Measuring the Causal Dynamics of Facial Interaction with Convergent Cross Mapping

Eric Postma (e.o.postma@uvt.nl)
Tilburg center for Cognition and Communication
Tilburg, The Netherlands

Marie Postma-Nilsenová (m.nilsenova@uvt.nl)
Tilburg center for Cognition and Communication
Tilburg, The Netherlands

Abstract

The nature of the dynamics of nonverbal interactions is of considerable interest to the study of human communication and future human-computer interaction. Facial expressions constitute an important source of nonverbal social signals. Whereas most studies have focused on the facial expressions of isolated individuals, the aim of this study is to explore the coupling dynamics of facial expressions in social dyadic interactions. Using a special experimental set-up, the frontal facial dynamics of pairs of socially interacting persons were measured and analyzed simultaneously. We introduce the use of convergent cross mapping, a method originating from dynamical systems theory, to assess the causal coupling of the dyadic facial-expression dynamics. The results reveal the presence of bidirectional causal couplings of the facial dynamics. We conclude that convergent cross mapping yields encouraging results in establishing evidence for causal behavioral interactions.
Keywords: mimicry; convergent cross mapping; facial expressions

Introduction

In recent years, the development of automatic coding algorithms boosted the study of facial expressions (Littlewort et al., 2011). By automatically tracking the facial dynamics in response to experimental conditions, the nonverbal facial responses to a large variety of situations can be assessed. Human facial expressions constitute a complex dynamical system formed by the facial muscles that are controlled by brain dynamics which, in turn, are governed by endogenous and exogenous sources. Facial expressions have an important signalling function. By means of expressions, nonverbal information is transmitted to social partners and provides an important context to the verbal information (Smith, Cottrell, Gosselin, & Schyns, 2005). In successful dyadic interactions, the facial expressions of interlocutors may be expected to be coupled. For instance, in mother-child interactions, the smile of the mother may invoke a smiling response from the child (Cohn, 2010), which, in turn, reinforces the smile of the mother. In a behavioral context, such a transient positive feedback loop is associated with behavioral synchrony or mimicry, the tendency of interacting humans to synchronize or mimic their behaviors (Louwerse, Dale, Bard, & Jeuniaux, 2012). In a mathematical context, such feedback loops can be modeled in terms of a coupled dynamical system.

The aim of this paper is to study behavioral mimicry by means of dynamical system theory. More specifically, the goal is to determine if dynamical system theory can be used successfully to measure the presence and direction of causality in the interaction dynamics of dyadic facial expressions. Detecting synchrony or mimicry in behavioral interactions is a challenge. Simply measuring the correlation between two dynamic time-series representing non-verbal behaviors of the two interlocutors (e.g., their smiling intensities) falls short for two reasons: (i) there may be causation without correlation, and (ii) there may be correlation without causation. An example of causation without correlation is a case in which the causing effect equally often induces a positive result and a negative result. An example illustrating correlation without causation is the presence of an external effect that induces synchronous behavior. For example, a sudden flash of light induces the synchronized closing of the eyes of both conversation partners.

Many methods have been proposed to measure synchrony and mimicry in behavioral interactions, for instance: windowed cross correlation (Boker, Xu, Rotondo, & King, 2002), frame differencing co-occurrence (Paxton & Dale, 2012), Granger causality (Kalimeri, Lepri, Kim, Pianesi, & Pentland, 2011), and cross recurrence quantification (Shockley, 2005). For a more complete overview, the interested reader is referred to a relatively recent survey of methods (Delaherche et al., 2012). The main limitation of these methods is that they fail to measure causality in an appropriate manner. Windowed cross correlation suffers from the aforementioned limitations of correlation. Frame differencing co-occurrence only considers the simultaneity of time-specific derivatives of pixel intensity associated with the visually discernible behaviors of two persons. Granger causality has its merits for linear dynamics, but fails to account for the prevalent nonlinear dynamics of human behavior. Finally, cross recurrence quantification has a strong foundation in dynamic systems theory and provides many measures for characterizing the interaction dynamics. However, in itself it is not suitable to establish a causal coupling of temporal sequences. The only way to determine causality is to manipulate the alleged causing stimulus and measure its effect (Richardson & Dale, 2005).

A relatively recent method called Convergent Cross Mapping (CCM) (Sugihara et al., 2012) originates from the domain of complex system dynamics and allows for measuring causal couplings. In contrast to cross recurrence quantification, it provides means for the detection of causal relations between interacting nonverbal signals without suffering from the two shortcomings of straightforward correlation.

Detecting Causality

In the context of complex dynamical systems, the dynamics of two interacting systems is represented by an *attractor manifold*: i.e. the admissible states within the state space spanned by the two (possibly multidimensional) variables. In what follows, we briefly explain Convergent Cross Mapping in terms of attractor manifolds.

Convergent Cross Mapping

When N interacting dynamical variables describe a single attractor manifold in N-dimensional phase space (i.e., a coupled system of N differential equations), the temporal development of each individual variable contains information about the influence of the other $N-1$ variables. This may be compared to a musical ensemble in which the contributions of each of the constituent members depend on those of all the other members. The realization that each variable in a coupled dynamical system contains information about all other variables, led Flores Takens (Takens, 1981) to propose the reconstruction of an N-dimensional attractor by mapping a single variable against one or more time-delayed versions of itself. He showed that the resulting "shadow attractor" is highly similar ("diffeomorphic") to the original N-dimensional attractor manifold. We illustrate these concepts by means of the standard example: the $N=3$ dimensional Lorenz attractor defined by a system of three coupled differential equations:

$$\begin{aligned} dX/dt &= 10(Y-X), \\ dY/dt &= X(28-Z)-Y, \\ dZ/dt &= XY-8Z/3. \end{aligned} \tag{1}$$

The three variables X, Y, and Z, describe the time-varying system state, with t representing time. As can be seen in these equations, the dynamics of X depends on Y (and on X itself), those of Y on X and Z, and those of Z on X and Y. The top part of Figure 1 shows the attractor manifold M of the famous Lorenz system. The middle and the bottom part illustrates the shadow attractors M_X and M_Y formed by plotting $X(t)$ against its time-delayed versions $X(t+\tau)$ and $X(t+2\tau)$ and $Y(t)$ against its time-delayed versions $Y(t+\tau)$ and $Y(t+2\tau)$. By comparing the two shadow attractors, their similarities become apparent. These similarities reflect the fact that the two variables X and Y (and Z) are (by definition) causally related.

Convergent Cross Mapping exploits the same principle in the opposite direction. If two dynamical variables A and B are coupled, than if A has a causal relation to B, it is possible to predict the shadow attractor of A from the shadow attractor of B, and vice versa. In practice, the CCM algorithm measures the causal coupling between variables X and Y by selecting a point of shadow attractor M_X together with a number of its nearest neighbors. These are used to predict

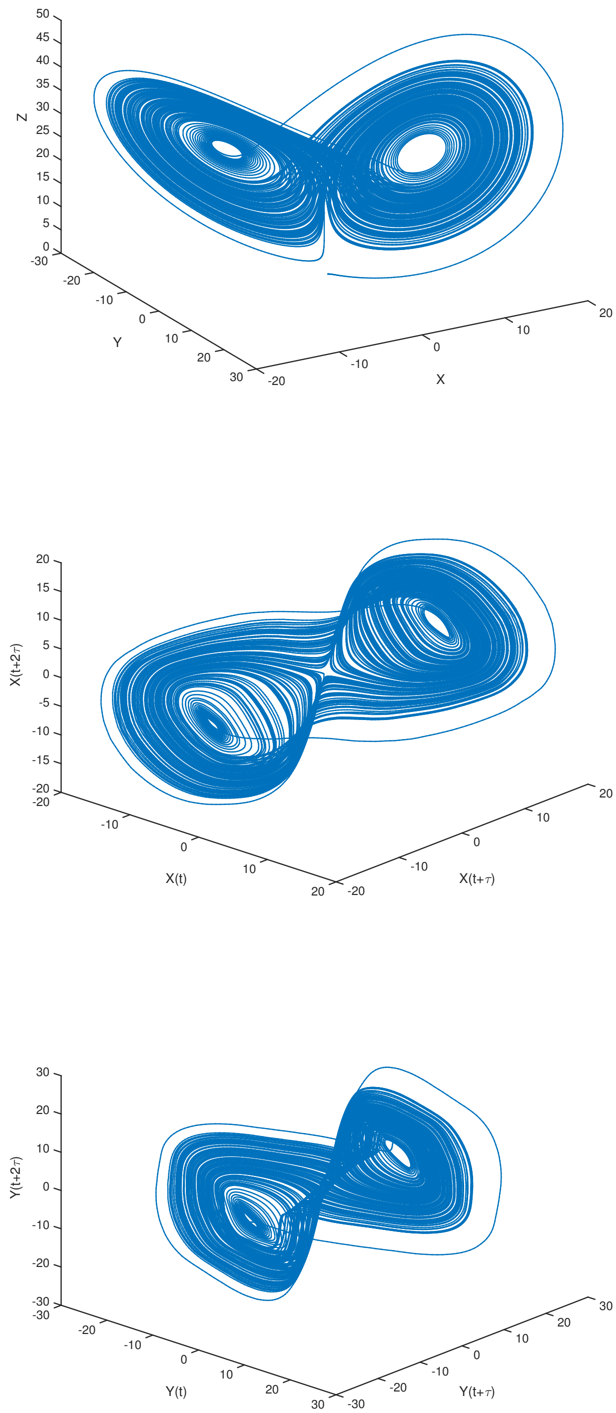

Figure 1: Top: Illustration of the Lorenz attractor manifold M. Middle: Shadow attractor manifold M_X obtained by plotting $X(t)$ against a time-delayed versions of itself: $X(t+\tau)$ and $X(t+2\tau)$ ($\tau = 30$). Bottom: Shadow attractor M_Y obtained for Y. The shadow attractors have a similar shape.

Table 1: Statements rated by the participants on the five-point scale "completely disagree", "disagree", "neutral", "agree", and "completely agree" . Only the ratings of statement 1 are used in the present study.

#	Statement
1	The conversation was good
2	I liked my conversation partner
3	My conversation partner seemed to like me
4	I listened carefully to my conversation partner
5	My conversation partner listened carefully to me

the temporally coupled points on shadow attractor M_Y. If M_X and M_Y are causally coupled, as is the case in the Lorenz attractor, increasing the number of nearest neighboring points in M_X leads to improved predictions of the points in M_Y. The improvement in prediction quality as a function of the number of points is a sign of an unidirectional causal coupling. The prediction quality is measured by means of a correlation value ρ_{CCM} of the actual and predicted values. Hence, the admissible values of ρ_{CCM} are confined to the unit interval.

Experimental Method

We applied CCM to measure causal facial interaction dynamics. The data was collected in an experiment studying mimicry and synchrony in a dyadic conversation. Twelve participants (10 female) were instructed to tell each other about the best and worst events they experienced in their lives. Throughout the session, one participant was assigned the role of the storyteller, whereas the other participant was assigned the role of the listener. After the session, each participant evaluated the quality of the interaction by filling in a brief survey. Participants were asked if they knew their conversation partner, and if so, how often they met. In addition, they rated five statements on a five-point scale ("completely disagree", "disagree", "neutral", "agree", "completely agree"). The five statements are listed in Table 1. The motivation for including these statements was to evaluate the relation between the perceived quality of the conversation and the CCM measurements. In the present study, only the ratings for statement 1 of the survey were used for a post-hoc comparison of the perceived quality of the interaction and the presence and strength of the causal coupling as determined by CCM.

Data Collection

Pairs of participants were each seated in front of an Eye Catcher system in separate rooms. An Eye Catcher system (Ex'ovision) is a recording and display device that by means of an internal see-through mirror, allows for the frontal recording of the face of a person watching the displayed face of his or her conversation partner. Two Eye Catcher systems were coupled in such a way that the pairs of participants could hear and see each other. Simultaneously, both participants were recorded during the session using digital recording software running on two Macbooks (OSX). All videos had a resolution of 640×400 pixels and a frame rate of 30 frames per second. For the current experiment, 6 pairs of sequences were selected for causal analysis. Four pairs involved female-female interactions and the other two mixed gender interactions. For each segment, two fragments of 3000 frames (100 seconds) were selected during which either participant spoke about a positive experience.

Table 2: Action units (AUs) extracted from the video sequences.

AU	Facial Expression
AU1	Inner Brow Raise
AU2	Outer Brow Raise
AU9	Nose Wrinkle
AU12	Lip Corner Pull
AU20	Lip Stretch

Facial Expressions Extraction

The resulting video sequences were processed using the Computer Expression Recognition Toolbox (CERT; Version 5.1, build 1208::867:869M) (Littlewort et al., 2011). CERT outputs for each frame of a facial video sequence estimates of action unit (AU) intensities. Action units are the building blocks of facial expressions as defined by Ekman and Friesen (1976). Each action unit describes the appearance or movement of a local region of the face. An example of an action unit is the "Inner Brow Raise", which is represented by its shorthand "AU1" (action unit 1). The action unit intensity estimates generated by CERT can be negative (evidence for absence) or positive (evidence for presence). The magnitude is proportional to the visual absence or presence of the action unit. Table 2 lists the five action units for which we extracted estimates for further analysis, cf. (Cohn, 2010). Two action units measure facial expressions in the upper facial region (AU1 and AU2), one in the central region (AU9), and two cover the lower facial region (AU12 and AU20).

Dynamical Systems Analysis

Convergent Cross Mapping was applied to paired sequences of the same AU intensities. (We did not examine causal couplings between different pairs of AUs.) We used the R implementation of Convergent Cross Mapping (*rEDM*_0.2.6) combined with Matlab routines for file input and output. CCM has two parameters: the embedding dimension and the time lag. The embedding dimension parameter specifies the number of time-delayed versions of a variable to reconstruct the shadow manifold. For instance, the shadow attractors displayed in Figure 1b and c were reconstructed with an embedding dimension of 3, corresponding to the original and two time-delayed version of X and Y, respectively (i.e., $X(t)$, $X(t+\tau)$, $X(t+2\tau)$ and $Y(t)$, $Y(t+\tau)$, $Y(t+2\tau)$). Its optimal value is highly dependent on the task. We determined the optimal value by varying the embedding dimension from 2 to 10. The time lag parameter defines the time lag separating the predict-

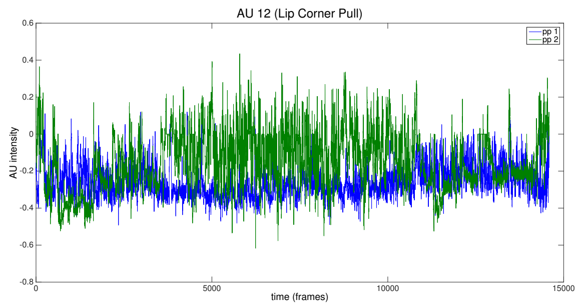

Figure 2: Example of paired time-series of action unit 12 estimates. The two series represent the two interacting participants of dyad 6, pp11 and pp12. The horizontal axis represents time in frames ($\frac{1}{30}$th seconds), the vertical axis represents the AU intensity.

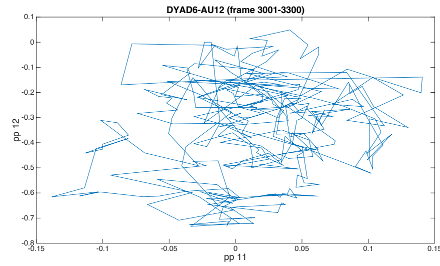

Figure 3: Visualization of the "attractor manifold" obtained by plotting the AU12 intensities of pp11 against those of pp12 for a sample of 300 frames (10 seconds). Adjacent frames are connected by lines.

ing and predicted AU intensity. Because we were interested in multiple time lags ranging from synchrony (time lag = 0) to mimicry (time lag > 0), we examined time lags ranging from 0 (synchrony) to 30 (mimicry) in steps of one. Each time step corresponds to $1/30$-th of a second. Hence, the time lags cover the interval from 0 to 1 second.

Results

Figure 2 shows an example of the AU12 estimates for the two interacting participants of dyad 6 (pp11 and pp12). Whenever a participant is listening, his or her AU12 intensity tends to be larger than that of the speaker. In the figure, participant pp12 is listening to pp11 in the middle time segment.

As an illustration of the interaction dynamics, Figure 3 shows the "attractor manifold" of a 10-second fragment of the AU12 signature of a dyadic interaction between pp11 and pp12. A circular structure represents a period coupling. Clearly, the interaction dynamics of pp11 and pp12 are more noisy and complicated than those of the Lorenz system. This is due to the noisy nature of the CERT action unit estimates and the complicated dynamics of behavioral interactions.

The results of applying the CCM algorithm to the paired time-series of the five action units revealed that only AU12 gave rise to causal coupling. For this action unit, we found that the prediction quality ρ_{CCM} increased as a function of the number of samples L for all dyads and in both directions. This indicates that all dyads exhibited a bidirectional causal coupling of their lip corner pull expression dynamics. The optimal strengths were obtained at time lags varying from 0 (synchrony) to 0.8 seconds (mimicry). The values of the prediction quality $\rho_{CCM}(L_{max})$ at the maximum number of samples examined L_{max} provide an indication of the coupling strengths.

Table 3 lists the strongest causal relations (the maximum values across the five action units examined) found for the 6 dyads examined. For each dyad, the Table lists the composition of each dyad, female-female (F-F), male-female (M-F), or female-male (F-M), the rating of statement 1 of the post-hoc evaluation of the conversation given by the participants ranging from strongly disagree (0) to strongly agree (5), the ratings by the two participants are separated by a /), and a specification of predictive quality $\rho_{CCM}(L_{max})$ obtained for the largest value of L examined, $L_{max} = 300$ in both directions: from the first participant to the second ($\rho_{CCM\rightarrow}$) and vice versa ($\rho_{CCM\leftarrow}$). As listed in the last column, the strongest causal relations were obtained for AU12 (Lip Corner Pull) which is a measure of (subtle) smiling. The results did not depend critically on the embedding dimension, although the best predictive qualities were obtained for the maximal value of the embedding dimension examined (10).

Table 3: CCM results for AU12. The first column specifies the dyad number. The second column indicates the gender composition of the dyad, where the first and second member of the dyad are separated by a hyphen (F = female, M = male). The third column lists the ratings of statement 1 in the survey ("The conversation was good") as a measure of the quality of the conversation (rating 1 represent "strongly disagree" and 5 "strongly agree"), via neutral (3) to very positive (5), the ratings by the two participants are separated by a /). The fourth and fifth columns list the predictive qualities $\rho_{CCM\rightarrow}$) and $\rho_{CCM\leftarrow}$) in both directions.

dyad	gender	evaluation	$\rho_{CCM\rightarrow}$	$\rho_{CCM\leftarrow}$
1	F-F	2/1	0.68	0.43
2	F-F	1/1	0.35	0.42
3	F-F	2/1	0.47	0.59
4	F-F	3/3	0.39	0.54
5	F-M	3/3	0.64	0.74
6	M-F	4/3	0.46	0.56

Figure 4 shows the results of applying Convergent Cross Mapping to paired time series of AU12, see also Table 3. The six plots depict the outcomes for dyads 1 to 6 (top to bottom). Each of the plots shows the prediction qualities ρ_{ccm} as a function of the number of samples (Library Size L) for the two directions of interaction between the two dyad members A and B. Increasing the number of samples L from the shadow manifold M_A for predicting the shadow manifold M_B leads to an increase in the correlation ρ_{ccm} for A $\rightarrow$ B (green curve). Similarly, increasing the number of samples from M_B to predict M_A, increases the correlation ρ_{ccm} for A $\leftarrow$ B (blue curve). The observation that for all dyads both values of ρ exhibit a steady increase with the number of samples, indicates a bidirectional coupling for all dyads. Importantly, such increases are not seen for arbitrary sequences taken from different dyads, i.e., from two persons that did not interact. When pairing the same AU sequences of two persons of different dyads, a small increase is observed up to a maximal value of $\rho_{ccm} \approx 0.2$.

The coupling strengths show some variation both within and between dyads. For our small sample of dyads, there does not seem to be clear relation between the composition of the dyad or the evaluation of the conversation and the prediction strengths (neither for the other survey statements).

Discussion and Conclusions

As stated in the introduction, the goal of the present study was to determine if dynamical system theory can be used to measure the presence and direction of causality in the interaction dynamics of dyadic facial expressions. We employed the method of Convergent Cross Mapping for determining a causal relation between nonverbal behavior in conversation partners. The fact that the method was able to predict the facial expressions of the interacting partners confirms that the presence of a causal coupling in nonverbal behavior can be measured with CCM. Importantly, the prediction improves with an increasing number of samples per conversation partner in a given dyad. In the CCM method, this is considered a true sign of causality.

Previous studies indicated an increased tendency of females as compared to males to exhibit nonverbal mimicry (Dimberg, 1990; Briton & Hall, 1995). LaFrance and colleagues (LaFrance, Hecht, & Paluck, 2003) found a clear gender difference in smiling especially for female dyads as compared to male dyads. Although we did not find an effect of the evaluation of the conversation on the coupling strength, evidence for such an effect has been reported (Chartrand & van Baaren, 2009). Future work aims at increasing the number of dyads submitted to CCM analysis to establish if the effect can be confirmed. Finally, the range of lags at which the evidence for causal coupling is strongest agrees with the range of values reported in the literature (Dimberg & Thunberg, 1998; Dimberg, Thunberg, & Elmehed, 2000; Sato & Yoshikawa, 2007; Achaibou, Pourtois, Schwartz, & Vuilleumier, 2008).

To the best of our knowledge, this is the first time that CCM has been applied to measure the presence, strength, and direction of causal coupling in nonverbal behavioral dynamics. Although our results are encouraging, additional study is re-

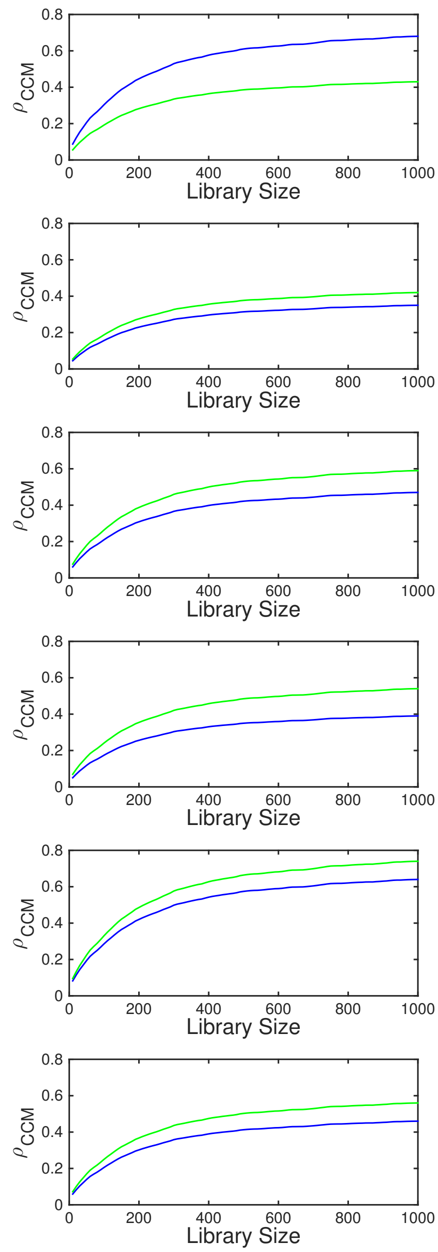

Figure 4: Illustration of the output of Convergent Cross Mapping applied to the paired AU12 time series of dyads 1 to 6 (top to bottom). The curves show the increase of the prediction quality ρ_{CCM} as a function of the number of samples (Library Size L). The increase is a qualitative measure of causal coupling. The green curve represents the causal effect of the first dyad member on the second, and the blue curve the causal effect in the reverse direction.

quired to establish the weaknesses and strengths of applying CCM in cognitive science. An important issue concerns the question to what extent behavioral coupling dynamics correspond to the complex system dynamics for which CCM and related methods were formulated. As a case in point, the "attractor manifold" depicted in Figure 3 appears quite chaotic when compared to the smooth attractor manifold of the Lorenz system in Figure 1. Although a difference in appearance is to be expected, given that the Lorenz system is a deterministic mathematical model and the dyadic AU12 dynamics reflect real-world behavioral measurements, a further analysis of coupled facial expressions is needed to establish the precise nature of the dynamics in relation to established mathematical formulations. Still, since CCM has been applied successfully to other noisy real-world time-series, such as the causal relation between galactic cosmic rays and annual variations in global temperature (Tsonis et al., 2015), we are confident about its validity as a measurement tool for establishing behavioral causality.

Notwithstanding our encouraging results, the relative small number of samples (6 dyads) does not warrant any far-reaching conclusions. In our future studies we will determine if the results generalize to a larger set of dyads, to other action units, and to other nonverbal behaviors. The main conclusion that we draw from our exploration is that Convergent Cross Mapping appears to provide a viable and fruitful alternative to existing measures of synchrony and mimicry.

References

Achaibou, A., Pourtois, G., Schwartz, S., & Vuilleumier, P. . (2008). Simultaneous recording of eeg and facial muscle reactions during spontaneous emotional mimicry. *Neuropsychologia*, *46*(4), 1104-1113.

Boker, S. M., Xu, M., Rotondo, J. L., & King, K. (2002, September). Windowed cross-correlation and peak picking for the analysis of variability in the association between behavioral time series. *Psychological Methods*, *7*(3), 338–55.

Briton, N., & Hall, J. (1995). Beliefs about female and male nonverbal communication. *Sex Roles*, *32*(1-2), 79-90.

Chartrand, T., & van Baaren, R. (2009). Human mimicry. *Advances in Experimental Social Psychology*, *41*, 219-274.

Cohn, J. F. (2010, November). Advances in behavioral science using automated facial image analysis and synthesis. *IEEE Signal Processing Magazine*, 128-133.

Delaherche, E., Chetouani, M., Mahdhaoui, A., Saint-Georges, C., Viaux, S., & Cohen, D. (2012, July). Interpersonal Synchrony: A Survey of Evaluation Methods across Disciplines. *IEEE Transactions on Affective Computing*, *3*(3), 349–365. doi: 10.1109/T-AFFC.2012.12

Dimberg, U. (1990, September). Facial electromyography and emotional reactions. *Psychophysiology*, *27*(5), 481-494.

Dimberg, U., & Thunberg, M. (1998). Rapid facial reactions to emotional facial expressions. *Scandinavian Journal of Psychology*, *39*(1), 39-45.

Dimberg, U., Thunberg, M., & Elmehed, K. (2000). Unconscious facial reactions to emotional facial expressions. *Psychological Science*, *11*(1), 86-89.

Ekman, P., & Friesen, W. V. (1976). Measuring facial movement. *Environmental Psychology and Nonverbal Behavior*, *1*(1), 56-75.

Kalimeri, K., Lepri, B., Kim, T., Pianesi, F., & Pentland, A. S. (2011). Automatic modeling of dominance effects using granger causality. In *Human behavior understanding* (p. 124-133). Springer Berlin Heidelberg. I.

LaFrance, M., Hecht, M. A., & Paluck, E. L. (2003). The contingent smile: a meta-analysis of sex differences in smiling. *Psychological Bulletin*, *192*(2), 305-334.

Littlewort, G., Whitehill, J., Wu, T., Fasel, I., Frank, M., Movellan, J., & Bartlett, M. (2011, March). The computer expression recognition toolbox (CERT). *Face and Gesture 2011*, 298–305. doi: 10.1109/FG.2011.5771414

Louwerse, M., Dale, R., Bard, E. G., & Jeuniaux, P. (2012). Behavior Matching in Multimodal Communication is Synchronized. *Cognitive Science*, *36*(8), 1–48.

Paxton, A., & Dale, R. (2012, October). Frame-differencing methods for measuring bodily synchrony in conversation. *Behavior research methods*. doi: 10.3758/s13428-012-0249-2

Richardson, D., & Dale, R. (2005). Looking To Understand: The Coupling Between Speakers' and Listeners' Eye Movements and Its Relationship to Discourse Comprehension. *Cognitive Science*, *29*(6), 1045–1060.

Sato, W., & Yoshikawa, S. (2007). Spontaneous facial mimicry in response to dynamic facial expressions. *Cognition*, *104*(1), 1-18.

Shockley, K. (2005). Cross Recurrence Quantification of Interpersonal Postural Activity. In *Tutorials in contemporary nonlinear methods for the behavioral sciences* (pp. 143–177).

Smith, M. L., Cottrell, G. W., Gosselin, F., & Schyns, P. G. (2005, March). Transmitting and decoding facial expressions. *Psychological Science*, *16*(3), 184–189. doi: 10.1111/j.0956-7976.2005.00801.x

Sugihara, G., May, R., Ye, H., Hsieh, C.-h., Deyle, E., Fogarty, M., & Munch, S. (2012, October). Detecting causality in complex ecosystems. *Science (New York, N.Y.)*, *338*(6106), 496–500. doi: 10.1126/science.1227079

Takens, F. (1981). Detecting strange attractors in turbulence. In *Dynamical systems and turbulence* (Vol. 898, pp. 366–381). Springer Verlag.

Tsonis, A. A., Deyle, E. R., May, R. M., Sugihara, G., Swansona, K., Verbeten, J. D., & Wang, G. (2015). Dynamical evidence for causality between galactic cosmic rays and interannual variation in global temperature. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(11), 3253-3256.