A Dyadic Active Inference Model of Conflict Conversations: An Exploratory Study on 16 Couples.

Tahmineh A. Koosha[1], Sören Pohlmann[1],
Melanie Fischer[1], Hideaki Shimazaki[2], Dominik Endres[1]

[1]Department of Psychology, Philipps University of Marburg, Marburg, Germany

[2]Graduate School of Informatics, Kyoto University, Kyoto, Japan

Abstract

We describe an active inference formulation of romantic couples during conflict conversations. We model partners' hidden states and the resulting observable sound and motion behaviors and their evolution in time. Agent models are equipped with an internal model of their own state and their partner's hidden state, i.e., a basic version of emotional theory of mind. This internal model enables the agent models to predict and subsequently act towards future behavioral observations, giving rise to patterns such as turn-taking or demand–withdraw-like interactions. We show that several, but not all model parameters can be recovered from simulated experimental data. Fitted parameter profiles from mixed-gender 16 couples suggest four couple-level clusters. These clusters differ in joint behavioral preferences and hidden state transition dynamics, exhibiting patterns consistent with female-demand & male-withdraw, male-demand & female-withdraw, and balanced behavioral engagement. We conclude that active inference is a promising modeling framework for interpersonal processes, even though parameter identifiability remains challenging, which limits interpretability.

Keywords: Interpersonal Dynamics; Active Inference; Multimodal Interaction; Couple Conflict; Computational Psychology

1 Introduction

1.1 Interdependence and Conflict Communication

Romantic relationships involve continuous interpersonal interdependence: As opposed to brief encounters between strangers where each person's individual needs and preferences might drive their actions, close and long-term relationships involve a shift towards consideration of both short- and long-term outcomes for oneself, one's partner, and the relationship as a whole (Kelley & Thibaut, 1978; Rusbult & Van Lange, 2003; VanderDrift & Agnew, 2019). This interdependence is particularly salient during couples' conversations, where each partner's behavior influences, and is influenced by, the other's responses. Given that conversations between romantic partners serve as an important context for emotional closeness, social support, and conflict resolution, couples' communication behaviors and interaction patterns have long been subject to investigation. In particular, an extensive tradition of observational research has highlighted the central role of constructive and maladaptive communication behaviors during couples' conflict discussions for long-term relationship functioning (Cohan & Bradbury, 1997; Leo et al., 2021; Woodin, 2011).

During conflict discussions, partners may display a variety of responses; they may engage with a disagreement, seek change, or disengage from the interaction. These responses unfold over time and have often been conceptualized in terms of different combinations of engagement and withdrawal, resulting in recurring patterns of mutual engagement, mutual withdrawal, or asymmetric engagement and withdrawal (Christensen & Shenk, 1991; Crenshaw et al., 2017; Leo et al., 2019).

However, engagement and withdrawal do not necessarily have the same meaning across interactions. Engagement may support constructive problem-solving or involve escalating confrontation, whereas disengagement may reflect avoidance, temporary regulation, or other interactional processes (Leo et al., 2019). Nevertheless, the demand-withdraw interaction pattern, in which one partner pressures for discussion or change while the other avoids or withdraws, has received particular attention because of its association with poorer relational and communicative outcomes (Eldridge et al., 2007; Schrodt et al., 2014). This pattern tends to exacerbate over time, with partners' responses becoming more polarized, which further prevents effective conflict resolution (Baucom & Atkins, 2013).

These recurring dyadic patterns raise a central question: how do partners' internal states and interpersonal goals jointly shape the moment-to-moment dynamics of conflict?

1.2 Theoretical and Computational Accounts of Couple Conflict

The interpersonal process model (IPM) (Baucom et al., 2015) links demand-withdraw dynamics to partners' emotional responses and their own and each other's behavior, highlighting the importance of cross-partner associations. Subsequent research has further examined the role of subjective emotional experience in these processes (Leo et al., 2021). From a motivational perspective, approach goals concern the pursuit of desirable social outcomes, whereas avoidance goals concern the prevention of undesirable outcomes (Gable, 2006). In romantic relationships, these goals have been linked to relationship satisfaction and responsiveness to partners' needs (Impett et al., 2010). Importantly, similar observable behaviors may reflect

different underlying motivations.

Interdependence theory (IT) provides a complementary perspective by emphasizing how partners' goals, anticipated outcomes, and behavioral choices jointly shape interpersonal interactions (Rusbult & Van Lange, 2003; VanderDrift & Agnew, 2019). Because the consequences of one partner's actions depend partly on the other's responses, partners may develop different expectations and preferences regarding the same interactional outcomes. Together, these perspectives suggest that understanding couple conflict requires considering both partners' internal states and preferences, rather than interpreting observable behavior in isolation.

These theoretical perspectives motivate computational approaches capable of representing how partners' behavior evolves through reciprocal interactions over time. Nonlinear dynamical models have represented reciprocal influences between partners (Cook et al., 1995; Gottman et al., 1999), while differential equation models have been fitted to dyadic affective data (Fischer, Loiselle, et al., 2024; Steele & Ferrer, 2011; Steele et al., 2014). Probabilistic approaches, including Bayesian nonparametric hidden semi-Markov models, have characterized affective interaction sequences (Griffin & Li, 2016), alongside statistical methods for analyzing dyadic temporal dependencies (Fuchs et al., 2017). Other research has examined partner-specific acoustic and vocal dynamics (Fischer, Baucom, et al., 2024; Fischer et al., 2017; Nasir et al., 2017; Wojda-Burlij et al., 2024), while recent computational models have simulated relationship conflict using two-agent frameworks (Olar et al., 2026) and multi-agent representations of demand-withdraw interactions in couples-therapy scenarios (Wang et al., 2026).

Together, these approaches provide complementary accounts of interpersonal dynamics but differ in how they represent internal states, reciprocal influences, and behavioral outcomes. A remaining challenge is to integrate these elements within a generative framework that links hidden states, beliefs about oneself and one's partner, and preferences over joint behavioral outcomes to action selection and observable sound and motion behaviors.

1.3 Coupled Active Inference for Interpersonal Interaction

Active inference (ActInf; Da Costa et al. (2020)) provides a probabilistic framework in which agents infer hidden states from observations and prior beliefs and select actions based on predictions of future outcomes and prior preferences. In interpersonal interactions, not only a partner's internal state but also one's own state may be partially hidden and must be inferred from observable behaviors.

In the present study, we equip each partner with two binary hidden states, one representing their own propensity to act, while the other represents their partner. These states may be interpreted as crude arousal representation in the spirit of modern versions of the James-Lange theory (Dalgleish, 2004; Dewey, 1894), where arousal leads to observable behavior, whose observation can in turn sustain this arousal. In our model, sound and motion behavior constitute observable outcomes. These behavioral observations allow each agent to infer their own and their partner's hidden state.

An important aspect of interpersonal inference is the ability to represent another person's internal states. Moutoussis et al. (2014) developed a formal model of interpersonal inference in a trust game, in which beliefs about a partner's prosocial or antisocial dispo-

sition inform predictions of their actions. Such representations are related to the broader psychological concept of Theory of Mind (ToM), the ability to attribute mental states to others (Baron-Cohen et al., 1997). Representing a partner's hidden state therefore provides a basic computational component of ToM, although it does not capture its full psychological complexity.

Active inference has also been applied to psychological expectations (Raghuveer & Endres, 2023) and interpersonal processes in computational psychiatry (Eckert et al., 2023; Kirchner et al., 2024). Related theoretical work has addressed reciprocal social interaction (Lehmann et al., 2024) and coordination through shared future expectations (Albarracin et al., 2024). More specifically, Friston and Frith (2015) introduced coupled active-inference models of communication, illustrating the emergence of turn-taking through simulated birdsong. Extending this perspective, Gwon et al. (2026) proposed a coupled active-inference model of dialogue in which agents infer their partners' policies to guide communication. These works mostly illustrate their proposed models through simulation studies without making contact with empirical data. First empirical applications have also been described: Arima and Okada (2026) modeled collaborative interaction using observed actions and gaze-related behaviors, but without explicitly representing the partner's hidden state.

Together, these developments provide a basis for modeling reciprocal interpersonal processes through coupled ActInf. However, their application to empirical data in general and recorded couple-conflict interactions in particular, with explicit representations of both partners' hidden states and preferences over joint behavioral outcomes, is missing.

1.4 The Present Study and Objectives

In the present study, we develop and empirically evaluate a coupled ActInf model of interpersonal dynamics during conflict conversations between romantic partners. We analyze approximately ten-minute discussions from 16 couples, using time-aligned binary sound and motion observations derived from audio and video recordings. These modalities capture complementary aspects of observable behavior: vocal behavior has been investigated in relation to emotional processes during couple conflict (Baucom et al., 2015), while movement characteristics can convey affective information (Roether et al., 2009). Together, these time-aligned observations describe patterns of sound and motion behavior during couple conflict conversations.

We address four objectives. First, we investigate whether coupled active inference can generate distinct dyadic behavioral patterns through variations in joint-outcome preferences and hidden state transitions and whether the model can account for the temporal dynamics of recorded sound and motion behaviors. We examine this through controlled simulations and observation-conditioned, one-step-ahead predictions.

Second, we evaluate the contributions of joint-outcome preferences and hidden state transitions by evaluating eight ActInf model variants with different parameter constraints via predictive model fits and approximate bayesian model comparison.

Third, we assess parameter recoverability through simulation-based analyses with known generating parameters, allowing us to examine the limits of the interpretability of the fitted parameters.

Fourth, we explore variation in fitted preferences and hidden state transition parame-

ters across couples. Using exploratory couple-level clustering and representative conversation trajectories, we examine whether different parameter configurations are associated with distinct observable interaction patterns, including behavioral asymmetries resembling demand–withdraw dynamics and patterns of balanced behavioral engagement.

Together, these analyses demonstrate the potential of coupled active inference to connect interpersonal beliefs, preferences, and state transitions with observed dyadic behavior. Our study establishes a computational basis for future work, such as testing the emotional mechanisms of the interpersonal process model viewed from an ActInf perspective.

2 Method

2.1 Participants and Procedure

The present analysis included $N = 16$ mixed-gender romantic couples participating in a larger clinical trial for the treatment of depression. Only baseline data collected before treatment were used. Patients were 68.75% female, 31.25% male, and $M = 41.44$ ($SD = 13.36$) years old. Partners were 68.75% male and 31.25% female, and were $M = 44.50$ ($SD = 13.70$) years old. 56.25% of couples were married and 93.75% were cohabiting.

Each couple completed a 10-minute conflict discussion. Before the discussion, partners independently rated the degree of conflict they experienced across several potential relationship problem areas. Couples were then instructed to share their thoughts and feelings about a topic that both partners had rated as moderately conflictual. The discussions were audio- and video-recorded to derive time-aligned behavioral observations for both partners.

All participants provided informed consent in accordance with the Declaration of Helsinki. The study procedures were approved by the relevant institutional ethics committee (details withheld for blind review).

2.2 Behavioral Measures and Preprocessing

Audio and video recordings were processed to obtain four time-aligned binary behavioral observations at one-second resolution: sound and motion for each partner. Although additional acoustic, visual, and text-derived features were extracted, only these two modalities were used in the present model.

Sound Observations. Each session contained two participant-specific close-microphone recordings and one room-microphone recording. Audio was converted to mono at its native sampling rate and divided into non-overlapping 250-ms windows. Root-mean-square (RMS) amplitude was calculated for each microphone and window. To reduce cross-talk, acoustic measurements were assigned using the relative RMS amplitude of the two close microphones:

$$\rho_1(t) = \frac{R_1(t)}{R_1(t) + R_2(t) + 10^{-10}}, \tag{1}$$

where $R_1(t)$ and $R_2(t)$ denote the respective microphone RMS amplitudes. Windows with $\rho_1(t) > 0.60$ were assigned to participant 1, and those with $\rho_1(t) < 0.40$ to participant 2. For

intermediate values, each participant's microphone measurement was retained.

Physical silence was identified separately when all three microphone RMS values fell below a session-specific threshold defined as the fifth percentile of room-microphone RMS. Thus, the assignment procedure was based on microphone amplitude rather than source separation or speaker diarization.

Assigned RMS values were log-transformed as $L(t) = \log_{10}(\mathrm{RMS}(t) + 10^{-10})$ and aggregated into one-second intervals using the median. Participant-specific measurements were then masked for intervals classified as non-speaking according to timestamped transcripts. The resulting binary sound observation was defined as

$$o_{i,t}^{\text{sound}} = \begin{cases} 1, & \text{if a valid log-RMS value was retained,} \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

Thus, sound indicated the availability of a retained participant-specific acoustic measurement after speaking-status masking, rather than vocal intensity. A zero did not necessarily indicate physical silence.

Motion Observations. Participant-specific movement was quantified using dense optical flow and body segmentation. MediaPipe Pose Landmarker generated participant masks at a nominal rate of 5 Hz. Masks were assigned according to horizontal torso position and the manually verified seating arrangement. The most recent valid mask was retained for intermediate frames for up to 0.40 s. Overlapping pixels were assigned to the participant with the higher segmentation probability, provided it was at least 0.50.

Dense Farnebäck optical flow was computed between consecutive grayscale frames resized to 50% of their original dimensions, using a pyramid scale of 0.5, three levels, a 15-pixel window, three iterations, poly_n = 5, and poly_sigma = 1.2. Motion energy was defined as the mean optical-flow magnitude within each participant's valid mask:

$$M_{i,t} = \frac{1}{|\Omega_{i,t}|} \sum_{p \in \Omega_{i,t}} \sqrt{u_p(t)^2 + v_p(t)^2}, \tag{3}$$

where $\Omega_{i,t}$ denotes the segmentation mask and $u_p(t)$ and $v_p(t)$ are the optical-flow components at pixel p. Frame-level values were averaged into one-second intervals, excluding frames without valid masks. This procedure reduced background contributions, although global camera motion was not explicitly compensated.

Internal gaps of up to three consecutive missing motion values were linearly interpolated; remaining missing observations were excluded from model fitting. Nonnegative motion energy was transformed as

$$z_{i,t} = \log\left(1 + \frac{M_{i,t}}{0.05}\right). \tag{4}$$

A two-component Gaussian mixture model (GMM) was fitted separately to each participant's transformed motion sequence, using ten initializations, full covariance, and covariance regularization of 10^{-5}. Each observation was assigned to its most probable component. The component with the higher mean was coded as motion = 1 (behavior), and the lower-mean

component as motion = 0 (no behavior). Thus, these categories represent participant-relative movement levels.

Temporal Alignment and Model Input. Audio, video, and transcript-derived speaking information were aligned using session identifiers, participant identities, and exact one-second interval boundaries. Sessions with incompatible timestamps or interval keys were excluded rather than temporally reconstructed.

The final model input comprised four synchronized binary behavior observation sequences per couple: sound and motion for each partner. Model fitting required complete observations for both partners at every retained time point.

2.3 Coupled Active-Inference Model

Model Overview. We represented each couple as two interacting active-inference agents, each equipped with a generative model of their own and their partner's behavior, see fig. 1. Active inference combines inference about hidden states with action selection based on predicted outcomes and prior preferences (Da Costa et al., 2020; Smith et al., 2022).

We used a discrete-state partially observable Markov decision process (POMDP) implemented in `pymdp` (Heins et al., 2022). Each agent's generative model comprised observation likelihoods (A), action-dependent state transitions (B), preferences over joint behavioral outcomes (C), and prior beliefs about initial states (D).

The agents were coupled through shared sound and motion observations of both partners. Unlike the symmetric-agent formulation proposed in (Friston & Frith, 2015), our agents maintained separate beliefs and could have different preferences and transition parameters.

Hidden States and Joint Observations. Each agent $i \in \{1, 2\}$ maintained two binary hidden-state factors representing its own and its partner's states:

$$\mathbf{s}_{i,t} = \left(s_{i,t}^{\text{self}}, s_{i,t}^{\text{other}}\right), \qquad s_{i,t}^{(\cdot)} \in \{0, 1\}. \tag{5}$$

The hidden (unobservable) arousal states were labeled inactive ($s = 0$) and active ($s = 1$). Sound and motion behavior observations depended on the same hidden state for each person. For each modality $m \in \{\text{sound}, \text{motion}\}$, the two binary behavioral observations were combined into a four-category joint outcome:

$$o_{i,t}^{m,\text{joint}} = 2o_{i,t}^{m,\text{self}} + o_{i,t}^{m,\text{other}}. \tag{6}$$

The categories were $0 : (0, 0)$, $1 : (0, 1)$, $2 : (1, 0)$, and $3 : (1, 1)$, where the first entry of each tuple refers to the agent's own behavior observation. Thus, the same dyadic behavior was represented from each partner's perspective.

Observation Likelihoods (A). The observation model specified how behavioral observations depended on hidden states. The same fixed likelihood matrix was used for sound and motion:

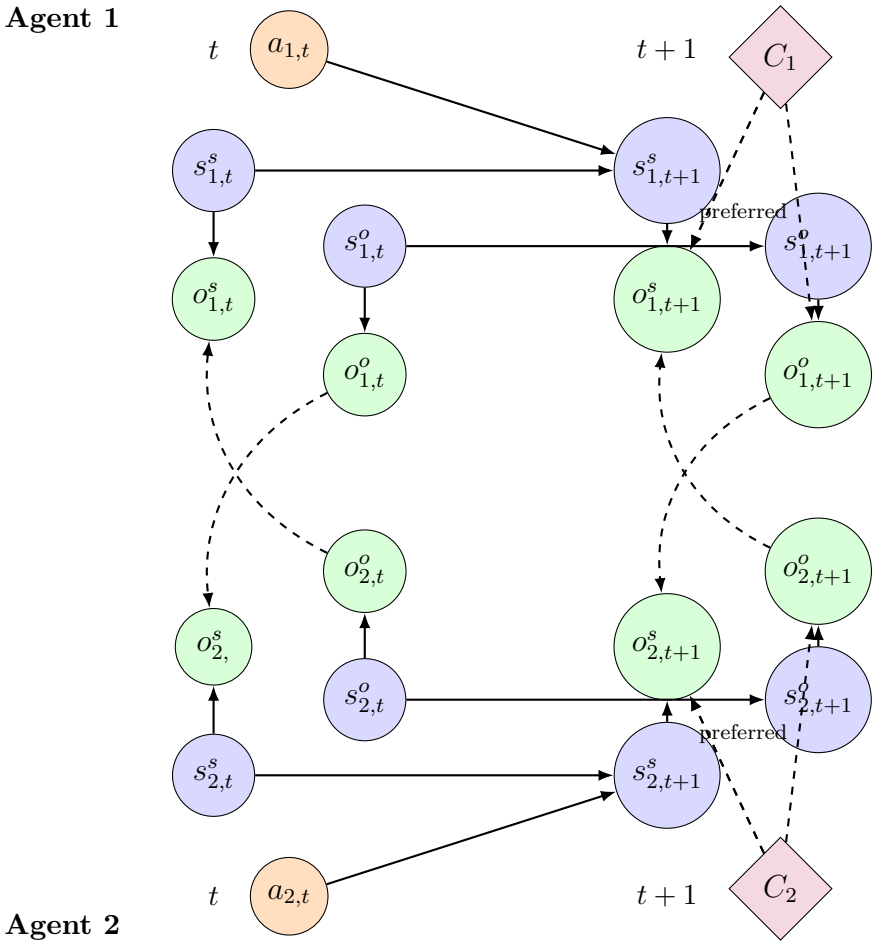


Figure 1: Coupled active-inference model of dyadic interaction. Each agent $1, 2$ represents its own $s^s_{1/2,t}$ and its partner's hidden states $s^o_{1/2,t}$, receives both partners' behavioral observations ($o^s_{1/2,t}$, $o^o_{1/2,t}$), and selects actions $a_{1/2,t}$ affecting its own state transitions. Preferences over joint behavioral outcomes C_1, C_2 guide action selection. Dashed connections indicate that the other partner o is observing self-generated behavior.

$$A^m_{o,s} = P(o^m_t = o \mid s_t = s), \qquad A^{\text{sound}} = A^{\text{motion}} = \begin{pmatrix} 0.90 & 0.05 \\ 0.10 & 0.95 \end{pmatrix}. \tag{7}$$

Rows correspond to the binary observation categories ($o = 0$: no-behavior; $o = 1$: behavior), and columns to the hidden states ($s = 0$: inactive; $s = 1$: active).

We assigned a probability of 0.90 to no-behavior given $s = 0$ and a probability of 0.95 to behavior given $s = 1$. This choice reflects the assumption that an active state will very probably lead to observable behavior, unless the behavior is down-regulated to a point where our thresholding procedure (cf. section 2.2) will no longer register it. Conversely, an inactive state will probably result in no behavior, with a 0.1 chance that a habitual action will be executed, e.g. comfort movements or throat cleaning sounds. These likelihoods were fixed modeling assumptions rather than parameters estimated from the recordings. For each modality, joint observation likelihoods were constructed as

$$P(o^{m,\text{self}}, o^{m,\text{other}} \mid s^{\text{self}}, s^{\text{other}}) = A^m_{o^{m,\text{self}}, s^{\text{self}}} A^m_{o^{m,\text{other}}, s^{\text{other}}}. \tag{8}$$

This formulation assumed conditional independence of the partners' observations given their hidden states and conditional independence of sound and motion given those states.

State Transitions (B). Each agent selected one of two abstract regulatory actions, $a_{i,t} \in \{0, 1\}$, corresponding to *stay* and *switch*. These actions probabilistically favored persistence or switching of the hidden state, respectively. Let

$$q_{i,a,s} = P(s^{\text{self}}_{i,t+1} = 0 \mid s^{\text{self}}_{i,t} = s, a_{i,t} = a). \tag{9}$$

The transition matrix for the agent's own hidden state was

$$B^{\text{self}}_i(a) = \begin{pmatrix} q_{i,a,0} & q_{i,a,1} \\ 1 - q_{i,a,0} & 1 - q_{i,a,1} \end{pmatrix}. \tag{10}$$

The unrestricted model estimated four transition probabilities per agent. In action-antisymmetric models (`Banti` variants; Table 1), we imposed

$$q_{i,1,0} = q_{i,0,1}, \qquad q_{i,1,1} = q_{i,0,0}, \tag{11}$$

reducing the number of free transition parameters to two per agent. This constraint makes the transition probabilities under the switch action correspond to those under the stay action with the previous hidden-state categories at timestep t exchanged.

Each agent represented its partner's state transitions using the no-switch dynamics of its own internal model, i.e. assuming that the partner would usually not execute internal regulatory actions:

$$B^{\text{other}}_i = B^{\text{self}}_i(a = 0). \tag{12}$$

This self–partner similarity assumption concerned the agent's internal model of its partner; it did not require the two agents to have identical fitted parameters. The partner's hidden-state factor was not directly controlled by the agent's actions.

Preferences (C) and Priors (D). For each modality m, agent i specified relative logit preferences over the four joint behavioral outcomes:

$$\mathbf{C}^m_i = (0, c^m_{i,(0,1)}, c^m_{i,(1,0)}, c^m_{i,(1,1)}), \tag{13}$$

where the joint no-behavior outcome $(0, 0)$ served as the reference category. Preferences influenced action selection by favoring policies expected to produce preferred observations. Full-C models estimated all three nonreference preferences per agent and modality, whereas restricted-C models estimated only $c^m_{i,10}$, corresponding to the self-behavior, partner-no-behavior outcome.

The default preference vector was $(0, 3, 3, -4)$ for both modalities, favoring outcomes in which exactly one partner shows behavior and assigning the lowest preference to joint

behavior. From the perspective of approach and avoidance motivations (Gable, 2006), under the default preference, both partners should approach turn-taking behavior and avoid talking or moving simultaneously.

Initial beliefs over each binary hidden-state factor were specified as

$$D = (0.95, 0.05), \tag{14}$$

assigning a prior probability of 0.95 to an inactive state $s = 0$ at the beginning of the interaction.

Belief Updating and Action Selection. At each time step, both agents received the recorded sound and motion observations of both partners and updated their beliefs about the hidden states.

Each agent independently evaluated one-step policies according to expected free energy, which combines the expected attainment of preferred outcomes (pragmatic value) with expected uncertainty reduction (epistemic value). The posterior probability of a policy π was expressed as

$$Q_i(\pi_i \mid o_{1:t}) \propto \exp\left[\alpha \cdot (\log(E_i(\pi)) - \gamma G_i(\pi))\right], \tag{15}$$

where $G_i(\pi)$ is expected free energy of partner i and $E_i(\pi)$ is the prior over policies, which was uniform. Policy precision was $\gamma = 1$, and action-sampling precision was $\alpha = 2$.

Both agents sampled their unobserved regulatory actions synchronously but independently. Following action selection, their beliefs were propagated through the corresponding transition models to predict the next behavioral observations. These one-step-ahead predictions were conditioned on the recorded interaction history.

2.4 Model Variants and Parameter Constraints

We fitted eight model variants with different constraints on joint behavioral preferences (C) and hidden state transitions (B), separately for each couple (Table 1). Parameters were agent-specific and k denote the total number of free parameters per dyad.

Table 1: Model variants and numbers of free parameters per dyad.

Model	Preferences C	Transitions B	k
`C`	Full	Fixed	12
`Cres`	Restricted	Fixed	4
`B`	Fixed	Full	8
`Banti`	Fixed	Antisymmetric	4
`BC`	Full	Full	20
`BantiC`	Full	Antisymmetric	16
`BantiCres`	Restricted	Antisymmetric	8
`BantiHighCres`	Restricted	Antisymmetric, tight bounds	8

For unrestricted B, the parameter bounds were

$$q_{i,0,0} \in [0.51, 0.99], \quad q_{i,0,1} \in [0.01, 0.49],$$
$$q_{i,1,0} \in [0.01, 0.49], \quad q_{i,1,1} \in [0.51, 0.99].$$

In antisymmetric variants, only the first two parameters were free, with the corresponding bounds. For `BantiHighCres`, these bounds were tightened to $[0.90, 0.99]$ and $[0.01, 0.10]$, respectively, to examine whether near-deterministic transitions improve preference recovery in simulations. Fixed transition probabilities were $(0.95, 0.05, 0.05, 0.95)$.

For C, the two single-behavior preferences $c^m_{i,(0,1)}$ and $c^m_{i,(1,0)}$ were bounded by $[0, 5]$, and the joint-behavior preference $c^m_{i,(1,1)}$ by $[-5, 0]$ during optimization, assuming that couples prefer non-escalating conversations in our experimental setup. Fixed parameters retained their default values specified above. Optimization was initialized at the applicable fixed values or the midpoints of the parameter bounds.

2.5 Parameter Estimation and Model Comparison

Each model variant was fitted separately to each couple using `pymdp` inference, JAX-compiled predictions, and bounded Py-BOBYQA optimization.

Model fitting minimized the negative log probability of the recorded one-step-ahead behavioral observations. At each objective evaluation, the observed sequence was processed using $R = 100$ stochastic rollouts, representing independent Monte Carlo realizations of the regulatory actions and behavior predictions conditioned on past observations of the same couple.

Let $p^{(r)}_{i,m,t}(o)$ denote the predicted probability of partner i's observation o in modality m at time t, conditioned on observations through time t in rollout r. For T aligned time points, the predictive loss was

$$\mathcal{L}_{\text{pred}}(\boldsymbol{\theta}) = -\sum_{i=1}^{2}\sum_{m} \frac{1}{R(T-1)} \sum_{r=1}^{R}\sum_{t=1}^{T-1} \log\Big[\max\left\{p^{(r)}_{i,m,t}(o^m_{i,t+1}), 10^{-12}\right\}\Big]. \tag{16}$$

The first observation was excluded because no preceding prediction was available. Since we average the rollouts outside the logarithm, this loss is actually an upper bound on the negative rollout-marginal log likelihood, comparable to variational free energy. To regularize parameter estimation, we added a weak quadratic penalty centered on the midpoint of each parameter's admissible range:

$$\mathcal{J}(\boldsymbol{\theta}) = \mathcal{L}_{\text{pred}}(\boldsymbol{\theta}) + 10^{-3}\sum_{j=1}^{k}\left[\frac{\theta_j - (u_j + \ell_j)/2}{u_j - \ell_j}\right]^2, \tag{17}$$

where ℓ_j and u_j denote the lower and upper bounds of parameter θ_j. Py-BOBYQA used bounded, noise-aware optimization with restarts, $2k + 1$ interpolation points, a final trust-region radius of 10^{-4}, and at most 2,000 function evaluations per fit. Optimization termination status was recorded.

For model comparison, we calculated an approximate Bayesian information criterion (BIC; Bishop (2006)):

$$\widetilde{\mathrm{BIC}} = 2(T-1)\mathcal{L}_{\mathrm{pred}} + k\log[4(T-1)]. \tag{18}$$

Here, k denotes the number of free parameters per dyad. Because the predictive loss was derived from stochastic predictions in a coupled dyadic model, rather than a conventional independent- observation likelihood, this criterion was used as a heuristic measure of predictive fit penalized for model complexity.

2.6 Simulation-Based Parameter Recovery

To assess parameter recoverability, we sampled known generating parameters uniformly within each examined model's bounds and generated synthetic dyadic sound and motion sequences. Recovery optimization was initialized at the midpoint of each parameter's admissible range, using the same objective function as empirical fitting and $R = 100$ stochastic rollouts.

Recovery was evaluated using the Pearson correlation coefficient between generating and recovered parameter values, calculated separately for each free parameter across simulation trials:

$$r_j = \frac{\sum_{n=1}^{N}(\theta^*_{j,n} - \bar{\theta}^*_j)(\hat{\theta}_{j,n} - \bar{\hat{\theta}}_j)}{\sqrt{\sum_{n=1}^{N}(\theta^*_{j,n} - \bar{\theta}^*_j)^2 \sum_{n=1}^{N}(\hat{\theta}_{j,n} - \bar{\hat{\theta}}_j)^2}}. \tag{19}$$

Here, $\theta^*_{j,n}$ and $\hat{\theta}_{j,n}$ denote the generating and recovered values of parameter j in trial n, respectively, and N is the number of simulation trials. We conducted 100 trials per examined model variant, each comprising 600 time points, with up to 2,000 function evaluations per fit.

Higher positive correlations indicate better preservation of between-trial parameter variation, but do not rule out systematic bias or estimation error. Recovery under matched generating and fitting assumptions therefore assesses ordinal interpretability of recovered parameters.

3 Results

3.1 Simulations

We conducted four free-running synthetic dyadic simulations to illustrate how transition dynamics (B) and joint-outcome preferences (C) shape interpersonal activity patterns. Each scenario was simulated for 600 time steps, with 100 stochastic realizations.

The gray traces show binary activity observations (0 = inactive, 1 = active) generated in one simulated realization. The orange traces show the corresponding model-predicted probability of activity, $P(o_t = 1)$, from the same realization. Thus, the gray traces represent sampled outcomes, whereas the orange traces represent probabilistic predictions. Because the model is stochastic, a sampled outcome may differ from the most probable outcome.

The joint-activity strips summarize the binary observations of both agents at each time step, distinguishing four states: neither active, man only, woman only, and both active. Separate strips are shown for sound and motion.



Figure 2: Illustrative dyadic interaction patterns generated by the active-inference model. A. Persistent speaker activity with occasional switching under default preferences and near-deterministic transitions. B. Frequent activity switching under less-deterministic transitions. C. Asymmetric activity between agents with dominant and submissive preferences. D. Sustained simultaneous activity when both agents have dominant preferences. Light gray lines show binary behavior (0/1) sampled from the model, while orange curves represent the model-predicted probability of sound activity for the same simulation run. The joint activity strips indicate whether neither agent, only the man, only the woman, or both agents are active. Each panel displays the first 300 one-second simulation steps. The movement modality evolves qualitatively similar to the sound modality. All parameters were specified rather than estimated from empirical observations. Orange curves show the agent's predicted probability of behavior ($o = 1$), while blue dashed curves show the corresponding partner's prediction.

The displayed trajectories in fig. 2 correspond to one realization, whereas summary measures, displayed as solid colored bars under the corresponding trajectories, were calculated across 100 realizations. We considered three preference vectors: default $C = (0, 3, 3, -4)$, dominant $C = (0, 0, 5, 2)$, and submissive $C = (0, 2, 1, -5)$, with entries ordered as $(00, 01, 10, 11)$. Near-deterministic transitions were specified as $B = (0.98, 0.02, 0.02, 0.98)$, whereas less-deterministic transitions used $B = (0.70, 0.30, 0.30, 0.70)$.

The trajectories fig. 2.A. were generated by the default C, which promotes joint observations (sound and motion) where one partner is active and the other one is quiet/still. These mutually exclusive activity states are preferred joint stillness, whereas the least preferable state is joint activity, e.g. both partners talking at the same time. State transitions are near deterministic, i.e. activity states are 'sticky' with high probability. This choice of C and B leads to *turn-taking behavior*: except for short interjections or gaps, one speaker tends to be active for a prolonged period of time, then hands off to the other speaker. This turn-taking behavior is a feature that persists on average, as can be seen in the solid colored bars.

When simulating with the default C but the less-deterministic B, we still obtain mostly

turn-taking, but with speaking bouts that are much shorter, see fig. 2.B. Thus, tuning of the state transition probabilities in B enable the model to accommodate a range of speaking and moving habits.

So far, preferences were symmetric between partners. Fig. 2.C. shows behavior resulting from a dominant partner 'man' interacting with a submissive partner 'woman.' We coded dominance as a preference for self-active, other-inactive observations, and vice versa for submission. B was near-deterministic. Consequently, it is mostly the dominant partner that shows both sound and motor output. The 'woman' tries to interject at some points in the simulated conversation, which leads to both active states, but 'he' just won't stop talking and 'she' quickly becomes inactive again. This might be seen as a computational realization of the *demand-withdraw* pattern (Baucom et al., 2015).

Finally, in fig. 2.D., a typical interaction between two dominant partners is shown. Here, both partners tend to be active most of the time, disregarding the other's sound/motor output. This might, e.g., be the case during an escalating conflict.

In summary, these simulations demonstrate that a range of interaction patterns can be reproduced by our model with suitable settings for B and C.

3.2 Model Comparison

We compared eight model variants using the Bayesian Information Criterion (BIC), where lower values indicate a better trade-off between predictive fit and model complexity.

Figure 3 shows the BIC values for each couple and their sum across all 16 couples. The combined B+C model achieved the lowest summed BIC, suggesting that jointly modeling preferences (C) and state transitions (B) provides a better overall account of the observed interactions than modeling either component alone. The worst BIC scores, which are mostly driven by the agent-averaged log-likelihoods, are reached by models C, Cres, and BantiHighCres. What these models have in common is a near-deterministic state transition matrix B (cf. table 1). This result therefore highlights the importance of stochasticity in the state transitions for the description of real conversation data.

Of the constrained models, BantiCres yields the lowest BIC score. Since this model is also reasonably recoverable from our experimental data (see next section), it will be the focus of our model-driven data evaluation.

3.3 Parameter Recovery

We assessed parameter recovery by comparing generating and recovered parameter values across 100 simulated datasets per examined model variant. Figure 4 summarizes the Pearson correlation coefficients for each free parameter, while Supplementary Figure S1 (ESM 1) shows the corresponding generating-versus-recovered scatterplots.

In the Cres model, preference estimates showed substantial dispersion, bias, and saturation near parameter bounds. However, a noisy monotonic relationship between generating and recovered parameters was obtainable; see also the correlation coefficients in fig. 4. This allows us to interpret at least the relative ordering of recovered Cres values.

In the combined BantiCres model, antisymmetric transition parameters B. showed clearer recovery patterns than restricted preferences C., which remained noisy and biased,

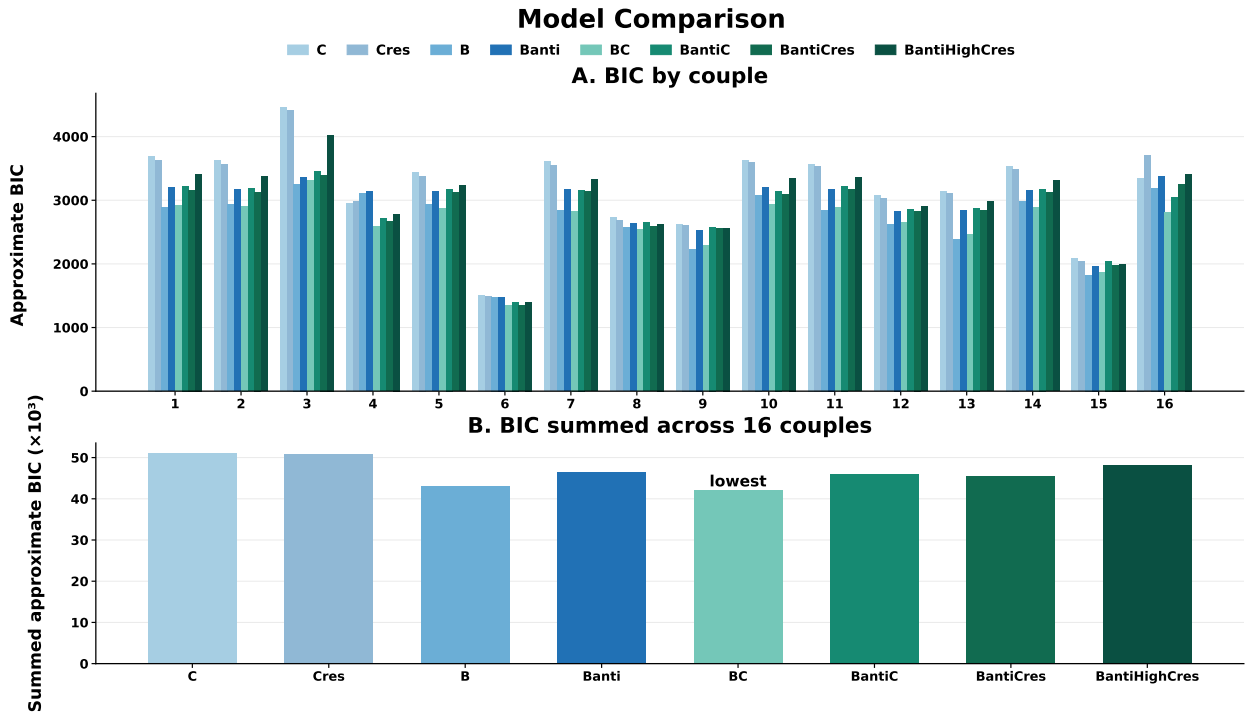


Figure 3: Model comparison across 16 couples under the synchronous action mode. A. Approximate BIC for each couple across eight model variants. B. Approximate BIC summed across couples. Lower values indicate better model fit after penalizing complexity.

and showed lower correlation coefficients than in the Cres model with near-deterministic B. This reduced preference recoverability is probably due to the construction of a typical active inference model: observation preference drives action choice through expected future observations. If these future observations are hard to predict due to a stochastic transition matrix B, then actions cannot bring about preferred observations with certainty, and therefore preferences are difficult to recover from observed behavior. Constraining B toward near-deterministic transitions in BantiHighCres did alleviate this difficulty to some degree, however at the expense of a model that is worse at describing the data, as detailed in the model comparison section.

These results indicate that improved predictive fit does not necessarily imply reliable parameter recovery. We therefore interpret fitted parameters cautiously, with greater emphasis on those showing stronger recovery.

3.4 Fitted Parameters

Figure 5 summarizes the fitted BantiCres parameters across 16 couples and the results of exploratory parameter clustering.

As shown in Fig. 5 B, the entries of the antisymmetric B matrix are fairly deterministic in many individuals, but not as high as our 'fixed B' model assumption (cf. table 1). However, in many instances the B entries are deterministic enough to allow for a cautious interpretability of the recovered Cres values. Generally speaking, inactivity persistence tended to be high,

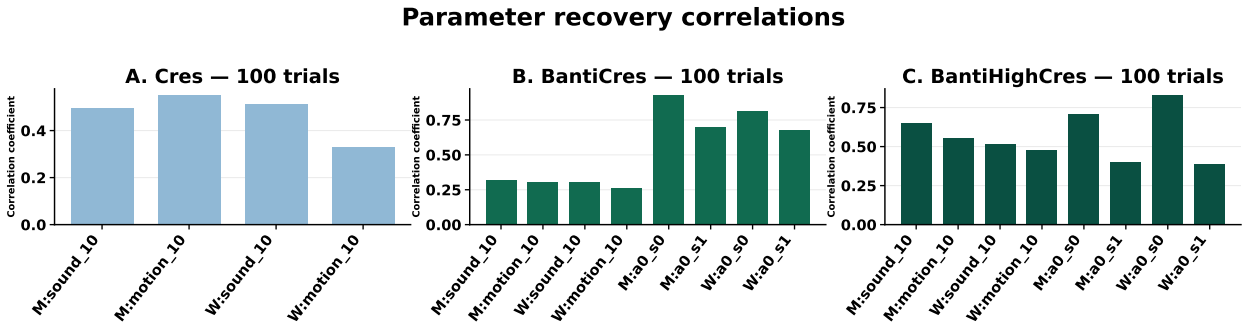


Figure 4: Parameter recovery correlations across 100 simulation trials for A. Cres, B. BantiCres, and C. BantiHighCres. Higher correlations indicate better correspondence between generating and recovered parameter values. Generating-versus-recovered scatterplots are provided in Supplementary Figure S1 (ESM 1).

with the values recovered from men being more tightly concentrated around 0.8 and women being more variable, i.e., male speaking bouts lengths were more stereotyped. Active-to-inactive transition probabilities were lower and more variable in general. The Cres values shown in fig. 5 represent preferences for self-behavior and partner-no-behavior outcomes in sound and motion. Fitted preferences varied substantially across individuals, with broad distributions in both partners and slightly higher behavior preferences for both sound and motion in men.

Given these broad distributions of preferences (and transitions), we wondered if the differences between individuals and dyads could be organized into interaction subtypes. To address this question, we applied Gaussian Mixture Model (GMM) clustering on the four fitted parameters per participant: two restricted C preferences and two action-0 transition probabilities. Parameter vectors from men and women were concatenated, yielding 16 eight-dimensional observations. The parameters were standardized before fitting GMMs with one to eight clusters. Due to the small number of couples and the absence of a hypothesis, this clustering analysis is exploratory.

As shown in Fig. 5B, BIC favored four clusters ($K = 4$), containing 4, 1, 3, and 8 couples, respectively.

Cluster-specific parameter means were examined separately for men and women. The clusters differed in their preference strengths and transition probabilities, with some gender-specific variation.

The PCA projection illustrates the within-cluster distribution in parameter space (Figure 5b). Clusters 1 and 3 are fairly tightly concentrated, which allows for an interpretation based on the mean. Cluster 4 is elongated, which entails a higher diversity between the couples. Cluster 2 contains only one couple, however it appears well separated from the other clusters.

The cluster centers showed distinct dyadic parameter configurations (Figure 5c,d). Cluster 1 exhibited very low self-behavior preferences (sound and motion) for men, and slightly higher preference for women. The transition probabilities are not at the bounds, which leads to short activity bouts, with the exception of the men's switch probability, which is near zero. In other words, men in this cluster are unlikely to become active. Together, the parameter

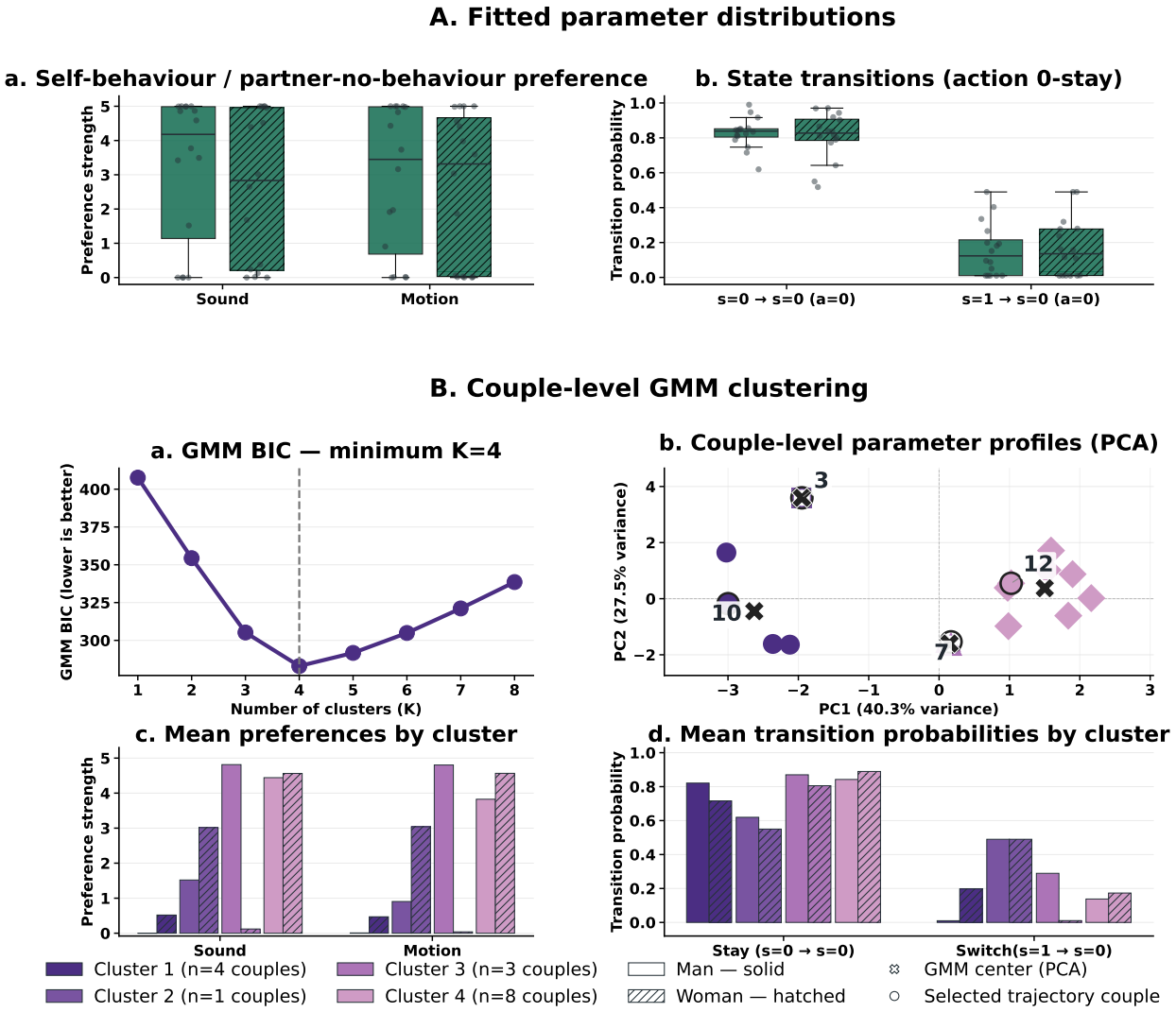


Figure 5: Fitted BantiCres parameters and exploratory clustering across 16 couples. **A.** Individual fitted-parameter distributions: a. restricted preferences for sound and motion outcomes; b. action-0 transition probabilities. Men are shown with solid fills and women with hatched fills. **B.** Pooled Gaussian mixture modeling: a. BIC across $K = 1$–8, with the minimum at $K = 4$; b. PCA projections of the cluster centers onto the dominant eigenvectors of the covariance matrix. c. cluster-specific sound and motion preference means; d. cluster-specific transition-probability means. GMMs were fitted to standardized parameter vectors from 16 couples. Bars in panels B(c–d) represent gender-specific arithmetic means of participants assigned to each cluster. Numbers in the PCA plot refer to the couples in section 3.5.

Table 2: Exploratory directions of within-cluster variation in fitted couple-level parameters. The three largest absolute loadings of the leading empirical covariance eigenvector are reported.

Cluster	n	Largest eigenvector loadings	Variance (%)
1	4	$B_{\text{w,ai}}$ +0.722; $B_{\text{w,ii}}$ −0.643; $C_{\text{w,motion}}$ +0.182	94.7
2	1	Not estimable	–
3	3	$B_{\text{m,ai}}$ +0.817; $B_{\text{m,ii}}$ +0.529; $C_{\text{m,motion}}$ −0.177	94.9
4	8	$B_{\text{m,ii}}$ +0.819; $C_{\text{m,motion}}$ −0.454; $B_{\text{w,ai}}$ −0.319	54.2

Covariance matrices were calculated from eight fitted parameters standardized across all analyzed couples. ii denotes inactive-to-inactive and ai active-to-inactive transitions under action 0. Variance is the proportion explained by the leading eigenvector within each cluster. Eigenvector signs are arbitrary. The single-couple cluster has no estimable sample covariance.

values of the couples in this group indicate a conversation dominated by the female partner, which is consistent with a *demanding* woman and a *withdrawing* man. The inspection of the within-cluster covariances in table 2 shows that the main (co)variability in cluster 1 is in the direction of the transition probabilities: women who are more likely to become inactive (Bw, ai) are also more likely to activate $(1 - B_{ii})$, i.e. faster vs. slower switching between activity states.

The couple in cluster 2 is described by intermediate self-behavior preferences and fairly stochastic transitions, which implies short activity bouts and unpredictable handoffs(Baucom & Atkins, 2013; Leo et al., 2019).

Cluster 3 showed asymmetric self-behavior preferences, which are substantially higher in men than women, and more deterministic transitions. This combination indicates a conversation in longer bouts dominated by the male partner, i.e. consistent with a *demanding* man and a *withdrawing* woman. The main within-cluster (co)variability is in the direction of male activity transitions: an increased active → inactive switching probability is accompanied by an increased probability of staying inactive, i.e. longer vs. shorter activity bout durations. Furthermore, the women in this couple are very unlikely to perform a spontaneous active-to-inactive transition.

Finally, in cluster 4 both partners are very eager to speak and move, with fairly deterministic transition probabilities (closer to the parameter bounds). This leads to a balanced conversation with long bouts, which is consistent with as productive *engagement*. Furthermore, men who are more likely to stay inactive ($B_{m,ii}$) also prefer less self-generated motion observations ($C_{m,motion}$).

3.5 Conversation Trajectories

To illustrate the temporal dynamics of the fitted `BantiCres` model, we selected the couple closest to the center of each of the four couple-level GMM clusters in standardized fitted-parameter space. Each couple was represented by the concatenated parameters of both partners.

Figure 6 shows the first 300 one-second bins of recorded sound behaviors and observation-conditioned model predictions for the four representative couples. The fitted parameters of these four representative couples are reported in Supplementary Table S2 (ESM 1). The examples illustrate differences in predicted activity patterns across the se-



Figure 6: Representative couple-level cluster trajectories. Recorded sound behavior and modeled probabilities for the couples nearest to the four GMM cluster centers: A. Cluster 1, Couple 10 ($n = 4$ couples); B. Cluster 2, Couple 3 ($n = 1$); C. Cluster 3, Couple 7 ($n = 3$); and D. Cluster 4, Couple 12 ($n = 8$). Light gray lines indicate recorded binary behaviors (0/1), whereas orange lines show the observation-conditioned predicted probability of activity from the fitted model. The joint sound behavior strips indicate whether neither partner, only the man, only the woman, or both partners were showing behavior. The first 300 one-second bins are displayed. Cluster assignments are based on the concatenated fitted parameters of both partners.

lected dyads. Couple 10 (cluster 1) shows a lower probability for vocal behavior by the man than by the woman, in line with the results from the cluster center analysis in the previous section. Consequently, the woman is talking most of the time. Couple 3 (cluster 2) shows nearly constant predicted activity probabilities, leading to short bouts and frequent, model-unpredictable handoffs. Couple 7 from cluster 3 is the inverse of cluster 1: here, the male partner strongly prefers self-behavior observations with fairly persistent activity states. Combined with a low female preference for self-behavior, this results in more predicted and more observed behavior from the male partner. The high self-behavior preferences combined with fairly deterministic transition probabilities in both partners lead to the long, predictable behavior bouts in couple 12 from cluster 4.

For space reasons, Figure 6 displays sound trajectories only. The corresponding sound and motion trajectories for all 16 couples are provided in Supplementary Figures S2–S17 (ESM 1), organized by parameter-target profiles. Supplementary Table S1 lists the profile-to-couple correspondence and selected fitted parameter values, while Supplementary Table S2 reports the BantiCres parameters of the four cluster-representative couples.

4 Discussion

We presented a dyadic active inference model of (conflict) conversations. Each agent is equipped with an internal model of hidden states of self and other, which represents our first step towards a computational description of emotional 'Theory of Mind' in the framework of active inference. Importantly, preferences for the observable behavior of self and other do not factorize across agents but reflect each partner's preferences about *joint* interaction behavior. Thus, our model is a computational formulation of the ideas put forth in interdependence theory (Rusbult & Van Lange, 2003; VanderDrift & Agnew, 2019) that close and long-term relationships involve a consideration of outcomes for oneself and one's partner simultaneously. This consideration, expressed as a coupled preference matrix C, drives partners' choices via the expected free energy (Da Costa et al., 2020).

The current model is deliberately minimal and should be understood as a starting point for the data-driven investigation of dyadic interactions with active inference agents. It features two binary behavioral observations, fixed sensory likelihoods A, and assumes that each agent's hidden state transitions B are mirrored in the internal model which an agent has about the partner. Control actions are regulatory and lead to a probabilistic switching of the arousal state.

Despite this fairly restrictive minimality, we were able to demonstrate that the dyad model can produce a range of interaction patterns, that are consistent with demand-withdraw, productive engagement or escalating conflict. Except for escalation, all patterns could be recovered from our sample of 16 couples with approximate bayesian model comparison, with a majority ($n = 8$) showing the engagement-consistent pattern. Whether the patterns that we recovered from the data are not just consistent with, but actually indicative of the aforementioned conversation patterns is a question for future research. Answering it will involve a richer space of observable signals, for example a finer-grained representation of amplitude and fundamental frequency than our binarized operationalizations of behavior, cf. (Baucom et al., 2015). The same applies to our simplistic binary arousal variable, which would clearly benefit from a multi-valued and possibly multi-dimensional representation.

To our knowledge, this is the first active inference model with two dyadically connected individual agents which is fitted to real interaction data.

4.1 Limitations and Future Work

Recovering the model's parameters proved to be challenging. We eventually settled for a transition matrix that antisymmetric with respect to the regulatory control action: the probability of transiting inactive $\rightarrow$ inactive under a 'stay' control action was constrained to be equal to the probability of active $\rightarrow$ inactive under a 'switch' action. The necessity for this constraint is probably due to the relative sparsity of 'switch' actions, which occur only in $\approx 10\%$ of timesteps. Hence, the part of the B matrix which describes state transitions after 'switch' actions is effectively constrained by only small fraction of the datapoints in a conversation trajectory, making the fitted value very noisy.

The recovery of C entries was even more challenging, probably for similar reasons. For example, in a normal, non-escalated conversation, it happens fairly seldom that both agents are active. Hence, the C entry which determines the preference for both agents behaving

(sound and/or movement) is underconstrained by the available conversation data. We could recover a noisy version of the preference for self-behavior and partner-no-behavior. While the relationship between generating and recovered parameter is far from linear, it appears at least approximately monotonic, allowing for an ordinal interpretation.

One important line of future work on dyadic active inference models is therefore the determination of data modalities, including observation cardinalities beyond simple binarizations, which are necessary to recover more of the model's parameters and thereby increase the data-driven interpretability. Particularly interesting would be a pitch variable with several levels, since the work on the IPM indicates that pitch is a reliable indicator for conflict escalation (Baucom et al., 2015). Parameter recoverability is a precondition for the use of such models as 'digial twins' (Grieves & Vickers, 2017), e.g. for the planning and *in silico* testing of therapeutic interventions.

We used an agent-averaged BIC for model comparison. Strictly speaking, bayesian model comparison schemes require one prediction per datapoint and model to compute the marginal likelihood, or approximations thereof. However, we have two predictions per model per datapoint. Averaging these two predictions is at best a heuristic that potentially makes the BIC approximation worse. New approaches have to be developed for model comparison in dyadic settings, e.g. combining predictions via product-of-experts. In this vein, the presented clustering results should be considered exploratory and in need of replication on a larger dataset.

Richer hidden state representation should also be considered. Future work will investigate arousal variables with more than two states, and possibly several state variables for the different observation modalities. It will be interesting to see if these enriched models can be made more informative by including relevant covariates into the recovery process, such as personality traits, or self-reports on relationship history.

Open Science Statement

Supplementary Figures S1–S17 and Supplementary Tables S1–S2 are provided as Electronic Supplementary Material 1 (ESM 1) accompanying this article. Analysis code, model specifications, and supplementary materials will be deposited in the TAM data hub Repository: https://tam-datahub.online.uni-marburg.de/home

Ethics Statement

All participants provided informed consent. The study procedures were approved by the local ethics committee of the Department of Psychology, Philipps University of Marburg, Germany.

Author Contributions

Conceptualization: TK, DE, MF, HS **Data Curation:** TK, SP **Formal Analysis:** TK, DE **Funding Acquisition:** DE, MF **Investigation:** MF **Methodology:** TK, DE, SP,

MF, HS **Project Administration:** TK, DE **Resources:** DE, MF **Software:** TK, DE, SP **Supervision:** DE **Validation:** DE, TK **Visualization:** TK, DE **Writing—original draft:** TK, DE **Writing—review and editing:** TK, DE, MF, HS

Conflict of Interest

The authors declare no conflicts of interest.

Funding

This work was supported by the DYNAMIC center, funded by the LOEWE program of the Hessian Ministry of Science and Arts (grant number: LOEWE1/16/519/03/09.001(0009)/98), the Deutsche Forschungsgemeinschaft (German Research Foundation, DFG) under Germany's Excellence Strategy (EXC 3066/1 "The Adaptive Mind", Project No. 533717223), the DFG GRK-RTG 2271 'Breaking Expectations' project number 29087897.

Electronic Supplementary Material

ESM 1. Supplementary Figures and Tables (ESM_1_Supplementary.pdf). Parameter-recovery scatterplots, observation-conditioned trajectories for all 16 couples, and supplementary fitted-parameter tables.

References

Albarracin, M., Pitliya, R. J., Smithe, T. S. C., Friedman, D. A., Friston, K., & Ramstead, M. J. D. (2024). Shared protentions in multi-agent active inference. *Entropy*, *26*(4), 303. https://doi.org/10.3390/e26040303

Arima, Y., & Okada, M. (2026). Active inference modeling of socially shared cognition in virtual reality. *Sensors*, *26*(2), 604. https://doi.org/10.3390/s26020604

Baron-Cohen, S., Jolliffe, T., Mortimore, C., & Robertson, M. (1997). Another advanced test of theory of mind: Evidence from very high functioning adults with autism or asperger syndrome. *Journal of Child Psychology and Psychiatry*, *38*(7), 813–822. https://doi.org/https://doi.org/10.1111/j.1469-7610.1997.tb01599.x

Baucom, B. R., & Atkins, D. C. (2013). Understanding marital distress: Polarization processes. In *Handbook of family theories* (pp. 145–166). Routledge.

Baucom, B. R., Dickenson, J. A., Atkins, D. C., Baucom, D. H., Fischer, M. S., Weusthoff, S., Hahlweg, K., & Zimmermann, T. (2015). The interpersonal process model of demand/withdraw behavior. *Journal of Family Psychology*, *29*(1), 80–90. https://doi.org/10.1037/fam0000044

Bishop, C. (2006). *Pattern recognition and machine learning* (Vol. 4). Springer New York. http://scholar.google.com/scholar.bib?q=info:jYxggZ6Ag1YJ:scholar.google.com/&output=citation&hl=en&as_sdt=0,5&as_vis=1&ct=citation&cd=0

Christensen, A., & Shenk, J. L. (1991). Communication, conflict, and psychological distance in nondistressed, clinic, and divorcing couples. *Journal of Consulting and Clinical Psychology*, *59*(3), 458–463. https://doi.org/10.1037/0022-006X.59.3.458

Cohan, C. L., & Bradbury, T. N. (1997). Negative life events, marital interaction, and the longitudinal course of newlywed marriage. *Journal of personality and social psychology*, *73*(1), 114–128. https://doi.org/10.1037/0022-3514.73.1.114

Cook, J., Tyson, R., White, J., Rushe, R., Gottman, J., & Murray, J. (1995). Mathematics of marital conflict: Qualitative dynamic mathematical modeling of marital interaction. *Journal of Family Psychology*, *9*(2), 110–130. https://doi.org/10.1037/0893-3200.9.2.110

Crenshaw, A. O., Christensen, A., Baucom, D. H., Epstein, N. B., & Baucom, B. R. (2017). Revised scoring and improved reliability for the communication patterns questionnaire. *Psychological assessment*, *29*(7), 913–925. https://doi.org/10.1037/pas0000385

Da Costa, L., Parr, T., Sajid, N., Veselic, S., Neacsu, V., & Friston, K. (2020). Active inference on discrete state-spaces: A synthesis. *Journal of Mathematical Psychology*, *99*, 102447. https://doi.org/10.1016/j.jmp.2020.102447

Dalgleish, T. (2004). The emotional brain. *Nature Reviews Neuroscience*, 582–589. https://doi.org/10.1038/nrn1432

Dewey, J. (1894). The theory of emotion: I: Emotional attitudes. *Psychological Review*, *1*(6), 553–569. https://doi.org/10.1037/h0069054

Eckert, A.-L., Pawlowski, J., Rief, W., Endres, D., & Kirchner, L. (2023). Active inference of interpersonal trust within and across mental disorders. *PsyArXiv*. https://doi.org/10.31234/osf.io/8aexf

Eldridge, K. A., Sevier, M., Jones, J., Atkins, D. C., & Christensen, A. (2007). Demand-withdraw communication in severely distressed, moderately distressed, and nondistressed couples: Rigidity and polarity during relationship and personal problem discussions. *Journal of Family Psychology*, *21*(2), 218–226. https://doi.org/10.1037/0893-3200.21.2.218

Fischer, M. S., Baucom, D. H., Baucom, B. R., Abramowitz, J. S., Kirby, J. S., & Bulik, C. M. (2017). Disorder-specific patterns of emotion coregulation in couples: Comparing obsessive compulsive disorder and anorexia nervosa. *Journal of Family Psychology*, *31*(3), 304–315. https://doi.org/10.1037/fam0000251

Fischer, M. S., Baucom, D. H., Weber, D. M., Bauer, D. J., Munion, A. K., Porter, L. S., Christensen, A., Bulik, C. M., Whisman, M. A., Abramowitz, J. S., et al. (2024). Interpersonal dynamics of vocal fundamental frequency in couples: Depressive symptoms, anxiety symptoms, and relationship distress. *Behaviour Research and Therapy*, *180*, 104571. https://doi.org/10.1016/j.brat.2024.104571

Fischer, M. S., Loiselle, R., Weber, D. M., Highlander, A., McCall, M. P., Cain, G. H., Forehand, R., & Jones, D. J. (2024). Parent–child emotion dynamics in families presenting for behavioral parent training: Is there a link with child behavior, parenting, and treatment outcome? *Journal of consulting and clinical psychology*, *92*(9), 543. https://doi.org/10.1037/ccp0000878

Friston, K., & Frith, C. (2015). Active inference, communication and hermeneutics. *Cortex*, *68*, 129–143. https://doi.org/10.1016/j.cortex.2015.03.025

Fuchs, P., Nussbeck, F. W., Meuwly, N., & Bodenmann, G. (2017). Analyzing dyadic sequence data—research questions and implied statistical models. *Frontiers in Psychology*, *8*, 429. https://doi.org/10.3389/fpsyg.2017.00429

Gable, S. L. (2006). Approach and avoidance social motives and goals. *Journal of Personality*, *74*(1), 175–222. https://doi.org/10.1111/j.1467-6494.2005.00373.x

Gottman, J., Swanson, C., & Murray, J. (1999). The mathematics of marital conflict: Dynamic mathematical nonlinear modeling of newlywed marital interaction. *Journal of Family Psychology*, *13*(1), 3–19. https://doi.org/10.1037/0893-3200.13.1.3

Grieves, M., & Vickers, J. (2017). Digital twin: Mitigating unpredictable, undesirable emergent behavior in complex systems. In F.-J. Kahlen, S. Flumerfelt, & A. Alves (Eds.), *Transdisciplinary perspectives on complex systems: New findings and approaches* (pp. 85–113). Springer International Publishing. https://doi.org/10.1007/978-3-319-38756-7_4

Griffin, W. A., & Li, X. (2016). Using bayesian nonparametric hidden semi-markov models to disentangle affect processes during marital interaction. *PLOS ONE*, *11*(5), e0155706. https://doi.org/10.1371/journal.pone.0155706

Gwon, Y., Murphy, E., & Chung, C. K. (2026). Evidence and intervention: A coupled active-inference extension of rational speech act models [Preprint]. https://arxiv.org/abs/2610.04347

Heins, C., Millidge, B., Demekas, D., Klein, B., Friston, K., Couzin, I., & Tschantz, A. (2022). Pymdp: A python library for active inference in discrete state spaces. *arXiv preprint arXiv:2201.03904*.

Impett, E. A., Gordon, A. M., Kogan, A., Oveis, C., Gable, S. L., & Keltner, D. (2010). Moving toward more perfect unions: Daily and long-term consequences of approach and avoidance goals in romantic relationships. *Journal of Personality and Social Psychology*, *99*(6), 948–963. https://doi.org/10.1037/a0020271

Kelley, H. H., & Thibaut, J. W. (1978). *Interpersonal relations: A theory of interdependence*. Wiley.

Kirchner, L., Eckert, A.-L., Berg, M., Endres, D., Straube, B., & Rief, W. (2024). An active inference approach to interpersonal differences in depression. *New Ideas in Psychology*, *74*, 101092. https://doi.org/10.1016/j.newideapsych.2024.101092

Lehmann, K., Bolis, D., Friston, K. J., Schilbach, L., Ramstead, M. J. D., & Kanske, P. (2024). An active-inference approach to second-person neuroscience. *Perspectives on Psychological Science*, *19*(6), 931–951. https://doi.org/10.1177/17456916231188000

Leo, K., Crenshaw, A. O., Hogan, J. N., Bourne, S. V., Baucom, K. J. W., & Baucom, B. R. W. (2021). A replication and extension of the interpersonal process model of demand/withdraw behavior: Incorporating subjective emotional experience. *Journal of Family Psychology*, *35*(4), 534–545. https://doi.org/10.1037/fam0000802

Leo, K., Leifker, F. R., Baucom, D. H., & Baucom, B. R. W. (2019). Conflict management and problem solving as relationship maintenance. In B. G. Ogolsky & J. K. Monk (Eds.), *Relationship maintenance: Theory, process, and context* (pp. 194–214). Cambridge University Press.

Moutoussis, M., Trujillo-Barreto, N. J., El-Deredy, W., Dolan, R. J., & Friston, K. J. (2014). A formal model of interpersonal inference. *Frontiers in Human Neuroscience*, *8*, 160. https://doi.org/10.3389/fnhum.2014.00160

Nasir, M., Baucom, B. R., Georgiou, P., & Narayanan, S. (2017). Predicting couple therapy outcomes based on speech acoustic features. *PLOS ONE*, *12*(9), e0185123. https://doi.org/10.1371/journal.pone.0185123

Olar, B. M., Treur, R. M., Hendrikse, S. C. F., & Treur, J. (2026). Relationship dynamics between anxious and avoidant attachment styles: Computational analysis and ai-generated video creation. *Cognitive Systems Research*, *97*, 101470. https://doi.org/10.1016/j.cogsys.2026.101470

Raghuveer, D., & Endres, D. (2023). Active inference and psychology of expectations: A study of formalizing violex. In C. L. Buckley, D. Cialfi, P. Lanillos, M. Ramstead, N. Sajid, H. Shimazaki, & T. Verbelen (Eds.), *Active inference* (pp. 235–250). Springer Nature Switzerland.

Roether, C. L., Omlor, L., Christensen, A., & Giese, M. A. (2009). Critical features for the perception of emotion from gait. *Journal of Vision*, *9*(6:15), 1–32. https://doi.org/10.1167/9.6.15

Rusbult, C. E., & Van Lange, P. A. M. (2003). Interdependence, interaction, and relationships. *Annual Review of Psychology*, *54*, 351–375. https://doi.org/10.1146/annurev.psych.54.101601.145059

Schrodt, P., Witt, P. L., & Shimkowski, J. R. (2014). A meta-analytical review of the demand/withdraw pattern of interaction and its associations with individual, relational, and communicative outcomes. *Communication Monographs*, *81*(1), 28–58. https://doi.org/10.1080/03637751.2013.813632

Smith, R., Friston, K. J., & Whyte, C. J. (2022). A step-by-step tutorial on active inference and its application to empirical data. *Journal of mathematical psychology*, *107*, 102632. https://doi.org/10.1016/j.jmp.2021.102632

Steele, J. S., & Ferrer, E. (2011). Latent differential equation modeling of self-regulatory and coregulatory affective processes. *Multivariate Behavioral Research*, *46*(6), 956–984. https://doi.org/10.1080/00273171.2011.625305

Steele, J. S., Ferrer, E., & Nesselroade, J. R. (2014). An idiographic approach to estimating models of dyadic interactions with differential equations. *Psychometrika*, *79*(4), 675–700. https://doi.org/10.1007/s11336-013-9366-9

VanderDrift, L. E., & Agnew, C. R. (2019). Interdependence perspectives on relationship maintenance. In B. G. Ogolsky & J. K. Monk (Eds.), *Relationship maintenance: Theory, process, and context* (pp. 15–28). Cambridge University Press. https://doi.org/10.1017/9781108304320.002

Wang, C., Chen, A., Bao, C., Jin, S., Chan, Y. K., Mindel, J. R., Xie, S., Swartz, H., Wu, T., Kraut, R. E., & Zhu, H. (2026). Modeling multi-party interaction in couples therapy: A multi-agent simulation approach [Preprint]. https://arxiv.org/abs/2601.10970

Wojda-Burlij, A. K., Baucom, D. H., Weber, D. M., & Christensen, A. (2024). The heterogeneity of intimate partner violence and its associations with distressed couples' emotional communication. *Journal of family psychology*, *38*(7), 1051–1062. https://doi.org/10.1037/fam0001242

Woodin, E. M. (2011). A two-dimensional approach to relationship conflict: Meta-analytic findings. *Journal of Family Psychology*, *25*(3), 325–335. https://doi.org/10.1037/a0023791