but can instead be detected by looking at the interaction between the signals of the interacting members, i.e. using cross-correlation measures between participants' inter-tap intervals [12].

While these synchronization strategies are well documented and prevalent in the tapping literature, there are multiple open questions as to their emergence and nature. For instance, to what extent are synchronization strategies dynamic, in the sense that an interaction may conceivably continuously swing between different strategies? Or, which parameters of individual behaviour in an otherwise equally constrained interaction contributes to the prevalence of one synchronization strategy over another?

In previous research, individual differences such as people's preferred tapping rate have been shown to influence synchronization, but their impact on the behaviour underlying synchronization remains unclear [16, 17]. Similarly, the dynamics of responses to perturbations in synchronization tasks such as rate changes in the synchronization target may impact synchronization strategies, and has been shown to depend on individual differences [18-21]. Interestingly, in a drumming task, toddlers (2- and 3-year-olds) were able to spontaneously coordinate rhythmic interaction with a peer, but only 4-year-old children exhibited continuous adaptability of their drumming [22]. This age distinction clearly links to the development of a cognitive self-other distinction as measured by the classic false-belief test that typical 3-year-olds fails but typical 4-year-olds pass [23]. These findings suggest that interpersonal synchronization is influenced both by development and individual differences, and crucially relies on the ability to distinguish between the self and the other.

Over the last four decades multiple researchers have proposed models aiming to capture the underlying processes and dynamics of sensorimotor synchronization, and recently also for interpersonal synchronization [8, 24-30]. Generally, these models can be classified as information processing event-based models or nonlinear coupled-oscillator dynamic models. Both approaches have shown promise, with a dual-agent version of the adaptation and anticipation model (ADAM) capturing leading-following and mutual adaptation behaviour [31], and with our Kuramoto-oscillators model also managing to account for leading-leading behaviour [24]. A third alternative also exists in the strong anticipation hypothesis, wherein anticipatory behaviour in dynamical systems is shown to

emerge from time-delays [30, 32, 33]. Nonetheless, these models have not explicitly focused on the underlying neural basis for how and why dyadic synchronization strategies emerge. In this paper we present a theoretical model that explains how simple predictive brain dynamics give rise to synchronization strategies. Furthermore, we show how to use the model for deriving testable hypotheses of dyadic synchronization behaviour.

The metastable attractor model of self-other integration

In previous work we, and others, have argued that intentional interpersonal synchronization may be framed through mechanisms of self-other integration [2, 24, 34-40]. Specifically, we have proposed that mutual adaptation occurs when two dyad members both integrate the self and other, that leading-following occurs when one of the two dyad members (the follower) integrates the other while the other (the leader) segregates, and that leading-leading occurs when both dyad members are segregating between the self and other [24]. Here, *self* refers to actions performed by oneself and their related consequences, whereas *other* refers to information stemming from the person one is interacting with. This argument is based on observations of overlapping brain networks processing both perceived/observed and performed actions, wherein action-perception loops blur the difference between self and other may emerge [34, 41-44]. To explain how these processes account for not just synchronization, but also synchronization strategies, here we present a model where self-other integration is conceptualised as a continuous process wherein action and perception are constantly evaluated as either connected or unconnected.

The foundational assumption of the metastable attractor model of self-other integration (MEAMSO) is that the brain maintains predictive cognitive models for the *self*, i.e. the actions performed, and for the *other*, i.e. the perception of actions performed by the person one is interacting with [45]. In joint finger tapping, the self-model initiates finger taps, and then receives somatosensory and auditory feedback when their finger hits the tapping device. The other-model relies on auditory feedback from the other dyad member's taps. If the dyad's goal is only to synchronize, then each dyad member needs to collectively minimize the difference between event predictions formed by their self- and other-model [2]. We propose that this process of integrating or segregating between the self- and other-model is the result of a continuous evaluation of percepts as either being attributed to the self or to the other; or, in other words, whether perceptions are causally linked to actions. We hypothesize that self-other integration and segregation act as attractors, or transiently stable states, in the configuration of brain network dynamics, underpinning synchronization strategies. By metastable we mean a system exhibiting stable states other than the state of least energy, operationalized in the brain as neural oscillations existing in a dynamic system which switches between states due to both external perturbations or internal fluctuations [46-50]. The goal of the MEAMSO is therefore to

formalize the qualities and dynamics of a dyadic interaction resulting in the dyad members either being in a state of self-other integration or segregation.

Mapping the dynamics of self-other integration

The first step in mapping the dynamics of self-other integration in joint action is to understand how we are aware that an action and its perceptual consequence are self-generated. In the last two decades, it has become increasingly evident that sensory predictions form a crucial part of determining the likelihood of an action and a concurrent perception being self-generated [35]. In a seminal study from the late 90s, Fournieret and Jeannerod showed that participants are more aware of intended movements rather than actual deviated movements, but only up until a crucial threshold [51]. In addition, people are prone to confuse self-produced hand movements with those performed by someone else given a high enough similarity [52, 53]. Within the framework of predictive coding, these observations may be explained by predictions being formed about the sensory consequences of an action, where only prediction violations of a certain magnitude traverse high enough in the prediction error hierarchy to reach conscious appraisal [54]. It is also likely that top-down control modulates these decisions, through imposing a certain leniency on the conscious appraisal of a perceptual event deviating substantially from the predicted sensory consequences [55]. For instance, if only one isolated event is surprising in a series of predictable events, the surprising event may be considered just a lapse in attention [56].

An interesting avenue of exploring the latter component can be found in amplified musical performance settings. Here musicians usually receive an amplified signal of their instrument as a monitor signal, in order to be able to self-monitor their performance and integrate with other musicians [57]. Due to the inherent delays in both digital and analogue audio equipment, and acoustic propagation delay, this signal may deviate by multiple milliseconds from the original output signal of the musical instrument. In these cases, performers are consciously aware that the auditory perceptions they hear are the causally linked to their actions. Nonetheless, once the delay is sufficiently large, the situation becomes disruptive for the musicians, and negatively impacts their performance [58]. Interestingly, the thresholds here appear to be both experience- and instrument-specific, yet at a latency of between 10 to 15 ms most musicians report performance issues [59]. This can be interpreted as the edge of a time window wherein self-produced actions and perceptions are likely to be perceived as causally linked.

The second step in mapping the dynamics of self-other integration in joint action is understanding how perception, in this case sound, is related to another person's intentions and goals. A core tenet in social neuroscience is the ability to understand that others may hold mental states dissimilar to

oneself, and that people's mental states form the basis of their behaviour [60, 61]. In joint action this entails understanding that the person we are interacting with may not hold the same idea of the task, or even the same goal. Nonetheless, experimental manipulation of underlying predictive models in a polyrhythmic interpersonal synchronization task shows that participants quickly harmonize their predictive models leading to synchronized behaviour [10]. Interestingly, if one considers two interacting brains wherein their respective sensory input is the output of their interacting partner's actions, this forms a reciprocally coupled system wherein mutual prediction, e.g. a harmonization of the predictive models, is achieved as an emergent property of the coupling [45]. In the MEAMSO we do not explicitly model a three-way separation of perception (i.e. perceiving something as linked to the self, to the other, or as unrelated to both), nor does the model explicitly model the combined joint action outcome, such as when two musicians are playing a duet together (wherein the music individually performed may differ in both melody and rhythm) and hence requiring monitoring of the shared goal [11, 62]. Instead, we work on the premise that if perceptions stemming from the other (in this case the sound from their taps) cannot be causally attributed to one's own actions, they are instead sorted into the predictive model of the other. While this is a simplification, it is also a necessary step in order to reduce the number of comparisons and calculations performed by the MEAMSO.

Metastable attractor model of self-other integration (MEAMSO)

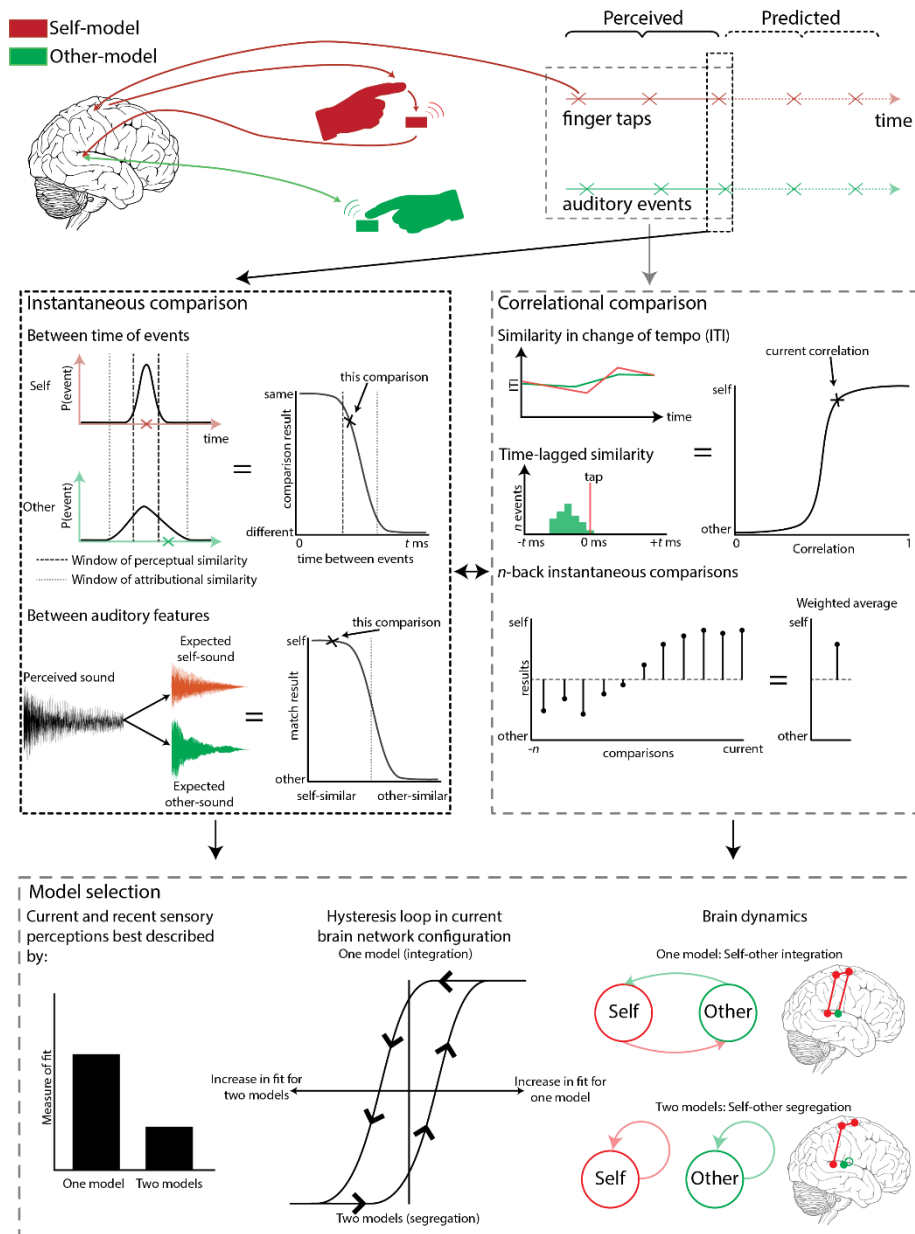


Figure 1 – The Metastable Attractor Model of Self-Other Integration (MEAMSO). A) An overview of the process that MEAMSO models, here illustrated for a joint finger tapping task. MEAMSO assumes that a participant in the task maintains two predictive models, one self-model governing the actions performed, and an other-model governing the perceptions from the actions of the other. Together these predictive models control the perception of sensations, and predict upcoming sensations. We operationalize the MEAMSO in three modules: B) Instantaneous comparisons performed between events occurring close in time. We highlight two main comparisons here, one between the relative time of events, and another between auditory features of the sounds of events. This comparisons both lead to a value indicating a likelihood of perceptions being linked to actions. C) Correlational comparisons are performed over successive events, including a measure of similarity of tempo (measured by the intertap interval in the case of an isochronous rhythm), a time-lagged similarity to account for simple delays between a tap and a resulting sound, as well as an accumulation of previous instantaneous comparisons. D) The two comparison modules are piped into a model selection wherein the corroboration for the current event best being described by one model (a collapse of the predictive self- and other-model), or by two separate predictive model, is collected. The most

likely current state (integration or segregation) is fed through a hysteresis loop, which then determines the current state of brain dynamics best describing the stream of actions and perceptions.

MEAMSO modules

The MEAMSO (illustrated in Figure 1) consists of three primary interconnected modules: 1. instantaneous comparisons, 2. correlational comparisons, and 3. model selection. It takes two time series as input. In our example joint finger tapping task, one of these time series represents the actions performed by the model's target person, and the other time series represents the perceptions stemming from the tapping partner. Throughout the description of the model we refer to these as *actions* and *perceptions*, respectively. The main outcome from the model is a state indicating whether two time series are best described by one predictive model (self-other integration), or by two separate predictive models (self-other segregation). In the following paragraphs we describe key parts of the model. A detailed overview including equations can be found in the Supplementary Appendix 1.

1. Instantaneous comparisons

This module handles the more-or-less instantaneous comparison between two events occurring simultaneously or in rapid succession. The first component is a comparison between the time of events, leading to a decision or likelihood of whether the two events are causally linked. Under the assumption that two predictive models are active, one for self-produced actions (the self-model), and one for other-produced actions (the other-model), this comparison is perceptually a measure of the asynchrony between the self-performed finger tap, and the auditory perception of the other-performed finger tap. We propose that there are two time-windows at play here, one wherein there may exist a perceptual similarity, i.e. when the asynchrony is low, and another time-window for medium asynchronies where there may exist an attributional similarity, i.e. where top-down processes actively attribute the asynchronous event as being causally linked to the originating event. Computationally we can describe this by calculating the difference in time between events, and passing the value through a logistic function. As we can assume that a finger tap preceding an auditory signal is more likely to be causally related than the reverse chronology, we construct a transfer function taking this into account by using a reversed logistic function for asynchronies > 0 , and a regular logistic function for asynchronies < 0 . By applying a min-max normalization to the input asynchrony, we can express this process with the following two equations:

$$async_s = \begin{cases} t_{max} & abs(async) \geq t_{max} \\ \frac{2(abs(async) - t_{min})}{t_{max} - t_{min}} - 1 & t_{min} < abs(async) < t_{max} \\ t_{min} & abs(async) \leq t_{min} \end{cases} \quad (1)$$

$$comp_{time} = \begin{cases} \frac{2}{1 + e^{k_1(async_s)}} - 1 & async > 0 \\ 0 & async = 0 \\ \frac{2}{1 + e^{-k_2(async_s)}} - 1 & async < 0 \end{cases} \quad (2)$$

Here, $async$ is the time between events from the point of view of the finger tap, which is scaled to a value between -1 and +1, where t_{max} indicates a cut-off point in milliseconds representing the maximum length of the time-window wherein action and perception may be linked, and t_{min} is equal to 0. Dependent on the sign of the $async$ this goes through a logistic function, where k_1 is the steepness of the curve for positive asynchronies, and k_2 for negative asynchronies. Their respective values are likely individual, and influenced by experience and expertise [63]. We can then construct a function where when $comp_{time}$ approaches 1, we assume that action and perception are causally linked, and when it approaches -1, we assume they are not linked.

The second component is a comparison between auditory features. In a normal musical interaction, musicians are intimately familiar with the sounds of their own instrument. This relies on experience and learning where one builds a cognitive template of what perceptions match the actions performed on an instrument, and consists of multiple acoustic features such as the frequency spectrum and envelope (known as timbre). There exist multiple perceptual models for how we identify and classify individual sounds from a combined sound source, which could be integrated into the MEAMSO [64-67]. Due to the complexity of this comparison, we set its value to zero for this iteration of our model, indicating that the comparison does not influence the later decisions made by model.

2. Correlational comparisons

We know from multiple joint finger-tapping studies that the intertap intervals (the time between successive taps) of the dyads rarely reach a high positive correlation value at lag 0 (regular correlation) [10, 12-14]. Instead, positive correlation values are found at both lag +1 and lag -1 for the mutual adaptation strategy, at either lag +1 or lag -1 for the leading-following strategy, or with weak positive values on all lags in the leading-leading strategy. Yet, and especially in mutual adaptation, even with alternating auditory feedback from their own tapping, participants still report a sense of agency over their perceptual feedback [68, 69]. Hence, it is likely that a correlational similarity between perceptions and actions may, at a certain threshold, contribute evidence towards perceiving a causal relationship between the two. For instance, if considering the intertap intervals we presume that a

constant event-lagged similarity may instil a likelihood of ascribing a causal relationship between actions and perceptions. As an example, we can consider a slow increase in tempo, an *accelerando*. Here, the intertap interval is decreased for each subsequent tap usually in a non-linear matter. We assume that this behaviour may contribute to a sensation of causal links between action and perception, even if the changes are lagged one event behind the current action. To capture this, we calculate a cross-correlation at lag -1 and lag 0 between the intertap intervals of own actions vs. others' actions, and select the highest correlation value. As this calculation necessitates a certain number of events before it becomes informative, we start the calculation only when there are four intertap intervals available (at event number five), and implement a sliding window wherein only the last eight events are included, hence building up the stability of the calculation as the length of the interaction increase.

As previously discussed, while the musicians will complain of the latency, they are capable of recognizing and performing with the perceptions of their actions being delayed by a certain time period. This is a time-lagged similarity, and we deem it likely that even a certain jitter in the time-lagged perception is acceptable in terms of causally relating the perception to the action. This calculation involves estimating the mean asynchrony between an action and a perception event, and its variability. If the variability is low, indicating a stationary latency, this calculation contributes towards linking action and perception.

The last correlational comparison included in the MEAMSO is an n-back instantaneous comparison. Here, we assume that a weighted average of previous comparisons is held, to decrease the influence of erroneous instantaneous comparisons. Here we use a Bayesian average, which is updated by the four previous instantaneous comparisons, as shown in the following equation: $\bar{x} = \frac{Cm + \sum_{i=1}^n x_i}{C+n}$, with a prior m , and a constant C [70]. This is equal to padding the average with C datapoints of m . For every update of the calculation, we update the prior to be equal to its current output, thus creating a smoothed weighted average. Notably, the correlational comparisons described above assume that the rhythm performed is isochronous, and would not function for complex rhythms, or rhythms with a high degree of syncopation. To do so one would need to implement an expectation template that accounts for multiple event types, due to the finding that perturbations to individual events within a complex rhythm differentially affects perception [71]. A promising approach to building such an expectation template can be found in the recently developed Phase Inference from Point Process Event Timing framework, which could serve as a filter prior to the correlational comparisons described above [72].

3. Model selection

All the above-mentioned processes are constructed to produce a numeric output ranging from -1 to 1, with values approaching -1 indicating that actions and perceptions are not causally linked, and values approaching 1 indicating a causal link between actions and perceptions. We can then sum all values > 0 as a measure of fit for self-other integration, and all values < 0 for self-other segregation. By using a softmax normalization on the absolute measures of fit followed by an argmax to determine whether the fit is best for integration or segregation, we get an integer representing the best fitting state for the current event. This integer is then inserted into the same type of Bayesian averaging as described in the previous paragraph, and the rounded value of the averaged number is taken to indicate the current network configuration of the brain with the value 1 representing integration and the value 2 representing segregation. Together, this causes a hysteresis loop wherein switching from one configuration necessitates both comparison modules to yield evidence for the other configuration over multiple events before a switch occurs, dependent on the parameter values chosen for the Bayesian averaging (for an example see Supplementary Figure 1). In addition, depending on the parameters chosen for the comparison, we can bias the model so that maintaining segregation becomes a metastable state reflecting the increased cognitive cost of holding two predictive models.

Implementation

To exemplify how MEAMSO can be used on behavioural data, we implemented a minimal version of the model in MATLAB (code is available at github.com/OleAd/MEAMSO). Here we first created an example joint finger tapping trial exhibiting a transition in synchronization strategy, by combining a section of mutual adaptation tapping with a section of leading-following tapping from a previous study as illustrated in Figure 2 [12]. To show the change in synchronization strategy, we calculated cross-correlations in non-overlapping windows of 15 taps (see Figure 2). We then ran the behavioural data through the MEAMSO, both from the perspective of dyad member 1 and dyad member 2 (Figure 2). For dyad member 1 we see that the brain state output from the model starts at state 2 (segregation of self and other), before quickly transitioning to state 1 (integration of self and other). It remains in state 1 until a few taps into the leading-following section, wherein dyad member 1 becomes the leading participant. It then switches to brain state 2, indicating that the perceptions are no longer causally integrated with the actions. Towards the end of the trial, it rapidly switches, before settling back to state 1. For dyad member 2, the brain state output starts in state 2, and rapidly settles in state 1 throughout the trial, indicating integration of self and other. As dyad member 2 is the follower in this interaction, this is the expected behaviour of the model.

Application of the MEAMSO on tapping data

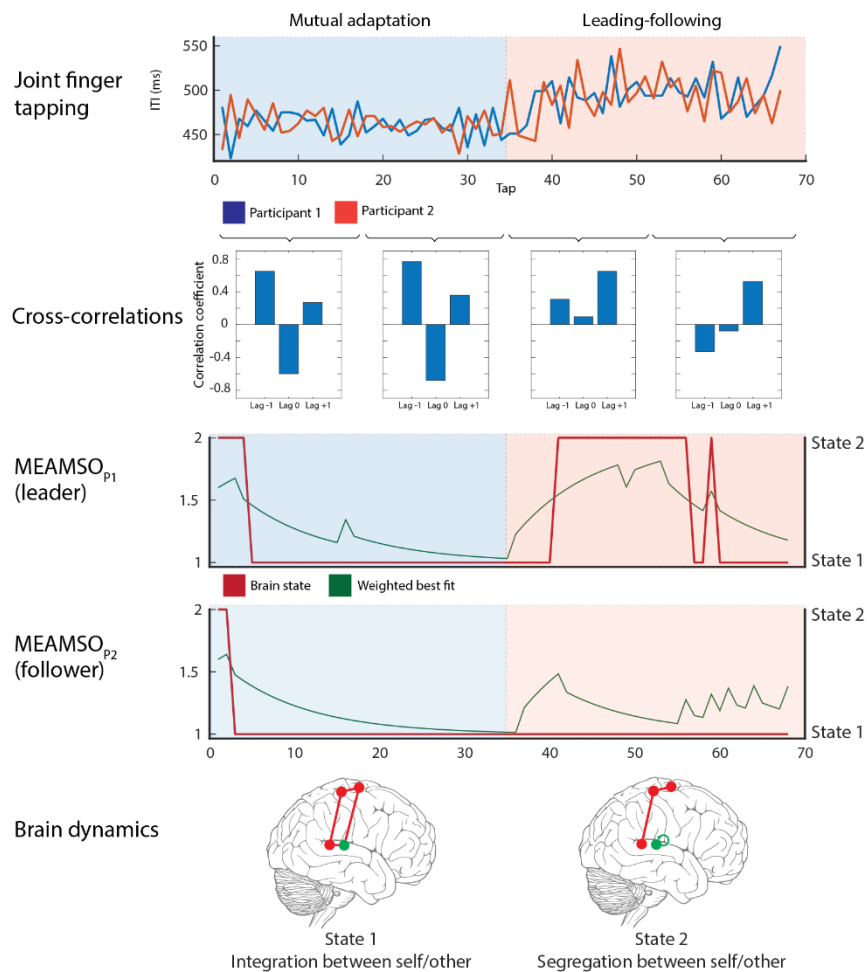


Figure 2 – Application of the MEAMSO on tapping data. Here we create an example interaction by combining joint finger tapping data from a mutual adaptation and leading-following interaction. By looking at cross-correlation coefficients we can see how the interaction shifts. We run the MEAMSO from both participant’s perspective, and plot the weighted best fit and the outcome brain state from the model. The model is initiated in state 2 (segregation of self/other). For dyad member 1 (P1), the leader, we can see the state switch shortly after the transition to leading-following, indicating a separation between self-performed actions and the sounds perceived from the other. For dyad member 2 (P2), the follower, we see that the state rapidly switch to state 1 (integration of self/other), and remains there throughout the interaction. In the last panel we illustrate the brain dynamics associated with the two states.

What does MEAMSO explain, and what does it predict?

The main purpose of our conceptual model is to collate, develop, test, and extend theory [73]. With MEAMSO we show how synchronization strategies can be framed from a cognitive viewpoint, and in the following paragraphs we describe how the MEAMSO explains the emergence of dyadic synchronization strategies, and highlight some key predictions from the model.

Dyadic synchronization strategies

The MEAMSO posits that there exist two metastable attractors in terms of brain dynamics during dyadic interpersonal synchronization, with one state indicative of self-other integration (maintaining the same predictive model for action and perception), and the other indicative of self-other segregation (maintaining two separate predictive models for action and perception). We can calculate the number of possible unique group combinations of these states by using the formula $states = \frac{(n+s-1)!}{n!(s-1)!}$, where n is the number of participants in a group, and s is the number of individual states. In our case, where $s = 2$, the formula reduces to $states = n + 1$, for $n > 1$. When considering a dyad, there then exist three possible combinations, which map onto the three known synchronization strategies in dyadic interaction: leading-leading, leading-following, and mutual adaptation (illustrated in Figure 3). Leading-leading may occur when both participants separate self from the other, and hence do not use information from the other to influence their own actions. Leading-following necessitates that the leader separates the other's actions from their own actions, whereas the follower integrates (by adapting their tapping based on the preceding tap from the leader). In mutual adaptation both participants integrate self and other, so that each other's actions reciprocally influence upcoming actions.

The emergence and selection of synchronization strategies

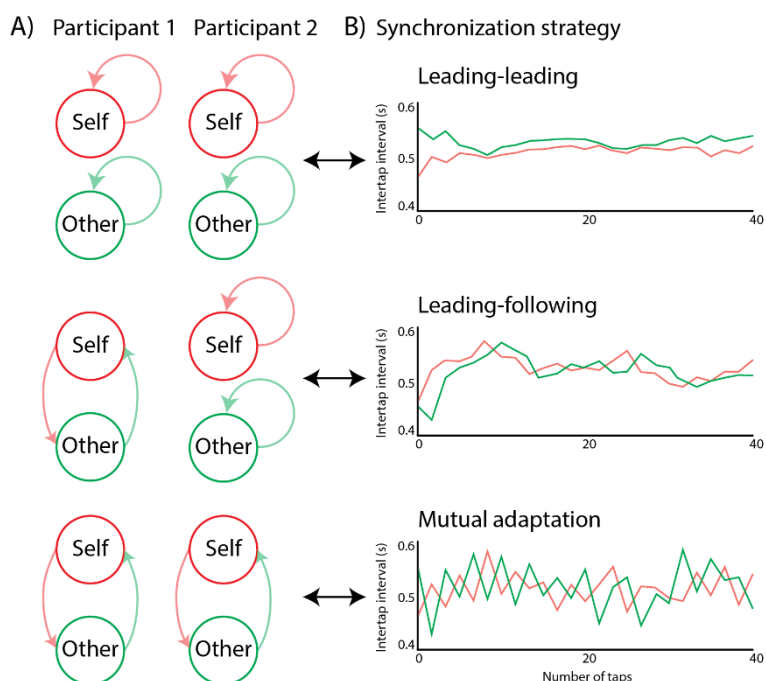


Figure 3 – The emergence and selection of synchronization strategies. In A we show the the configuration of self-other integration/segregation for pairs of participants, and in B the corresponding synchronization strategy. We propose that leading-leading occurs when both dyad members hold two separate predictive models through self-other segregation. In

leading-following, one dyad member integrates and the other separates. In mutual adaptation both dyad members integrate, which when viewing the dyad as a whole collapses the dyad members' predictive models into one shared model.

Group synchronization dynamics

Following the equation for the number of states listed above, MEAMSO would predict that the number of possible synchronization strategies is always $n + 1$ where n is the number of interacting people. As our model is based on dyadic interaction it will likely not capture all the intricacies present in group synchronization dynamics. In particular, as a group of interacting persons reach intermediate size such as in the Human Firefly experiment, rich and diverse synchronization dynamics emerges [25, 74]. In addition, the topology of feedback influences the group's synchronization behaviour [75, 76]. Given our model's predictions for dyadic synchronization strategies, we would hypothesize that as the size of the group grows so does the instability of the strategies, meaning that there may be fewer viable and stable synchronization strategies for large groups. There are two main factors leading us to this conclusion. First, a leader may have many followers, but a follower can only follow one leader. This means that certain constellations (e.g., two leaders and one follower), are unlikely to produce a stable synchronized interaction. Second, there is likely a limit to the size of a group when individuals can no longer be individually monitored, and instead a measure of the group's performance as a whole, or subgroups within the group, need to be monitored [77]. This limit is likely dependent on the individuals' ability to perform auditory stream segregation, a complex ability that depends on multiple characteristics of the sounds being heard [78-80]. In addition, a recent study with 16 violinists performing a simple melody together, perturbed by the introduction of auditory delays between the violinists, showed that the musicians were able to rapidly ignore conflicting signals [76]. In our model, one way of incorporating this ability would be to add an attention weight to signals in groups of more than 2, such that a signal that is kept segregated over a certain period of time is discarded from the computation until changes in the signal necessitate a re-evaluation. One way to explore this would be to design experiments wherein experienced groups of musicians perform joint finger tapping with designated roles (e.g. leader and follower) to evaluate the stability of interactions, and to investigate groups of musicians and non-musicians to evaluate how signals are weighted in a group interaction.

Synchronization strategies are dynamic

MEAMSO predicts that synchronization strategies are dynamic, yet that they are transiently stable. By this we mean that the strategy may change during an interaction, for instance switching between a leading-following strategy to a mutual adaptation strategy. This would occur if there is a sudden change in comparison calculations indicating that the current state is no longer supported by perceptual evidence. Yet, due to the hysteresis inherent in the MEAMSO calculations, such change would need to surpass a critical threshold before a re-evaluation of the best fitting state is performed.

In addition, if the brain tends towards computational efficiency, one would predict mutual adaptation to be more frequent and stable than the other synchronization strategies. This is due to both interacting brains only needing to maintain one predictive model of the task, which has been proposed to be more computationally efficient [2].

To empirically test this prediction, one could design a paradigm wherein an adaptive virtual tapping partner gradually introduces perturbations in the interaction. Here we would predict that small, isolated, perturbations would not disrupt a mutual adaptation interaction, but larger and sustained perturbations would switch the interaction to a leading-following synchronization strategy. The inherent metastability in the MEAMSO would also point towards it being easier to go from leading-following to mutual adaptation, than the other way around. Here there is also an opportunity for incorporating the MEAMSO with our Kuramoto-based coupled oscillators model [24]. In the Kuramoto-based model the coupling strengths that govern the behaviour of the model were fixed throughout the interaction. By using the MEAMSO to dynamically update the coupling strengths throughout a simulated tapping trial, we can formalize predictions on the stability of synchronization strategies.

Similar sounds are more likely to lead to self-other integration

The instantaneous comparison between auditory features is a technical and cognitively complex component of the MEAMSO, whose inner workings we have not explored in detail in this paper. Nonetheless, the ability to match an action with a predicted auditory perception is an interesting aspect of the model. Here, the model predicts that a precise predictive model of action-related auditory feedback will increase the likelihood of self-other segregation. This may entail that dyads of musicians, who through training develop fine-tuned predictive models due to their instrument-specific expertise [81-83], are more likely than the general public to exhibit leading-following and leading-leading behaviour. Another prediction would be that self-other integration should be increased in interactions where participants are unsure about the auditory feedback from their actions. A possible example of this can be found in multiple musical settings, such as ritual music and group singing, which promote group cohesion and facilitate prosocial behaviour [84, 85]. Here, the auditory output produced by individuals is often highly similar or embedded in the group's auditory output. We would predict this contributes to self-other integration, and perhaps to the sense of belonging and meaningfulness sometimes attributed to group singing [86].

MEAMSO in the brain

The output from MEAMSO is a state indicating whether self-other integration or segregation is currently taking place, and should be identifiable on a neural level. One of the proposed neural

mechanisms involved in self-other integration is alignment of neural self/other representations [2]. This view finds support in recent research on interacting mice wherein populations of neurons in the dorsomedial prefrontal cortex selectively encode self-related and other-related behaviour [87]. Notably, this activity may underlie the interbrain synchronization reported between interacting people. Nonetheless, human social interaction inherently relies on many complex cognitive tasks, as evident by the diverse brain regions involved in social tasks [44, 60]. In a controlled experimental setting, such as with a joint finger tapping task, some of this variability is constrained by placing limits on the actions performed and on the perceptual feedback (often just simple transient sounds are used) [10, 12, 14]. Hence, tightly controlled joint action paradigms offer a promising avenue towards identifying the neural correlates of self-other integration and segregation.

In previous work we used a data-driven approach to identify a functional brain network selectively associated with mutual adaptation and leading-leading synchronization strategies [88]. Mutually adapting participants exhibited a higher occurrence of synchronized activity within this network, perhaps representing the proposed synchronization between self- and other-representations in the brain. Information flow within this network, measured with directed phase transfer entropy, indicated a key role of the right precuneus and its connections to the supramarginal gyrus, auditory cortex, and temporoparietal junction. As such, coherent activity within this network and between its key regions may be a candidate for a neural correlate of the MEAMSO's output, an avenue we intend to pursue in future work.

Summary

In this paper, we have introduced the MEAMSO. This event-based model proposes an interconnected set of mechanisms describing how self-other integration occurs in dyadic interaction, and how this process gives rise to metastable synchronization strategies. We use a minimal implementation of the MEAMSO to illustrate how the model could capture the dyadic imbalance in a transition between mutual adaptation and leading-following. The MEAMSO is therefore a model which can encompass dynamic shifts in synchronization strategies over time. Furthermore, MEAMSO can be used to formulate key predictions on the dynamics of rhythmic interpersonal synchronization and to point to neural mechanisms underlying interpersonal synchronization. It is our hope that the MEAMSO will prove useful in approximating the complexity of rhythmic interpersonal synchronization.

References

1. Clark H.H. 2006 Social actions, social commitments. *Roots of human sociality: Culture, cognition and interaction*, 126-150.

2. Koban L., Ramamoorthy A., Konvalinka I. 2017 Why do we fall into sync with others? Interpersonal synchronization and the brain's optimization principle. *Soc Neurosci*, 1-9.
3. van Ulzen N.R., Lamoth C.J., Daffertshofer A., Semin G.R., Beek P.J. 2008 Characteristics of instructed and uninstructed interpersonal coordination while walking side-by-side. *Neurosci Lett* **432**(2), 88-93.
4. Néda Z., Ravasz E., Brechet Y., Vicsek T., Barabási A.-L. 2000 The sound of many hands clapping. *Nature* **403**(6772), 849-850.
5. Peltokorpi J., Niemi E. 2018 Differences between worker pairs in manual assembly: a case study. *Procedia Manufacturing* **25**, 535-542.
6. Repp B.H. 2006 Musical synchronization. *Music, motor control, and the brain*, 55-76.
7. Keller P.E., Koch I. 2008 Action planning in sequential skills: Relations to music performance. *Q J Exp Psychol* **61**(2), 275-291.
8. Van Der Steen M.C., Keller P.E. 2013 The ADaptation and Anticipation Model (ADAM) of sensorimotor synchronization. *Frontiers in human neuroscience* **7**, 253.
9. Koelsch S., Vuust P., Friston K. 2019 Predictive processes and the peculiar case of music. *Trends in Cognitive Sciences* **23**(1), 63-77.
10. Heggli O.A., Konvalinka I., Kringelbach M.L., Vuust P. 2019 Musical interaction is influenced by underlying predictive models and musical expertise. *Scientific Reports* **9**(1), 11048.
11. Loehr J.D., Kourtis D., Vesper C., Sebanz N., Knoblich G. 2013 Monitoring individual and joint action outcomes in duet music performance. *Journal of cognitive neuroscience* **25**(7), 1049-1061.
12. Konvalinka I., Vuust P., Roepstorff A., Frith C.D. 2010 Follow you, follow me: continuous mutual prediction and adaptation in joint tapping. *The Quarterly journal of experimental psychology* **63**(11), 2220-2230.
13. Konvalinka I., Bauer M., Stahlhut C., Hansen L.K., Roepstorff A., Frith C.D. 2014 Frontal alpha oscillations distinguish leaders from followers: multivariate decoding of mutually interacting brains. *Neuroimage* **94**, 79-88.
14. Gebauer L., Witek M., Hansen N., Thomas J., Konvalinka I., Vuust P. 2016 Oxytocin improves synchronisation in leader-follower interaction. *Scientific reports* **6**, 38416.
15. Reidsma D., Nijholt A., Bos P. 2008 Temporal interaction between an artificial orchestra conductor and human musicians. *Computers in Entertainment (CIE)* **6**(4), 1-22.
16. Zamm A., Wellman C., Palmer C. 2016 Endogenous rhythms influence interpersonal synchrony. *J Exp Psychol Hum Percept Perform* **42**(5), 611.
17. Zamm A., Wang Y., Palmer C. 2018 Musicians' natural frequencies of performance display optimal temporal stability. *J Biol Rhythms* **33**(4), 432-440.
18. Repp B.H. 1998 The detectability of local deviations from a typical expressive timing pattern. *Music Perception* **15**(3), 265-289.
19. Rankin S.K., Limb C.J. 2014 Auditory-motor synchronization with temporally fluctuating sequences is dependent on fractal structure but not musical expertise. *Front Psychol* **5**, 970.
20. Loehr J.D., Large E.W., Palmer C. 2011 Temporal coordination and adaptation to rate change in music performance. *J Exp Psychol Hum Percept Perform* **37**(4), 1292.
21. Rankin S.K., Large E.W., Fink P.W. 2009 Fractal tempo fluctuation and pulse prediction. *Music perception* **26**(5), 401-413.
22. Endedijk H.M., Ramenzoni V.C., Cox R.F., Cillessen A.H., Bekkering H., Hunnius S. 2015 Development of interpersonal coordination between peers during a drumming task. *Dev Psychol* **51**(5), 714.
23. Steinbeis N. 2016 The role of self-other distinction in understanding others' mental and emotional states: neurocognitive mechanisms in children and adults. *Philosophical Transactions of the Royal Society B: Biological Sciences* **371**(1686), 20150074.
24. Heggli O.A., Cabral J., Konvalinka I., Vuust P., Kringelbach M.L. 2019 A Kuramoto model of self-other integration across interpersonal synchronization strategies. *PLoS computational biology* **15**(10). (doi:<https://doi.org/10.1371/journal.pcbi.1007422>).

25. Zhang M., Beetle C., Kelso J.S., Tognoli E. 2019 Connecting empirical phenomena and theoretical models of biological coordination across scales. *Journal of the Royal Society Interface* **16**(157), 20190360.
26. Haken H., Kelso J.S., Bunz H. 1985 A theoretical model of phase transitions in human hand movements. *Biol Cybern* **51**(5), 347-356.
27. Fuchs A., Jirsa V.K., Haken H., Kelso J.S. 1996 Extending the HKB model of coordinated movement to oscillators with different eigenfrequencies. *Biol Cybern* **74**(1), 21-30.
28. Konvalinka I., Vuust P., Roepstorff A., Frith C.D. 2009 A coupled oscillator model of interactive tapping. In *ESCOM 2009: 7th Triennial Conference of European Society for the Cognitive Sciences of Music* (Jyväskylä, Finland, University of Jyväskylä).
29. Large E.W., Palmer C. 2002 Perceiving temporal regularity in music. *Cogn Sci* **26**(1), 1-37.
30. Roman I.R., Washburn A., Large E.W., Chafe C., Fujioka T. 2019 Delayed feedback embedded in perception-action coordination cycles results in anticipation behavior during synchronized rhythmic action: A dynamical systems approach. *PLoS computational biology* **15**(10), e1007371.
31. Harry B., Keller P.E. 2019 Tutorial and simulations with ADAM: An adaptation and anticipation model of sensorimotor synchronization. *Biol Cybern* **113**(4), 397-421.
32. Stepp N., Turvey M.T. 2010 On strong anticipation. *Cogn Syst Res* **11**(2), 148-164.
33. Stepp N., Turvey M.T. 2008 Anticipating synchronization as an alternative to the internal model. *Behavioral and Brain Sciences* **31**(2), 216.
34. Farrer C., Frith C.D. 2002 Experiencing oneself vs another person as being the cause of an action: the neural correlates of the experience of agency. *Neuroimage* **15**(3), 596-603.
35. Blakemore S.-J., Frith C. 2003 Self-awareness and action. *Current opinion in neurobiology* **13**(2), 219-224.
36. Van der Meer L., Groenewold N.A., Nolen W.A., Pijnenborg M., Aleman A. 2011 Inhibit yourself and understand the other: neural basis of distinct processes underlying Theory of Mind. *Neuroimage* **56**(4), 2364-2374.
37. Novembre G., Sammler D., Keller P.E. 2016 Neural alpha oscillations index the balance between self-other integration and segregation in real-time joint action. *Neuropsychologia* **89**, 414-425.
38. Huberth M., Dauer T., Nanou C., Román I., Gang N., Reid W., Wright M., Fujioka T. 2018 Performance monitoring of self and other in a turn-taking piano duet: A dual-EEG study. *Soc Neurosci*, 1-13.
39. Varlet M., Nozaradan S., Nijhuis P., Keller P.E. 2020 Neural tracking and integration of 'self' and 'other' in improvised interpersonal coordination. *Neuroimage* **206**, 116303.
40. Dumas G., Moreau Q., Tognoli E., Kelso J.S. 2020 The Human Dynamic Clamp reveals the fronto-parietal network linking real-time social coordination and cognition. *Cereb Cortex* **30**(5), 3271-3285.
41. Mitchell J.P., Banaji M.R., MacRae C.N. 2005 The link between social cognition and self-referential thought in the medial prefrontal cortex. *Journal of cognitive neuroscience* **17**(8), 1306-1315.
42. Weed E., McGregor W., Nielsen J.F., Roepstorff A., Frith U. 2010 Theory of Mind in adults with right hemisphere damage: What's the story? *Brain Lang* **113**(2), 65-72.
43. Hasson U., Frith C.D. 2016 Mirroring and beyond: coupled dynamics as a generalized framework for modelling social interactions. *Philosophical Transactions of the Royal Society B: Biological Sciences* **371**(1693), 20150366.
44. Frith C.D., Frith U. 1999 Interacting minds--a biological basis. *Science* **286**(5445), 1692-1695.
45. Friston K., Frith C. 2015 A duet for one. *Conscious Cogn* **36**, 390-405.
46. Deco G., Kringelbach M.L. 2016 Metastability and coherence: extending the communication through coherence hypothesis using a whole-brain computational perspective. *Trends Neurosci* **39**(3), 125-135.
47. Tognoli E., Kelso J.S. 2014 The metastable brain. *Neuron* **81**(1), 35-48.

48. Kelso J.S., Tognoli E. 2009 Toward a complementary neuroscience: metastable coordination dynamics of the brain. In *Downward causation and the neurobiology of free will* (pp. 103-124, Springer).
49. Kelso J.S. 2012 Multistability and metastability: understanding dynamic coordination in the brain. *Philosophical Transactions of the Royal Society B: Biological Sciences* **367**(1591), 906-918.
50. Tognoli E., Kelso J.S. 2009 Brain coordination dynamics: true and false faces of phase synchrony and metastability. *Prog Neurobiol* **87**(1), 31-40.
51. Fourneret P., Jeannerod M. 1998 Limited conscious monitoring of motor performance in normal subjects. *Neuropsychologia* **36**(11), 1133-1140.
52. Daprati E., Franck N., Georgieff N., Proust J., Pacherie E., Dalery J., Jeannerod M. 1997 Looking for the agent: an investigation into consciousness of action and self-consciousness in schizophrenic patients. *Cognition* **65**(1), 71-86.
53. Daprati E., Sirigu A. 2002 Laterality effects on motor awareness. *Neuropsychologia* **40**(8), 1379-1386.
54. Friston K., Mattout J., Kilner J. 2011 Action understanding and active inference. *Biol Cybern* **104**(1-2), 137-160.
55. Brown H., Adams R.A., Parees I., Edwards M., Friston K. 2013 Active inference, sensory attenuation and illusions. *Cogn Process* **14**(4), 411-427.
56. Jones M.R., Boltz M. 1989 Dynamic attending and responses to time. *Psychol Rev* **96**(3), 459.
57. Eargle J., Foreman C. 2002 *Audio engineering for sound reinforcement*, Hal Leonard Corporation.
58. Chafe C., Gurevich M., Leslie G., Tyan S. 2004 Effect of time delay on ensemble accuracy. In *Proceedings of the international symposium on musical acoustics* (p. 46, ISMA Nara).
59. Lester M., Boley J. 2007 The effects of latency on live sound monitoring. In *Audio Engineering Society Convention 123* (Audio Engineering Society).
60. Frith C.D. 2007 The social brain? *Philosophical Transactions of the Royal Society of London B: Biological Sciences* **362**(1480), 671-678.
61. Dennett D.C. 1989 *The intentional stance*, MIT Press.
62. Loehr J.D., Vesper C. 2016 The sound of you and me: novices represent shared goals in joint action. *Q J Exp Psychol* **69**(3), 535-547.
63. Sorati M., Behne D.M. 2020 Audiovisual Modulation in Music Perception for Musicians and Non-musicians. *Front Psychol* **11**(1094). (doi:10.3389/fpsyg.2020.01094).
64. Elliott T.M., Hamilton L.S., Theunissen F.E. 2013 Acoustic structure of the five perceptual dimensions of timbre in orchestral instrument tones. *The Journal of the Acoustical Society of America* **133**(1), 389-404.
65. Sperling G., Spelman R.G. 1970 Acoustic similarity and auditory short-term memory: Experiments and a model. *Models of human memory*, 151-202.
66. Bregman A.S. 1994 *Auditory scene analysis: The perceptual organization of sound*, MIT press.
67. Almonte F., Jirsa V.K., Large E.W., Tuller B. 2005 Integration and segregation in auditory streaming. *Physica D: Nonlinear Phenomena* **212**(1-2), 137-159.
68. Sato A. 2008 Action observation modulates auditory perception of the consequence of others' actions. *Conscious Cogn* **17**(4), 1219-1227.
69. Bolt N.K., Poncelet E.M., Schultz B.G., Loehr J.D. 2016 Mutual coordination strengthens the sense of joint agency in cooperative joint action. *Conscious Cogn* **46**, 173-187.
70. Yang X., Zhang Z. 2013 Combining prestige and relevance ranking for personalized recommendation. In *Proceedings of the 22nd ACM international conference on Information & Knowledge Management* (pp. 1877-1880).
71. Witek M.A., Clarke E.F., Kringelbach M.L., Vuust P. 2014 Effects of polyphonic context, instrumentation, and metrical location on syncopation in music. *Music Perception: An Interdisciplinary Journal* **32**(2), 201-217.

72. Cannon J.J. 2020 PIPPET: A Bayesian framework for generalized entrainment to stochastic rhythms. *bioRxiv*.
73. Smaldino P. 2020 How to translate a verbal theory into a formal model. *Soc Psychol* **51**(4).
74. Zhang M., Kelso J.S., Tognoli E. 2018 Critical diversity: Divided or united states of social coordination. *PLoS One* **13**(4), e0193843.
75. Alderisio F., Fiore G., Salesse R.N., Bardy B.G., di Bernardo M. 2017 Interaction patterns and individual dynamics shape the way we move in synchrony. *Scientific reports* **7**(1), 1-10.
76. Shahal S., Wurzburg A., Sibony I., Duadi H., Shniderman E., Weymouth D., Davidson N., Fridman M. 2020 Synchronization of complex human networks. *Nature communications* **11**(1), 1-10.
77. Wing A.M., Endo S., Bradbury A., Vorberg D. 2014 Optimal feedback correction in string quartet synchronization. *Journal of The Royal Society Interface* **11**(93), 20131125.
78. Deutsch D. 1999 Grouping mechanisms in music. In *The psychology of music* (pp. 299-348, Elsevier).
79. Rankin J., Rinzel J. 2019 Computational models of auditory perception from feature extraction to stream segregation and behavior. *Current opinion in neurobiology* **58**, 46-53.
80. Rankin J., Sussman E., Rinzel J. 2015 Neuromechanistic model of auditory bistability. *PLoS computational biology* **11**(11), e1004555.
81. Vuust P., Pallesen K.J., Bailey C., Van Zuijen T.L., Gjedde A., Roepstorff A., Østergaard L. 2005 To musicians, the message is in the meter: pre-attentive neuronal responses to incongruent rhythm are left-lateralized in musicians. *Neuroimage* **24**(2), 560-564.
82. Vuust P., Brattico E., Seppänen M., Näätänen R., Tervaniemi M. 2012 The sound of music: differentiating musicians using a fast, musical multi-feature mismatch negativity paradigm. *Neuropsychologia* **50**(7), 1432-1443.
83. Koelsch S., Vuust P., Friston K. 2018 Predictive processes and the peculiar case of music. *Trends in Cognitive Sciences*.
84. Kirschner S., Tomasello M. 2010 Joint music making promotes prosocial behavior in 4-year-old children. *Evolution and Human Behavior* **31**(5), 354-364.
85. Cirelli L.K. 2018 How interpersonal synchrony facilitates early prosocial behavior. *Current opinion in psychology* **20**, 35-39.
86. Stewart N.A.J., Lonsdale A.J. 2016 It's better together: The psychological benefits of singing in a choir. *Psychology of Music* **44**(6), 1240-1254.
87. Kingsbury L., Huang S., Wang J., Gu K., Golshani P., Wu Y.E., Hong W. 2019 Correlated neural activity and encoding of behavior across brains of socially interacting animals. *Cell* **178**(2), 429-446. e416.
88. Heggli O.A., Konvalinka I., Cabral J., Brattico E., Kringelbach M.L., Vuust P. 2021 Transient brain network underlying interpersonal strategies during synchronized action. *Social Cognitive and Affective Neuroscience* **16**(1-2), 11. (doi:<https://doi.org/10.1093/scan/nsaa056>).