Taking a Turn for The Better:
Conversation Redirection Throughout the Course of Mental-health Therapy

Vivian Nguyen, Sang Min Jung, Lillian Lee, Thomas D. Hull, Cristian Danescu-Niculescu-Mizil

Try it on your data ConvoKit

Conversation Flow: a continuous negotiation

At every **utterance juncture**, participants engage in a **joint act** that determines where the conversation will be heading next: either keeping the same course, or **redirecting** the discussion.

Cf. discourse-level notion of shift that does not require joint acts (e.g., topic shift)

Redirection in Mental-health Therapy

Mental-health therapy involves a complex conversation flow, e.g., **therapists** may redirect to keep the therapeutic process on track, **patients** may redirect to switch focus to recent issues.

The **joint** nature of redirection makes it well suited for examining the **patient-therapist relationships:**
their long-term development and quality.

Do patients *steer* more in **successful** relations?

Measuring Redirection

redirection = **redirection attempt** + **acceptance**

Needs: **point of reference**

Redirection of **therapist's** k'th utterance $\mathbf{R(T_k)}$
(note in general both parties can redirect)

1. Q_k : Likelihood of **patient's** reply $\boldsymbol{C_{k+1}}$ in a hypothetical scenario where **therapist** attempts no redirection by **repeating** their previous utterance $\boldsymbol{T_{k-1}}$ (used as a **point of reference**):

$$\mathcal{Q}_k(C_{k+1}) \triangleq P(C_{k+1}|C_k, T_{k-1})$$

2. Compare to P_k : the likelihood of **reply** C_{k+1} given **therapist actual** utterance $\boldsymbol{T_k}$:

$$\mathcal{P}_k(C_{k+1}) \triangleq P(C_{k+1}|C_k, T_k)$$

Intuition: If $P_k >> Q_k$, then:
$\mathbf{T_{k-1}}$ must be quite different from $\mathbf{T_k}$,
so there was a **redirection attempt** and
$\mathbf{C_{k+1}}$ is much more likely as a reply to $\mathbf{T_k}$ than to (the hypothetical) $\mathbf{T_{k-1}}$,
so the patient **accepts the redirection attempt**.

If $P_k << Q_k$, then the **redirection attempt was rejected**.

If there was no redirection attempt, then $P_k \approx Q_k$

3. *Redirection $R(T_k)$ is then* the log-odds ratio of P_k and Q_k

$$R(T_k) \triangleq \log\left(\frac{\mathcal{P}_k(C_{k+1})}{1-\mathcal{P}_k(C_{k+1})} \Big/ \frac{\mathcal{Q}_k(C_{k+1})}{1-\mathcal{Q}_k(C_{k+1})}\right)$$

Probabilities computed by fine-tuning the Gemma-2B model (Mesnard et al. 2024)

Comparison to Related Measures

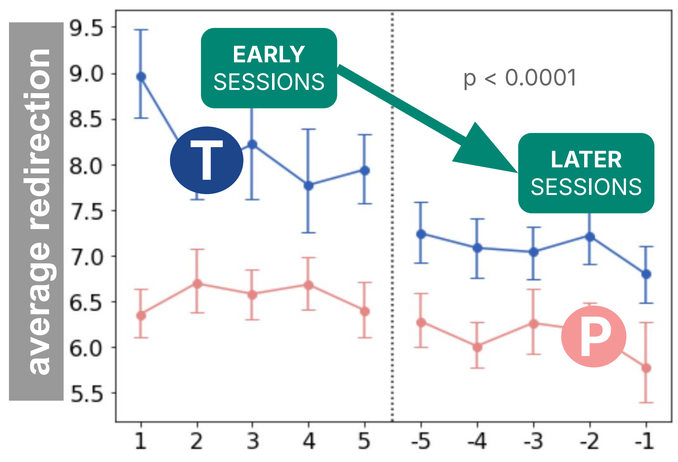
Orientation (Zhang & DNM '20) Captures *only the attempt* to shift the focus of the conversation

Similarity Difference Captures *only similarity (≠ probability)* between the reply and previous utterances

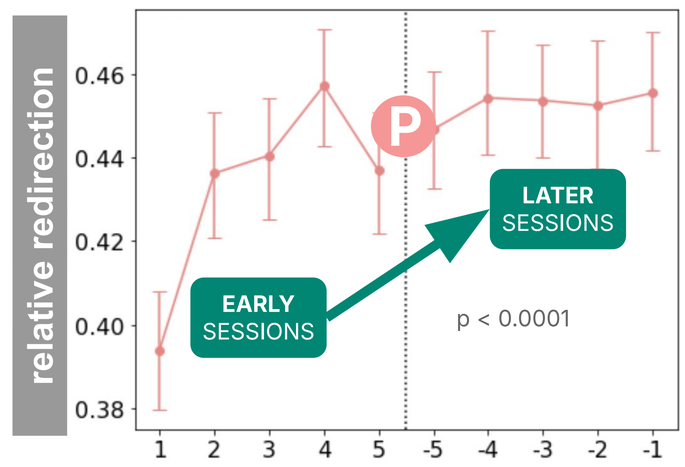
Uptake (Demszky et al. '21) Captures *only the probability* between the immediate utterance and the reply, does not *take a contextual point of reference,* but a random one.

Human intuition of redirection aligns more closely with our measure
Empirically, none of these distinguish unsuccessful therapies.

Evolution of the Therapeutic Relation



Both **therapist** and **patient's** **average redirection** **decreases ↓** as therapy progresses, indicating a more **focused** and **natural** conversation flow.



Patient **relative redirection** **increases ↑** relative to the **therapist**, suggesting **patients** **gain more control** over the conversation from the therapist.

The difference disappears after shuffling utterance order, indicating it's **tied to conversation dynamics**.

How About Unsuccessful Relationships?

Some patients eventually express dissatisfaction with their therapist *and* ask to switch therapists or cancel.

EARLY SESSIONS	Unsuccessful relationships		Control	p-Value
Actual	6.06	<	6.91	**0.02***
Shuffled	0.14	≈	0.04	0.17

Patients **redirect less** in the early sessions of such **unsuccessful** relationships. This shows the importance of **patient agency** and of **therapist willingness to follow** the patient's lead.

Reduced patient redirection is not due to lack of patient redirection attempts (orientation is the same), but rather to the therapist unwillingness to accept those attempts.

MARCH 2
MARCH 9
APRIL 16
MAY 12

Cornell Talkspace