Toward Causal Understanding of Therapist-Client Relationships: A Study of Language Modality and Social Entrainment

Alexandria K. Vail
avail@cs.cmu.edu
Carnegie Mellon University
Pittsburgh, Pennsylvania, USA

Jeffrey M. Girard
jmgirard@ku.edu
University of Kansas
Lawrence, Kansas, USA

Lauren M. Bylsma
bylsmalm@upmc.edu
University of Pittsburgh
Pittsburgh, Pennsylvania, USA

Jeffrey F. Cohn
jeffcohn@pitt.edu
University of Pittsburgh
Pittsburgh, Pennsylvania, USA

Jay Fournier
jay.fournier@osumc.edu
The Ohio State University
Columbus, Ohio, USA

Holly A. Swartz
swartzha@upmc.edu
University of Pittsburgh
Pittsburgh, Pennsylvania, USA

Louis-Philippe Morency
morency@cs.cmu.edu
Carnegie Mellon University
Pittsburgh, Pennsylvania, USA

ABSTRACT

The relationship between a therapist and their client is one of the most critical determinants of successful therapy. The *working alliance* is a multifaceted concept capturing the collaborative aspect of the therapist-client relationship; a strong working alliance has been extensively linked to many positive therapeutic outcomes. Although therapy sessions are decidedly multimodal interactions, the language modality is of particular interest given its recognized relationship to similar dyadic concepts such as rapport, cooperation, and affiliation. Specifically, in this work we study *language entrainment*, which measures how much the therapist and client adapt toward each other's use of language over time. Despite the growing body of work in this area, however, relatively few studies examine *causal* relationships between human behavior and these relationship metrics: does an individual's perception of their partner affect how they speak, or does how they speak affect their perception? We explore these questions in this work through the use of structural equation modeling (SEM) techniques, which allow for both multilevel and temporal modeling of the relationship between the quality of the therapist-client working alliance and the participants' language entrainment. In our first experiment, we demonstrate that these techniques perform well in comparison to other common machine learning models, with the added benefits of interpretability and causal analysis. In our second analysis, we interpret the learned models to examine the relationship between working alliance and language entrainment and address our exploratory research questions. The results reveal that a therapist's language entrainment can have a significant impact on the client's perception of the working alliance, and that the client's language entrainment is a strong indicator of their perception of the working alliance. We discuss the implications of these results and consider several directions for future work in multimodality.

CCS CONCEPTS

• **Applied computing → Psychology**; *Health informatics.*

ACM Reference Format:
Alexandria K. Vail, Jeffrey M. Girard, Lauren M. Bylsma, Jeffrey F. Cohn, Jay Fournier, Holly A. Swartz, and Louis-Philippe Morency. 2022. Toward Causal Understanding of Therapist-Client Relationships: A Study of Language Modality and Social Entrainment. In *INTERNATIONAL CONFERENCE ON MULTIMODAL INTERACTION (ICMI '22), November 7–11, 2022, Bengaluru, India.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3536221.3556616

1 INTRODUCTION

Evidence suggests that the quality of the relationship between a client and their therapist is one of the most critical factors in determining treatment success [18, 28]. Concretely, much of the current psychological literature on the client-therapist relationship focuses on what is known as the *working alliance* [17]. This concept aims to capture the *collaborative* aspect of the therapist-client relationship. The working alliance is generally considered to consist of three components: agreement on the overall goal of the treatment, agreement on the tasks required to reach that goal, and the feeling of emotional bond between the participants. A positive working alliance between client and therapist plays a crucial role in fostering numerous positive therapeutic outcomes, including reduction of the client's symptoms and concerns [12, 17, 18], reduced drug abuse and recidivism [27] and improved medication compliance [11]. Of particular note is the recognized relationship between the quality of the working alliance and client dropout [11, 23, 37]. Proactive detection of a poor working alliance is especially valuable in this case: by the time a client has decided to quit therapy, the time for potential intervention has already passed. Understanding the complexity of the therapist-client relationship is crucial for informed treatment decision-making.

ICMI '22, November 7–11, 2022, Bengaluru, India

ACM ISBN 978-1-4503-9390-4/22/11.
https://doi.org/10.1145/3536221.3556616

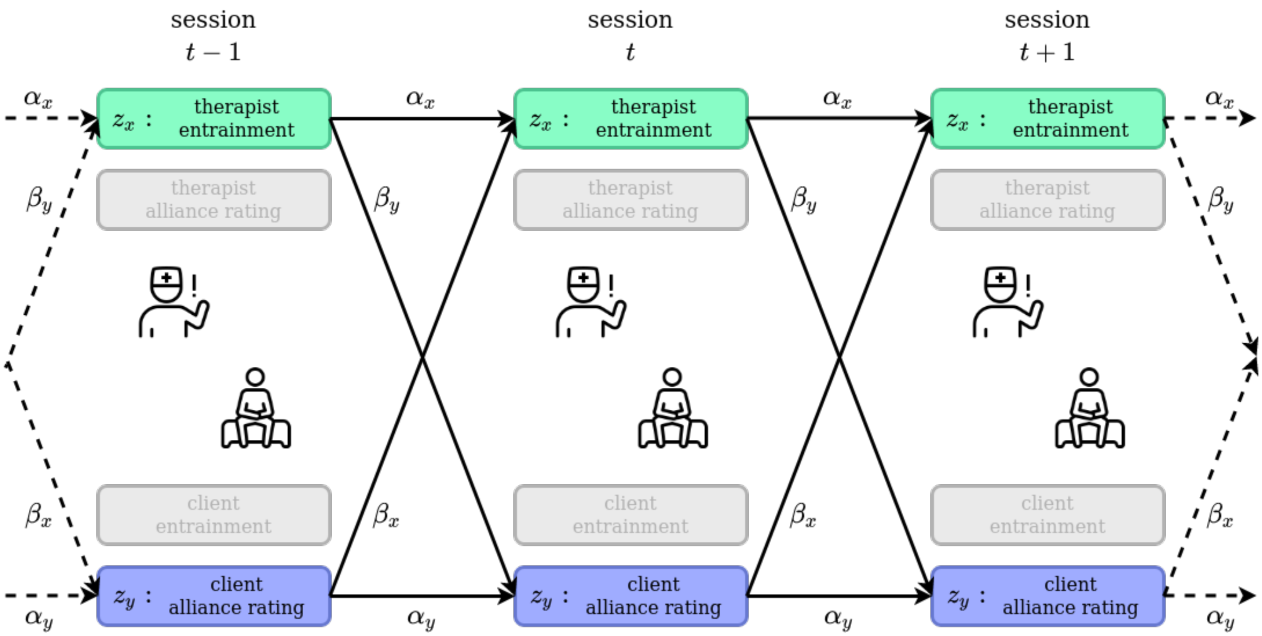


Figure 1: An example illustration of the overall structure of the therapist-entrainment/client-alliance analysis. During each session, we calculated an entrainment score (style and content) based on each participant's behavior, and at the conclusion of each session, each participant provided a rating of the working alliance (goal, task, and bond subscales). Edge labels (α_x, α_y, β_x, β_y) and node labels (z_x, z_y) correspond to the parameters introduced in Section 5 and Fig. 2. A similar structure was mirrored for the therapist-entrainment/therapist-alliance, client-entrainment/client-alliance, and client-entrainment/therapist-alliance analyses.

While working alliance and therapist-client relationships are decidedly multimodal concepts, the modality of language use is of particular interest given its importance in understanding similar forms of dyadic interaction [8, 21, 22, 32]. Relatively few studies have examined approaches for evaluating the working alliance beyond explicit questionnaires. More importantly, no previous work has studied the causal direction of the relationship between language and working alliance. Studying this relationship through the lens of causality allows us to go beyond correlation and address a broader range of research questions, such as the ones we focus on in this paper: does language behavior affect how the working alliance is perceived, or does working alliance perception affect how language is used?

This paper builds upon structural equation modeling (SEM) techniques to investigate the causal relationship between language use and working alliance. In particular, we introduce a specific method of structuring this model that allows us to study both relationships over time (temporal modeling) and patterns within individuals (multilevel modeling). Given the highly social nature of therapy sessions, we focus on *entrainment* in participant language. Linguistic entrainment is the process of multiple interlocutors (in our case, a client and their therapist) converging toward each other's use of language. We study linguistic entrainment in terms of both stylistic properties and content properties.

The overall structure of this paper consists of seven sections. In the next section, we review previous literature on behavior detection, working alliance, and linguistic entrainment (Section 2), and the following section provides a brief overview of the dataset used in this analysis (Section 3). Section 4 describes the definition and computation of our features and labels. We then devote Section 5 to an in-depth explanation of the SEM-based model we use in our analysis. The primary contributions of the paper lie in the next two sections: Section 6 evaluates the performance of this model in relation to other commonly-used modeling techniques, while Section 7 interprets the model's conclusions and discusses the implications of these results for behavioral research. The final section summarizes the main findings of this work and identifies areas for further research.

2 RELATED WORK

Interpersonal coordination is a behavioral phenomenon where multiple interacting persons adapt their behavior together over time [38], which can take many different forms [6]. Previous research has demonstrated that humans will coordinate their movements [3], voices [20, 34], and other communicative behaviors [26] to match each other during an interaction. A considerable amount of work has been published on the relationship between prosocial outcomes and behavioral coordination: increased interpersonal coordination during interaction leads to improved cooperation and collaboration [39], as well as higher self-reported ratings of rapport [36] and affiliation [19].

Despite this growing body of literature, relatively little work has focused on the role of interpersonal coordination in psychotherapy (cf. [1, 2, 33, 40]). Within this area of research, most prior work on therapy sessions has focused primarily on movement synchrony [24, 35]. In this analysis, we draw from related literature in social psychology that examines the role of *language entrainment* as a predictor of prosocial outcomes. Significant evidence exists to suggest that increased language style matching, in particular, leads to higher ratings of social intimacy, stability, and involvement [21, 22]. Language entrainment has also been linked to increased perception of support [32] and the general positivity of the interaction in question [8]. In long-term social relationships, language entrainment has also been shown to significantly predict child attachment security in parental relationships [5]. Inspired by this adjacent literature, this analysis explores whether language entrainment can also serve as a reliable and objective indicator of the quality of the therapeutic working alliance.

3 THERAPIST-CLIENT INTERACTION DATASET

Audiovisual recordings were collected from 266 therapy sessions between 39 unique clients and 11 unique therapists. Each therapist met with an average of 3.6 unique clients, and each client participated in an average of 6.8 sessions lasting between 40 and 60 minutes each (average 50.3 minutes).

Potential participants were recruited from a research registry, printed material advertising the study, and word-of-mouth. To be included in the study, participants had to be adults aged 18–65, meet DSM-5 criteria for a major depressive disorder[1], currently experience at least moderate depressive symptoms (as measured by a Hamilton Rating Scale for Depression score $\geq$ 14; [14]), and be willing and able to provide informed consent. Individuals with a comorbid psychotic disorder, active suicidal or homicidal ideation, chronic depression, or current substance or alcohol abuse were excluded from the study. If an individual was suspected of experiencing psychosis or active suicidal ideation with intent or plan to harm themselves, the investigator terminated the screening interview and ensured that the individual obtained appropriate care, including but not limited to a referral to the psychiatric emergency room.

Included clients ranged from 22 to 65 years of age; 77% identified as female, and 62% identified as White. Clients were randomly assigned to an eight-session brief course of one of two empirically-supported psychotherapy conditions: cognitive behavioral therapy (CBT; 21 clients, 6 therapists) or interpersonal psychotherapy (IPT; 18 clients, 5 therapists).[2]

[1]The Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5; [4]) is a taxonomy of psychiatric disorders published by the American Psychiatric Association. This manual serves as the primary diagnostic tool for psychiatric diagnosis and treatment in the United States.

[2]There were no statistically significant differences in working alliance ratings observed between the two treatment conditions.

4 LANGUAGE ENTRAINMENT AND WORKING ALLIANCE

4.1 Ratings of Working Alliance

Following the conclusion of each therapy session, both therapist and client participants completed the therapist and client versions of the revised short-form Working Alliance Inventory (WAI; [15]), a widely-used measure of alliance in therapy. The WAI consists of three subscales capturing three aspects of a working alliance:

- the *goal* subscale, which assesses the individual's belief that participants agree on the overall objectives of the treatment;
- the *task* subscale, which assesses the individual's belief that participants agree on the steps required to reach the goals mentioned above; and
- the *bond* subscale, which assesses the individual's respect and trust for the other participant in an emotional sense.

Each subscale consists of statements that the individual rates on a five-point Likert-type scale ranging from 'seldom true' to 'always true'; the inventory contains 12 items for the client and 10 items for the therapist. Representative items for each subscale are presented in Table 1.

4.2 Language Style and Content Metrics

Language entrainment is the process of multiple interlocutors adapting toward each other's use of language throughout an interaction. Although there exist many operational definitions to measure this construct, we leverage and expand upon a metric called *reciprocal linguistic style matching* (rLSM; [29]). The original definition of rLSM utilizes the Linguistic Inquiry and Word Count dictionary (LIWC; [30]), a well-validated and established lexicon that organizes approximately 6,400 English words into several semantically and/or functionally similar categories. In particular, we use LIWC "function word" categories: pronouns, articles, prepositions, auxiliary verbs, adverbs, conjunctions, and negations. Function words are useful to examine because they are independent of context, and their use is often less conscious. The benefit of rLSM over other metrics is the reciprocal component, which aims to measure how much the interlocutors change toward each other over time, rather than how much they may coincidentally speak with a similar style.

The rLSM score is initially calculated at the utterance level. Consider a therapist's response (T) to an utterance by the client (C): we aim to calculate the rLSM metric for the therapist's utterance. Since utterance T is a response to utterance C, we define rLSM_T as follows:

$$\text{rLSM}_T(S) = 1 - \frac{|S_C - S_T|}{S_C + S_T + 0.0001} \tag{1}$$

Here S represents any LIWC category score (e.g., negations) computed for client and therapist utterances (S_C and S_T, respectively). The total rLSM score for a statement is the average score of all function word categories. This score is calculated for each utterance during the session, and all utterance scores from each session are then averaged to determine each participant's session-level rLSM score.

We also propose an extension to rLSM, which studies the "content" component of language for contrast against the "style" component of language. We approximate this content component using

Goal Subscale	Task Subscale	Bond Subscale
[Therapist] and I collaborate on setting goals for my therapy.	What I am doing in therapy gives me new ways of looking at my problem.	I believe [Therapist] likes me.
[Therapist] and I have established a good understanding of the kind of changes that would be good for me.	[Therapist] and I agree on what is important for me to work on.	I feel that [Therapist] appreciates me.
We are working towards mutually agreed upon goals.	[Client] and I agree about the steps to be taken to improve his/her situation.	I feel [Therapist] cares about me even when I do things that he/she does not approve of.
[Client] and I have a common perception of his/her goals.	[Client] and I both feel confident about the usefulness of our current activity in therapy.	I appreciate [Client] as a person.
		[Client] and I respect each other.

Table 1: Sample items from both therapist and client versions of the Working Alliance Inventory.

the following LIWC categories: affective words; social words (family, friends); words relating to cognitive, perceptual, and biological processes (seeing, feeling; health); and words relating to motivation/drives and personal concerns (risk, reward; leisure, religion). We term this new metric rLCM — reciprocal linguistic content matching.

5 CAUSAL MODEL INTRODUCTION

Our model was designed with a number of desired principles in mind. First and foremost, we needed our model to be *interpretable*. Although the nature of the present analysis is primarily exploratory, we begin with some degree of expert domain knowledge and initial hypotheses as to the underlying structure of the data. For example, we expect that some individuals will adapt their language more than others ([31]; requiring multilevel modeling) and that working alliance ratings tend to increase over time ([16]; requiring temporal modeling).

In order to leverage these existing theoretical foundations, we turn to structural equation modeling (SEM) techniques [10]. SEM is a set of multivariate techniques that are generally confirmatory in nature, aiming to test whether a particular model structure fits a given dataset [25]. Unlike traditional machine learning models, SEM primarily leverages not the raw data provided to it but the covariance matrix: the goal is to minimize the distance between the observed and model-implied matrices. SEM also offers some advantages in our particular case. Given the additional overhead and sensitivity required to collect rich healthcare data, such as our own dataset introduced in Section 3, these healthcare datasets are often of a smaller size than those in other domains of multimodal research. In reducing the number of estimated variables by imposing a theoretical structure, SEM also allows us to explicitly account for the variance due to the inevitable measurement error present in psychological data. These features allow us to attain greater statistical power with fewer samples.

Given that we pursue the use of SEM for our analysis, we must design the underlying structure fundamental to these techniques. We intend to evaluate the relationship between the participants' perception of the working alliance and the adaptation of their language use toward their conversational partner, and in particular, the direction of this relationship: we are interested in *causality* in the data. Given the longitudinal nature of our dataset, the standard practice is to turn to the family of cross-lagged panel models (CLPMs; [7]). Finally, we must consider that our observations follow the same individuals over time, so we must also include consideration for participant-level patterns. We expect that participants will differ in their personal tendencies simply due to personality or other individual characteristics; for example, some people may be more inclined to adapt their language than others. This final consideration leads us to a modern hierarchical extension of the CLPM: random intercept cross-lagged panel modeling (RI-CLPM; [13]). The following subsections describe the intuitions and definitions of the classic CLPM (subsection 5.1) as well as the improvements and benefits introduced by the RI-CLPM extension (subsection 5.2).

5.1 Cross-Lagged Panel Modeling

Cross-lagged panel models (CLPM; [7]) involve the evaluation of the effect of two (or more) variables on each other over time. Consider x and y as two distinct variables (e.g., entrainment score and working alliance rating) from participant i measured over multiple time points (sessions) t. We aim to evaluate the relationship between x and y. The first important intuition behind CLPM techniques is the idea that a measured variable x (or y) is composed of a mean and a variation from that mean. This intuition can be formulated as follows (see Fig. 2a for an illustrated breakdown):

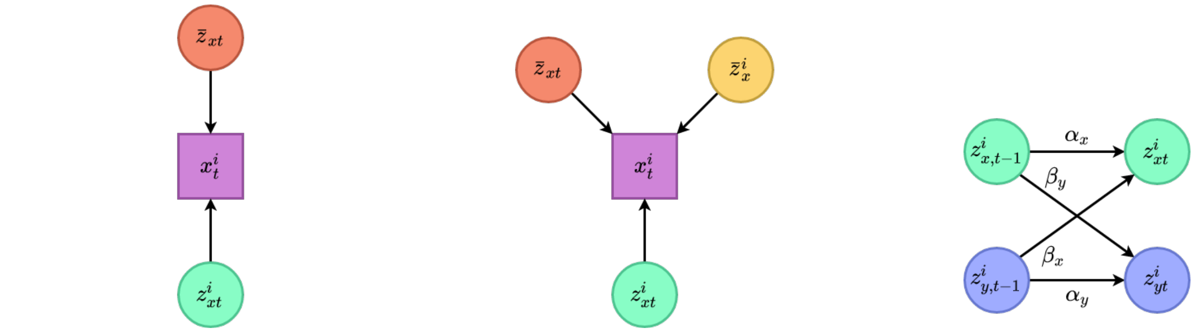
$$x_t^i = \bar{z}_{xt} + z_{xt}^i; \qquad y_t^i = \bar{z}_{yt} + z_{yt}^i; \tag{2}$$

where z_{xt}^i and z_{yt}^i represent the participant's temporal deviations from the temporal group means $\bar{z}_{xt}$ and $\bar{z}_{yt}$, respectively.

The second important intuition behind this model is that these temporal deviations z_{xt}^i are affected not only by previous temporal instances of itself, but also previous temporal instances of the other variable, z_{yt}^i; the same concept applies symmetrically for temporal variations of the other measured variable. This intuition is where the "cross-lagged" term in this approach originates. We can formally model these temporal deviations on the latent variables z_{xt}^i and z_{yt}^i as follows (Fig. 2c):

$$z_{xt}^i = \alpha_x z_{x,t-1}^i + \beta_x z_{y,t-1}^i + e_{xt}^i, \tag{3}$$

$$z_{yt}^i = \alpha_y z_{y,t-1}^i + \beta_y z_{x,t-1}^i + e_{yt}^i. \tag{4}$$



(a) $x_t^i = \bar{z}_{xt} + z_{xt}^i$ The CLPM measurement component for x_t^i. A similar structure is mirrored for y_t^i.

(b) $x_t^i = \bar{z}_{xt} + \bar{z}_x^i + z_{xt}^i$ The RI-CLPM measurement component for x_t^i. A similar structure is mirrored for y_t^i.

(c) $z_{xt}^i = \alpha_x z_{x,t-1}^i + \beta_x z_{y,t-1}^i + e_{xt}^i$ The temporal deviation component common to both CLPM and RI-CLPM.

Figure 2: A breakdown of the essential components of the CLPM and RI-CLPM techniques for a given session t.

The parameters α_x and α_y are autoregressive parameters that account for the temporal stability of these constructs: that is, the closer these parameters are to one, the more stable the rank order of individuals across time points. The parameters e_{xt}^i and e_{yt}^i represent residuals. The cross-lagged parameters β_x and β_y are fundamental to this family of models — by comparing the crossed effects of x on y (and vice versa), we can identify evidence to suggest the causal predominance of one direction over the other.

5.2 Random Intercept Cross-Lagged Panel Modeling

Following Hamaker et al. [13], we use an extension of CLPM that allows each participant to have their own individual variation compared to the group-level means $\bar{z}_{xt}$ and $\bar{z}_{yt}$. This model is named the random intercept cross-lagged panel model (RI-CLPM). RI-CLPM is a multilevel model where observations are nested within individuals. This model includes a random intercept that allows it to account not only for temporal stability, but also trait-level stability. With this in mind, Equation 2 can be rewritten as follows (see Fig. 2b for an illustrated breakdown):

$$x_t^i = \bar{z}_{xt} + \bar{z}_x^i + z_{xt}^i, \qquad y_t^i = \bar{z}_{yt} + \bar{z}_y^i + z_{yt}^i, \tag{5}$$

where the added parameters $\bar{z}_x^i$ and $\bar{z}_y^i$ represent the participant's individual trait-level deviations from the existing temporal group means. In this case, the parameters z_{xt}^i and z_{yt}^i now represent the participant's temporal deviations from their personalized expected scores (i.e., $\bar{z}_{xt} + \bar{z}_x^i$ and $\bar{z}_{yt} + \bar{z}_y^i$) as opposed to deviation from the temporal group mean (i.e., $\bar{z}_{xt}$ and $\bar{z}_{yt}$). We can now express these deviations as follows (Fig. 2c):

$$z_{xt}^i = \alpha_x z_{x,t-1}^i + \beta_x z_{y,t-1}^i + e_{xt}^i, \tag{6}$$

$$z_{yt}^i = \alpha_y z_{y,t-1}^i + \beta_y z_{x,t-1}^i + e_{yt}^i. \tag{7}$$

The autoregressive parameters α_x and α_y no longer represent merely the rank order of participants over time, but the degree of the within-person carry-over effect. For example, if this parameter is positive, it suggests that if a participant scored higher than their expected score at time point t, they are likely to also score higher than their expected score at time point $t + 1$.

One advantage of using the RI-CLPM over the CLPM is that it is effectively a generalization of the CLPM: if the additional elements are determined to be unnecessary, the additions tend toward zero and the model essentially 'collapses' to the base CLPM. Furthermore, in the case of the present analysis, we can reasonably assume that the effect the variables x and y have on each other over time remains stable: our observed time points are roughly evenly spaced, and we do not perform any midpoint 'intervention' that would suggest that any particular interval differs from the other intervals. As a result, we tie parameters (i.e., α and β) across time points, providing us with many more degrees of freedom in our model and parameters that are more straightforward to interpret.

6 PREDICTION EXPERIMENT

Our first set of experiments compares RI-CLPM performance against other commonly-used models, such as neural networks. As a reminder, an important goal when designing our model based on RI-CLPM was to leverage domain knowledge to reduce complexity and hopefully improve performance. Our model integrates inductive biases (domain knowledge) for both the temporal and the multilevel aspects of the data.

6.1 Baseline Models

We compare our model with several commonly-used machine learning models. We begin with neural networks: given the small number of data samples, we constrained ourselves to multilayer perceptrons. We included two variants with one or two hidden layers (MLP-1 and MLP-2, respectively). To study the relative importance of the two inductive biases we included in our model, we included as baselines a multilevel linear model (MLM) and the standard CLPM. The comparison with the CLPM allows us to evaluate the importance of including the random intercept component. All models were compared in terms of the performance of a simple linear model (LM), which can also perform effectively with small datasets.

6.2 Prediction Metrics

One of the challenges when evaluating all of these models is selecting a metric that will be fair and comparable across models.

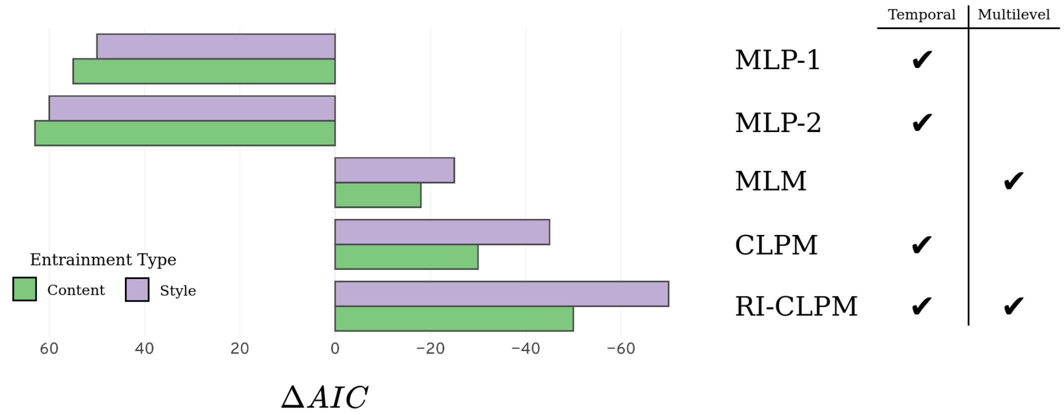


Figure 3: Comparative performance of baseline models relative to the linear model (LM). Note that AIC is a relative metric, and has no meaning in absolute terms: there are no "good" or "bad" AIC scores, only "better" or "worse" than another. Therefore, lower ΔAIC scores (further right in the chart) are better.

Although many commonly-used models (such as MLP models) are generally trained and evaluated in terms of their predictive performance (e.g., accuracy), SEM-based models have no directly corresponding notion of "prediction". Therefore, for this comparison, we rely on a metric revolving around *model fit*: Akaike's information criterion (AIC; [9]), which evaluates how well a given model's implied data matches a given dataset. Rather than providing an "absolute" score, it instead offers evidence for the preference of one model over a set of others: in other words, there are no "good" or "bad" AIC scores, only scores that are "better" or "worse" than that of another model. This metric can be expressed as follows:

$$\text{AIC} = 2k - 2\ln(\hat{L}), \tag{8}$$

where k is the number of estimated parameters in the model and $\hat{L}$ is the maximum value of its likelihood function.

6.3 Results and Discussion

Fig. 3 presents an overview of the performance of all of the models. Given that AIC is a relative metric, all scores are interpreted in terms of difference from the baseline model, the linear model. From this figure, it becomes apparent that the general pattern of better performance is achieved with the addition of temporal and multilevel elements — for such a relatively small but rich dataset, the importance of leveraging expert knowledge of both domain and dataset structure is evident.

7 LANGUAGE ANALYSIS

Our second set of experiments analyses the learned cross-lagged parameters (β_x and β_y) of the RI-CLPM model. Our goal is to study the relative effects of a participant's perception of the working alliance on their linguistic entrainment behavior. One benefit of our approach is the ability to distinguish directional effects — that is, whether working alliance perception affects linguistic entrainment, or if linguistic entrainment affects working alliance perception.

Working alliance ratings were collected from both client and therapist at the end of each session: these working alliance ratings are divided into agreement on goals, agreement on tasks, and agreement on bond. We also calculated both a stylistic entrainment score and a content entrainment score for each participant during each session (see Section 4 for more details on the calculation of these metrics). We fit an RI-CLPM to each combination of language behavior and working alliance ratings. From these fitted models, we primarily examine the cross-lagged parameters that estimate the relationship between the two measured variables: see Section 5 for more details on the model.

7.1 Results

Highlighted results are presented in Fig. 4. A number of significant effects can be observed from these results. In general, the client's perception of the working alliance results in an increase in their style and content entrainment (Fig. 4b). In particular, the client's perception of bond results in an increase in their stylistic entrainment, while their perception of the goal and task aspects of the working alliance result in an increase in their content entrainment.

From Fig. 4a, we can see that the client's perception of bond is significantly influenced by both content and stylistic linguistic entrainment on the part of the therapist. On the other hand, the therapist's perception of the working alliance appears less impacted by linguistic entrainment: the only significant association observed is that an increase in the client's content matching results in an increase in the therapist's perception of task agreement ($\beta = 0.1179$).

7.2 Discussion

The present analysis was designed to determine the effect of language entrainment during therapy sessions on the participants' perception of the working alliance, and vice versa. The results provide preliminary evidence to suggest a bidirectional but asymmetric relationship between these two constructs.

Stylistic entrainment is generally associated with perception of bond, while content matching is generally associated

(a) Estimated cross-lagged parameters (β) evaluating the effect of therapist entrainment behavior on client alliance ratings. ***How does the therapist's language affect the client's perception of the working alliance?***

(b) Estimated cross-lagged parameters (β) evaluating the effect of client alliance ratings on client entrainment behavior. ***How does the client's perception of the working alliance affect their language?***

Figure 4: Highlighted results from the language analysis described in Section 7. Asterisks (*) indicate parameters statistically significantly different from zero ($p < 0.05$).

with perception of task and goal. By examining working alliance ratings at this granular level, we can observe that stylistic entrainment seems associated mainly with the perception of bond. In contrast, content matching appears primarily associated with the perception of task and goal.

Therapy clients express their perception of the working alliance through linguistic entrainment. Perhaps the most compelling finding to emerge from this analysis is the suggestion that the client appears to demonstrate their current perception of the working alliance through their linguistic entrainment behavior, as seen in Fig. 4b.

Therapist linguistic entrainment has a notable impact on the client's perception of the working alliance bond. Finally, a third notable takeaway is that the therapist's language entrainment behavior seems to have a substantial impact on the client's perception of the working alliance, and particularly their impression of bond (Fig. 4a).

These results, particularly those discussed in the latter two points, also demonstrate the importance of considering causality when investigating these relationships. A model that explores only correlation, as most commonly-used models, would be unable to ascertain, for example, whether a client's linguistic entrainment affects their perception of the alliance or if their perception affects their entrainment.

8 CONCLUSION

The working alliance is a multifaceted concept that captures the collaborative aspect of the relationship between a therapist and their client. We use structural equation modeling (SEM) techniques to study the causal relationship between working alliance and language entrainment behaviors. We demonstrate that this kind of modeling can achieve excellent performance compared to other standard machine learning models, with the added benefit of interpretability and causal analysis. Interpretation of the model reveals valuable insights into the dyadic interaction between therapist and client during therapy. In general, the language entrainment of the therapist can have an impact on the client's perception of the alliance, and the client's perception of the alliance is often reflected in their own language use.

Future work includes exploring the relationship between working alliance and other social behaviors, such as gestures, prosody, and facial expression; bringing these modalities together into a multimodal approach is also of great interest. Examining the relationship between these behaviors and the alliance throughout a single interaction at a more granular level may also reveal exciting relationships. Such findings could eventually be implemented in the form of a computer-mediated feedback system, aiding the therapist in recognizing the deterioration of the working alliance in the moment and allowing for more immediate intervention to address client concerns. Multimodal behavior analysis in therapy has many promising future paths: the ensuing enhancement of therapeutic interaction will help ensure that more people seeking therapy receive the treatment they need.

ACKNOWLEDGEMENTS

This material is based upon work partially supported by U.S. National Science Foundation awards 1722822 and 1750439, and U.S. National Institute of Health awards R01MH125740, R01MH096951, and U01MH116925. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the sponsors, and no official endorsement should be inferred.

REFERENCES

[1] Katie Aafjes-van Doorn and Lena Müller-Frommeyer. 2020. Reciprocal Language Style Matching in Psychotherapy Research. *Counselling and Psychotherapy Research* 20, 3 (Sept. 2020), 449–455. https://doi.org/10.1002/capr.12298

[2] Katie Aafjes-van Doorn, John Porcerelli, and Lena Christine Müller-Frommeyer. 2020. Language Style Matching in Psychotherapy: An Implicit Aspect of Alliance. *Journal of Counseling Psychology* 67, 4 (July 2020), 509–522. https://doi.org/10.1037/cou0000433

[3] Drew H. Abney, Alexandra Paxton, Rick Dale, and Christopher T. Kello. 2015. Movement Dynamics Reflect a Functional Role for Weak Coupling and Role Structure in Dyadic Problem Solving. *Cognitive Processing* 16, 4 (Nov. 2015), 325–332. https://doi.org/10.1007/s10339-015-0648-2

[4] American Psychiatric Association. 2013. *Diagnostic and Statistical Manual of Mental Disorders* (fifth edition ed.). American Psychiatric Association. https://doi.org/10.1176/appi.books.9780890425596

[5] Jessica L. Borelli, Kizzann A. Ramsook, Patricia Smiley, David Kyle Bond, Jessica L. West, and Katherine H. Buttitta. 2017. Language Matching Among Mother-child Dyads: Associations with Child Attachment and Emotion Reactivity: Mother-child LSM. *Social Development* 26, 3 (Aug. 2017), 610–629. https://doi.org/10.1111/sode.12200

[6] Emily A. Butler. 2011. Temporal Interpersonal Emotion Systems: The "TIES" That Form Relationships. *Personality and Social Psychology Review* 15, 4 (Nov. 2011), 367–393. https://doi.org/10.1177/1088868311411164

[7] Donald T. Campbell and Julian C. Stanley. 1963. *Experimental and Quasi-Experimental Designs for Research.* Rand McNally, Chicago.

[8] Kaitlin Cannava and Graham D. Bodie. 2017. Language Use and Style Matching in Supportive Conversations between Strangers and Friends. *Journal of Social and Personal Relationships* 34, 4 (June 2017), 467–485. https://doi.org/10.1177/0265407516641222

[9] J. deLeeuw. 1992. Introduction to Akaike (1973) Information Theory and an Extension of the Maximum Likelihood Principle. In *Breakthroughs in Statistics*, Samuel Kotz and Norman L. Johnson (Eds.). Springer New York, New York, NY, 599–609. https://doi.org/10.1007/978-1-4612-0919-5_37

[10] Otis Dudley Duncan. 1975. *Introduction to Structural Equation Models.* Academic Press, New York.

[11] Arlene F. Frank. 1990. The Role of the Therapeutic Alliance in the Treatment of Schizophrenia: Relationship to Course and Outcome. *Archives of General Psychiatry* 47, 3 (March 1990), 228. https://doi.org/10.1001/archpsyc.1990.01810150028006

[12] Louise Gaston, Charles Marmar, Dolores Gallagher, and Larry Thompson. 1991. Alliance Prediction of Outcome Beyond In-Treatment Symptomatic Change as Psychotherapy Processes. *Psychotherapy Research* 1, 2 (Aug. 1991), 104–112. https://doi.org/10.1080/10503309112331335531

[13] Ellen L. Hamaker, Rebecca M. Kuiper, and Raoul P. P. P. Grasman. 2015. A Critique of the Cross-Lagged Panel Model. *Psychological Methods* 20, 1 (2015), 102–116. https://doi.org/10.1037/a0038889

[14] M. Hamilton. 1960. A Rating Scale for Depression. *Journal of Neurology, Neurosurgery & Psychiatry* 23, 1 (Feb. 1960), 56–62. https://doi.org/10.1136/jnnp.23.1.56

[15] Robert L. Hatcher and J. Arthur Gillaspy. 2006. Development and Validation of a Revised Short Version of the Working Alliance Inventory. *Psychotherapy Research* 16, 1 (Jan. 2006), 12–25. https://doi.org/10.1080/10503300500352500

[16] Anne Grete Hersoug, Jon T. Monsen, Odd E. Havik, and Per Høglend. 2002. Quality of Early Working Alliance in Psychotherapy: Diagnoses, Relationship and Intrapsychic Variables as Predictors. *Psychotherapy and Psychosomatics* 71, 1 (2002), 18–27. https://doi.org/10.1159/000049340

[17] Adam O. Horvath and Leslie S. Greenberg. 1986. The Development of the Working Alliance Inventory. In *The Psychotherapeutic Process: A Research Handbook.* Guilford Press, New York, NY, US, 529–556.

[18] Adam O. Horvath and B. Dianne Symonds. 1991. Relation between Working Alliance and Outcome in Psychotherapy: A Meta-Analysis. *Journal of Counseling Psychology* 38, 2 (1991), 139–149. https://doi.org/10.1037/0022-0167.38.2.139

[19] Michael J. Hove and Jane L. Risen. 2009. It's All in the Timing: Interpersonal Synchrony Increases Affiliation. *Social Cognition* 27, 6 (Dec. 2009), 949–960. https://doi.org/10.1521/soco.2009.27.6.949

[20] Zac E. Imel, Jacqueline S. Barco, Halley J. Brown, Brian R. Baucom, John S. Baer, John C. Kircher, and David C. Atkins. 2014. The Association of Therapist Empathy and Synchrony in Vocally Encoded Arousal. *Journal of Counseling Psychology* 61, 1 (2014), 146–153. https://doi.org/10.1037/a0034943

[21] Molly E. Ireland and James W. Pennebaker. 2010. Language Style Matching in Writing: Synchrony in Essays, Correspondence, and Poetry. *Journal of Personality and Social Psychology* 99, 3 (2010), 549–571. https://doi.org/10.1037/a0020386

[22] Molly E. Ireland, Richard B. Slatcher, Paul W. Eastwick, Lauren E. Scissors, Eli J. Finkel, and James W. Pennebaker. 2011. Language Style Matching Predicts Relationship Initiation and Stability. *Psychological Science* 22, 1 (Jan. 2011), 39–44. https://doi.org/10.1177/0956797610392928

[23] Anna M. Kokotovic and Terence J. Tracey. 1990. Working Alliance in the Early Phase of Counseling. *Journal of Counseling Psychology* 37, 1 (1990), 16–21. https://doi.org/10.1037/0022-0167.37.1.16

[24] Sander L. Koole and Wolfgang Tschacher. 2016. Synchrony in Psychotherapy: A Review and an Integrative Framework for the Therapeutic Alliance. *Frontiers in Psychology* 7 (June 2016), 1–17. https://doi.org/10.3389/fpsyg.2016.00862

[25] Brett Paul Laursen, Todd D. Little, and Noel A. Card (Eds.). 2012. *Handbook of Developmental Research Methods.* Guilford Press, New York, NY.

[26] Max M. Louwerse, Rick Dale, Ellen G. Bard, and Patrick Jeuniaux. 2012. Behavior Matching in Multimodal Communication Is Synchronized. *Cognitive Science* 36, 8 (Nov. 2012), 1404–1426. https://doi.org/10.1111/j.1551-6709.2012.01269.x

[27] Lester Luborsky. 1985. Therapist Success and Its Determinants. *Archives of General Psychiatry* 42, 6 (June 1985), 602. https://doi.org/10.1001/archpsyc.1985.01790290084010

[28] Daniel J. Martin, John P. Garske, and M. Katherine Davis. 2000. Relation of the Therapeutic Alliance with Outcome and Other Variables: A Meta-Analytic Review. *Journal of Consulting and Clinical Psychology* 68, 3 (2000), 438–450. https://doi.org/10.1037/0022-006X.68.3.438

[29] Lena C. Müller-Frommeyer, Niels A. M. Frommeyer, and Simone Kauffeld. 2019. Introducing rLSM: An Integrated Metric Assessing Temporal Reciprocity in Language Style Matching. *Behavior Research Methods* 51, 3 (June 2019), 1343–1359. https://doi.org/10.3758/s13428-018-1078-8

[30] James W. Pennebaker, Ryan L. Boyd, Kayla Jordan, and Kate Blackburn. 2015. The Development and Psychometric Properties of LIWC2015. (2015). https://doi.org/10.15781/T25P41

[31] Bernd Puschner, Stephanie Bauer, Leonard M. Horowitz, and Hans Kordy. 2005. The Relationship between Interpersonal Problems and the Helping Alliance. *Journal of Clinical Psychology* 61, 4 (April 2005), 415–429. https://doi.org/10.1002/jclp.20050

[32] Stephen A. Rains. 2016. Language Style Matching as a Predictor of Perceived Social Support in Computer-Mediated Interaction Among Individuals Coping With Illness. *Communication Research* 43, 5 (July 2016), 694–712. https://doi.org/10.1177/0093650214565920

[33] Fabian Ramseyer and Wolfgang Tschacher. 2011. Nonverbal Synchrony in Psychotherapy: Coordinated Body Movement Reflects Relationship Quality and Outcome. *Journal of Consulting and Clinical Psychology* 79, 3 (2011), 284–295. https://doi.org/10.1037/a0023419

[34] Catherine M. Reich, Jeffrey S. Berman, Rick Dale, and Heidi M. Levitt. 2014. Vocal Synchrony in Psychotherapy. *Journal of Social and Clinical Psychology* 33, 5 (May 2014), 481–494. https://doi.org/10.1521/jscp.2014.33.5.481

[35] Katharina C. H. Reinecke, Peter Joraschky, and Hedda Lausberg. 2022. Hand Movements That Change during Psychotherapy and Their Relation to Therapeutic Outcome: An Analysis of Individual and Simultaneous Movements. *Psychotherapy Research* 32, 1 (Jan. 2022), 104–114. https://doi.org/10.1080/10503307.2021.1925989

[36] Miriam Rennung and Anja S. Göritz. 2016. Prosocial Consequences of Interpersonal Synchrony: A Meta-Analysis. *Zeitschrift für Psychologie* 224, 3 (July 2016), 168–189. https://doi.org/10.1027/2151-2604/a000252

[37] Stephen M. Saunders, Kenneth I. Howard, and David E. Orlinsky. 1989. The Therapeutic Bond Scales: Psychometric Characteristics and Relationship to Treatment Effectiveness. *Psychological Assessment: A Journal of Consulting and Clinical Psychology* 1, 4 (1989), 323–330. https://doi.org/10.1037/1040-3590.1.4.323

[38] M. T. Turvey. 1990. Coordination. *American Psychologist* 45, 8 (1990), 938–953. https://doi.org/10.1037/0003-066X.45.8.938

[39] Scott S. Wiltermuth and Chip Heath. 2009. Synchrony and Cooperation. *Psychological Science* 20, 1 (Jan. 2009), 1–5. https://doi.org/10.1111/j.1467-9280.2008.02253.x

[40] Travis J. Wiltshire, Johanne Stege Philipsen, Sarah Bro Trasmundi, Thomas Wiben Jensen, and Sune Vork Steffensen. 2020. Interpersonal Coordination Dynamics in Psychotherapy: A Systematic Review. *Cognitive Therapy and Research* 44, 4 (Aug. 2020), 752–773. https://doi.org/10.1007/s10608-020-10106-3