Balancing Objectives in Counseling Conversations: Advancing Forwards or Looking Backwards

Justine Zhang
Cornell University
jz727@cornell.edu

Cristian Danescu-Niculescu-Mizil
Cornell University
cristian@cs.cornell.edu

Abstract

Throughout a conversation, participants make choices that can orient the flow of the interaction. Such choices are particularly salient in the consequential domain of crisis counseling, where a difficulty for counselors is balancing between two key objectives: advancing the conversation towards a resolution, and empathetically addressing the crisis situation.

In this work, we develop an unsupervised methodology to quantify how counselors manage this balance. Our main intuition is that if an utterance can only receive a narrow range of appropriate replies, then its likely aim is to advance the conversation forwards, towards a target within that range. Likewise, an utterance that can only appropriately follow a narrow range of possible utterances is likely aimed backwards at addressing a specific situation within that range. By applying this intuition, we can map each utterance to a continuous *orientation* axis that captures the degree to which it is intended to direct the flow of the conversation forwards or backwards.

This unsupervised method allows us to characterize counselor behaviors in a large dataset of crisis counseling conversations, where we show that known counseling strategies intuitively align with this axis. We also illustrate how our measure can be indicative of a conversation's progress, as well as its effectiveness.

1 Introduction

Participants in a conversation constantly shape the flow of the interaction through their choices. In psychological crisis counseling conversations, where counselors support individuals in mental distress, these choices arise in uniquely complex and high-stakes circumstances, and are reflected in rich conversational dynamics (Sacks, 1992). As such, counseling is a valuable context for computationally modeling conversational behavior (Atkins et al., 2014; Althoff et al., 2016; Pérez-Rosas et al., 2018; Zhang et al., 2019). Modeling the conversational choices of counselors in this endeavor is an important step towards better supporting them.

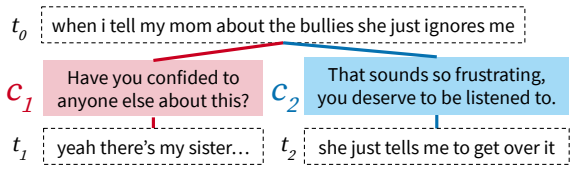

Figure 1: Two possible exchanges in a counseling conversation, illustrating key objectives that a counselor must balance: c_1 aims to *advance* the conversation towards a discussion of possible confidants; c_2 aims to *address* the emotion underlying the preceding utterance.

Counselors are driven by several objectives that serve the broader goal of helping the individual in distress; two key objectives are exemplified in Figure 1.[1] The counselor must *advance* a conversation towards a calmer state where the individual is better equipped to cope with their situation (Mishara et al., 2007; Sandoval et al., 2009): in c_1, the counselor prompts the individual to brainstorm options for social support. The counselor must also empathetically *address* what was already said, "coming to an empathic understanding" of the individual (Rogers, 1957; Hill and Nakayama, 2000): in c_2, the counselor validates feelings that the individual has just shared.

Balancing both objectives is often challenging, and overshooting in one direction can be detrimental to the conversation. A counselor who leans too much on advancing *forwards* could rush the conversation at the expense of establishing an empathetic connection; a counselor who leans too much *backwards*, on addressing what was already said, may fail to make any progress.

[1]These examples are derived from material used to train counselors in our particular setting, detailed in Section 2.

In this work, we develop a method to examine counselor behaviors as they relate to this balancing challenge. We quantify the relative extent to which an utterance is aimed at advancing the conversation, versus addressing existing content. We thus map each utterance onto a continuous backwards-forwards axis which models the balance of these objectives, and refer to an utterance's position on this axis as its *orientation*.

At an intuitive level, our approach considers the range of content that is expected to follow or precede a particular utterance. For an utterance like c_1 that aims to advance the conversation towards an intended target, we would expect a narrow range of appropriate replies, concentrated around that target (e.g., suggestions of possible confidants). We would likewise expect an utterance like c_2 that aims to address a previously-discussed situation to only be an appropriate reply for a narrow range of possible utterances, concentrated around that specific type of situation (e.g., disclosures of negative feelings). Starting from this intuition, we develop an unsupervised method to quantify and compare these expected forwards and backwards ranges for any utterance, yielding our orientation measure.

Using this measure, we characterize counselor behaviors in a large collection of text-message conversations from a crisis counseling service, which we accessed in collaboration with the service and with the participants' consent. We show how orientation meaningfully distinguishes between key conversational strategies that counselors are taught during their training. We also show that our measure tracks a conversation's progress and can signal its effectiveness, highlighting the importance of balancing the advancing and addressing objectives, and laying the basis for future inquiries in establishing potential causal effects.

In summary, we develop an unsupervised methodology that captures how counselors balance the conversational objectives of advancing and addressing (Section 4), apply and validate it in a large dataset of counseling conversations (Section 5), and use it to investigate the relation between a counselor's conversational behavior and their effectiveness (Section 5.4). While our method is motivated by a salient challenge in counseling, we expect similar balancing problems to recur in other conversational settings where participants must carefully direct the flow of the interaction, such as court trials and debates (Section 6).

2 Setting: Counseling Conversations

We develop our method in the context of Crisis Text Line, a crisis counseling platform which provides a free 24/7 service for anyone in mental crisis—henceforth *texters*—to have one-on-one conversations via text message with affiliated counselors. We accessed a version of this collection, with over 1.5 million conversations, in collaboration with the platform and with the consent of the participants. The data was scrubbed of personally identifiable information by the platform.[2] These conversations are quite substantive, averaging 25 messages with 29 and 24 words per counselor and texter message, respectively.

In each conversation, a crisis counselor's high-level goal is to guide the texter towards a calmer mental state. In service of this goal, all counselors first complete 30 hours of training provided by the platform, which draws on past literature in counseling to recommend best practices and conversational strategies. The first author also completed the training to gain familiarity with the domain.

While the platform offers guidance to counselors, their task is inevitably open-ended, given the emotional complexity of crisis situations. As such, the counselors are motivated by an explicit goal that structures the interaction, but they face a challenging flexibility in choosing how to act.

3 Background and Related Work

We now describe the conversational challenge of balancing between advancing the conversation forwards or addressing what was previously said. Our description of the challenge and our computational approach to studying it are informed by literature in counseling, on the platform's training material and on informal interviews with its staff.

A conversational balance. A crisis counselor must fulfill multiple objectives in their broader goal of helping a texter. One objective is guiding the texter through their initial distress to a calmer mental state (Mishara et al., 2007; Sandoval et al., 2009), as in Figure 1, c_1. Various strategies that aim to facilitate this *advancing* process are taught to counselors during training: for instance, a counselor may prompt a texter to identify a goal or cop-

[2]The data can be accessed by applying at https://www.crisistextline.org/data-philosophy/data-fellows/. The extensive ethical and privacy considerations, and policies accordingly implemented by the platform, are detailed in Pisani et al. (2019).

ing mechanism (Rollnick and Miller, 1995). As such, they attempt to move the conversation *forwards*, towards its eventual resolution.

The counselor must also engage with the texter's concerns (Rogers, 1957; Hill and Nakayama, 2000), as in c_2, via strategies that empathetically *address* what the texter has already shared (Rollnick and Miller, 1995; Weger et al., 2010; Bodie et al., 2015). For instance, counselors are taught to *reflect*, i.e., reframe a texter's previous message to convey understanding, or draw on what was said to affirm the texter's positive qualities. In doing so, the counselor looks *backwards* in the conversation.

Past work has posited the benefits of mixing between strategies that aim at either objective (Mishara et al., 2007). However, as the training acknowledges, striking this balance is challenging. Overzealously seeking to advance could cut short the process of establishing an empathetic connection. Conversely, focusing on the conversation's past may not help with eventual problem solving (Bodie et al., 2015), and risks stalling it. A texter may start to counterproductively rehash or *ruminate* on their concerns (Nolen-Hoeksema et al., 2008; Jones et al., 2009); indeed, prior psychological work has highlighted the thin line between productive reflection and rumination (Rose et al., 2007; Landphair and Preddy, 2012).

Orientation. To examine this balancing dynamic, we model the choices that counselors make at each turn in a conversation. Our approach is to derive a continuous axis spanned by advancing and addressing. We refer to an utterance's position on this axis, representing the relative extent to which it aims at either objective, as its *orientation* Ω. We interpret a *forwards-oriented* utterance with positive Ω as aiming to advance the conversation, and a *backwards-oriented* utterance with negative Ω as aiming to address what was previously brought up. In the middle, the axis reflects the graded way in which a counselor can balance between aims—for instance, using something the texter has previously said to help motivate a problem-solving strategy.

Related characterizations. While we develop orientation to model a dynamic in counseling, we view it as a complement to other characterizations of conversational behaviors in varied settings.

Prior work has similarly considered how utterances relate to the preceding and subsequent discourse (Webber, 2001). Frameworks like centering theory (Grosz et al., 1995) aim at identifying referenced entities, while we aim to more abstractly model interlocutor choices. Past work has also examined how interlocutors mediate a conversation's trajectory through taking or ceding control (Walker and Whittaker, 1990) or shifting topic (Nguyen et al., 2014); Althoff et al. (2016) considers the rate at which counselors in our setting advance across stages of a conversation. While these actions can be construed as forwards-oriented, we focus more on the interplay between forwards- and backwards-oriented actions. A counselor's objectives may also cut across these concepts: for instance, the training stresses the need for empathetic reflecting across all stages and topics.

Orientation also complements prior work on dialogue acts, which consider various roles that utterances play in discourse (Mann and Thompson, 1988; Core and Allen, 1997; Ritter et al., 2010; Bracewell et al., 2012; Rosenthal and McKeown, 2015; Prabhakaran et al., 2018; Wang et al., 2019). In counseling settings, such approaches have highlighted strategies like reflection and question-asking (Houck, 2008; Gaume et al., 2010; Atkins et al., 2014; Can et al., 2015; Tanana et al., 2016; Pérez-Rosas et al., 2017, 2018; Park et al., 2019; Lee et al., 2019; Cao et al., 2019). Instead of modeling a particular taxonomy of actions, we model how counselors balance among the underlying objectives; we later relate orientation to these strategies (Section 5). Most of these approaches use annotations or predefined labeling schemes, while our method is unsupervised.

4 Measuring Orientation

We now describe our method to measure orientation, discussing our approach at a high level before elaborating on the particular operationalization. The code implementing our approach is distributed as part of the ConvoKit library (Chang et al., 2020), at `http://convokit.cornell.edu`.

4.1 High-Level Sketch

Orientation compares the extent to which an utterance aims to advance the conversation forwards with the extent to which it looks backwards. Thus, we must somehow quantify how the utterance relates to the subsequent and preceding interaction.

Naive attempt: direct comparison. As a natural starting point, we may opt for a similarity-based approach: an utterance that aims to address its preceding utterance, or *predecessor*, should be similar

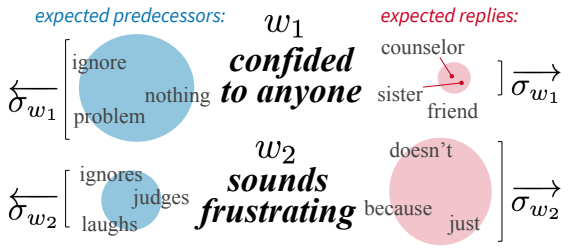

Figure 2: Words representative of replies and predecessors for utterances with two example phrasings, as observed in training data. Top row: observed replies to utterances with w_1 span a narrower range than observed predecessors (relative sizes of red and blue circles); w_1 thus has smaller *forwards-range* $\overrightarrow{\sigma_{w_1}}$ than *backwards-range* $\overleftarrow{\sigma_{w_1}}$ (i.e., it is forwards-oriented, $\Omega_{w_1} > 0$). Bottom row: observed predecessors to utterances with w_2 span a narrower range than replies; w_2 thus has smaller $\overleftarrow{\sigma_{w_2}}$ than $\overrightarrow{\sigma_{w_2}}$ (i.e., it is backwards-oriented $\Omega_{w_2} < 0$).

to it; an utterance that aims to advance the conversation should be similar to the reply that it prompts. In practice, having to make these direct comparisons is limiting: an automated system could not characterize an utterance in an ongoing conversation by comparing it to a reply it has yet to receive.

This approach also has important conceptual faults. First, addressing preceding content in a conversation is different from recapitulating it. For instance, counselors are instructed to *reframe* rather than outright restate a texter's message, as in Figure 1, c_2. Likewise, counselors need not advance the conversation by declaring something for the texter to simply repeat; rather than giving specific recommendations, counselors are instructed to prompt the texters to come up with coping strategies on their own, as in c_1. Further, texters are not bound to the relatively formal linguistic style counselors must maintain, resulting in clear lexical differences. Measuring orientation is hence a distinct task from measuring similarity.

Second, an utterance's *intent* to advance need not actually be realized. A counselor's cues may be rebuffed or misunderstood (Schegloff, 1987; Thomas, 1983): a texter could respond to c_1 by continuing to articulate their problem with t_2. Likewise, a counselor may intend to address a texter's concerns but misinterpret them. To model the balance in objectives that a counselor is aiming for, our characterization of an utterance cannot be contingent on its actual reply and predecessor.

Our approach: characterizing expectations. We instead consider the range of replies we might *expect* an utterance to receive, or the range of predecessors that it might follow. Intuitively, an utterance with a narrow range of appropriate replies aims to direct the conversation towards a particular target, moreso than an utterance whose appropriate replies span a broader range.[3] Likewise, an utterance that is an appropriate reply to only a narrow range of possible predecessors aims to address a particular situation. We draw on unlabeled data of past conversations to form our expectations of these ranges, and build up our characterizations of utterances from their constituent *phrasings*, e.g., words or dependency-parse arcs.

The intuition for our approach is sketched in Figure 2. From our data, we observe that utterances containing *confided to anyone* generally elicited replies about potential confidants (e.g., *sister*, *friend*), while the replies that followed utterances with *sounds frustrating* span a broader, less well-defined range. As such, we have a stronger expectation of what a reply prompted by a *new* utterance with *confided to anyone* might contain than a reply to a new utterance with *sounds frustrating*. More generally, for each phrasing w, we quantify the strength of our expectations of its potential replies by measuring the range spanned by the replies it has already received in the data, which we refer to as its *forwards-range* $\overrightarrow{\sigma_w}$. We would say that *confided to anyone* has a smaller $\overrightarrow{\sigma_w}$ than *sounds frustrating*, meaning that its observed replies were more narrowly concentrated; this is represented as the relative size of the red regions on the right side of Figure 2.

In the other direction, we observe in our data that *sounds frustrating* generally followed descriptions of frustrating situations (e.g., *ignores*, *judges*), while the range of predecessors to *confided to anyone* is broader. We thus have a stronger expectation of the types of situations that new utterances with *sounds frustrating* would respond to, compared to new utterances with *confided to anyone*. For a phrasing w, we quantify the strength of our expectations of its potential predecessors by measuring its *backwards-range* $\overleftarrow{\sigma_w}$, spanned by the predecessors we've observed. As such, *sounds frustrating* has a smaller $\overleftarrow{\sigma_w}$ than *confided to anyone*, corresponding to the relative size of the blue regions on the left side of Figure 2.

[3]Consider leading versus open-ended questions. When people ask leading questions, they intend to direct the interaction towards specific answers they have in mind; when people ask open-ended questions, they are more open to what answers they receive and where the interaction is headed.

The relative strengths of our expectations in either direction then indicate the balance of objectives. If we have a stronger expectation of w's replies than of its predecessors—i.e., smaller $\overrightarrow{\sigma_w}$ than $\overleftarrow{\sigma_w}$—we would infer that utterances with w aim to advance the conversation towards a targeted reply more than they aim to address a particular situation. Conversely, if we have stronger expectations of w's predecessors—i.e., smaller $\overleftarrow{\sigma_w}$—we would infer that utterances with w aim to address the preceding interaction, rather than trying to drive the conversation towards some target.

We thus measure orientation by comparing a phrasing's forwards- and backwards-range. The expectation-based approach allows us to circumvent the shortcomings of a direct comparison; we may interpret it as modeling a counselor's *intent* in advancing and addressing at each utterance (Moore and Paris, 1993; Zhang et al., 2017).

4.2 Operationalization

We now detail the steps of our method, which are outlined in Figure 3. Formally, our input consists of a set of utterances from counselors $\{c_i\}$, and a set of utterances from texters $\{t_i\}$, which we've observed in a dataset of conversations (Figure 3A). We note that each texter utterance can be a reply to, or a predecessor of, a counselor utterance (or both). We use this unlabeled "training data" to measure the forwards-range $\overrightarrow{\sigma_w}$, the backwards-range $\overleftarrow{\sigma_w}$ (Figures 3B-D), and hence the orientation Ω_w of each phrasing w used by counselors (Figure 3E). We then aggregate to an utterance-level measure.

For each counselor phrasing w, let $\overrightarrow{T_w}$ denote the subset of texter utterances which are replies to counselor utterances containing w (Figure 3A). As described above, the forwards-range $\overrightarrow{\sigma_w}$ quantifies the spread among elements of $\overrightarrow{T_w}$; we measure this by deriving vector representations of these utterances $\overrightarrow{U_w}$ (Figure 3B, detailed below), and then comparing each vector in $\overrightarrow{U_w}$ to a central reference point $\overrightarrow{u_w}$ (Figures 3C and 3D).[4] Likewise, $\overleftarrow{\sigma_w}$ quantifies the similarity among elements of $\overleftarrow{T_w}$, the set of predecessors to counselor utterances with w; we compute $\overleftarrow{\sigma_w}$ by comparing each corresponding vector in $\overleftarrow{U_w}$ to a central point $\overleftarrow{u_w}$.

[4]Using a central reference point to calculate the forwards-range, as opposed to directly computing pairwise similarities among replies in $\overrightarrow{U_w}$, allows us to account for the context of w in the utterances that prompted these replies (via tf-idf weighting, as subsequently discussed).

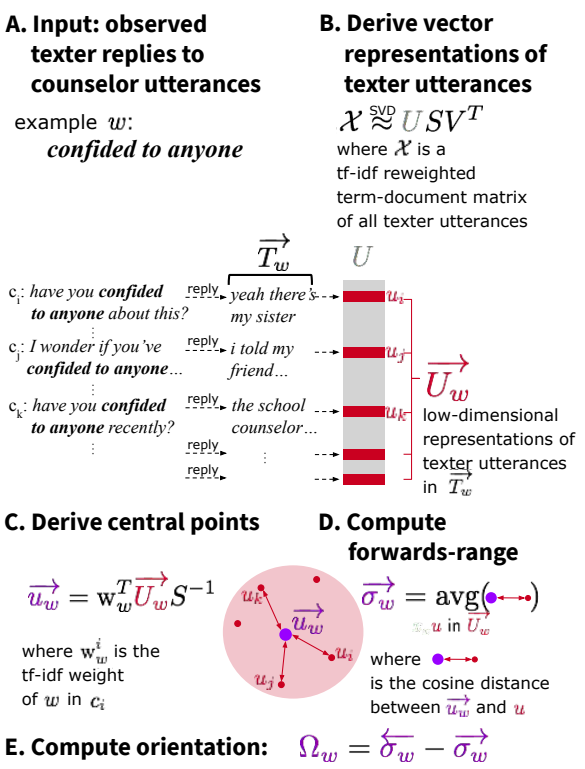

Figure 3: Outline of steps to compute orientation Ω_w of phrasing w, as described in Section 4.2. Panels A-D show the procedure for computing forwards-range $\overrightarrow{\sigma_w}$; the procedure for backwards-range $\overleftarrow{\sigma_w}$ is similar.

Deriving vector representations (Figure 3B). To obtain vectors for each texter utterance, we construct $\mathcal{X}$, a tf-idf reweighted term-document matrix where rows represent texter utterances and columns represent phrasings used by texters. To ensure that we go beyond lexical matches and capture conceptual classes (e.g., possible confidants, frustrating situations), we use singular value decomposition to get $\mathcal{X} \approx USV^T$. Each row of U is a vector representation u_i of utterance t_i in the induced low-dimensional space $\mathbb{T}$. $\overrightarrow{U_w}$ then consists of the corresponding subset of rows of U (highlighted in Figure 3B).

Deriving central points (Figure 3C). For each w, we take the central point $\overrightarrow{u_w}$ to be a weighted average of vectors in $\overrightarrow{U_w}$. Intuitively, a texter utterance t_i with vector u_i should have a larger contribution to $\overrightarrow{u_w}$ if w is more prominent in the counselor utterance c_i that preceded it. We let w_w^i denote the normalized tf-idf weight of w in c_i, and use w_w^i as the weight of the corresponding vector u_i. To properly map the resultant weighted sum $\sum \mathrm{w}_w^i u_i$ into $\mathbb{T}$, we divide each dimension by the corresponding singular value in S. As such, if w_w is a vector of weights w_w^i, we can calculate the central point $\overrightarrow{u_w}$

of $\overrightarrow{U_w}$ as $\overrightarrow{u_w} = \mathrm{w}_w^T \overrightarrow{U_w} S^{-1}$. In the other direction, we likewise compute $\overleftarrow{u_w} = \mathrm{w}_w^T \overleftarrow{U_w} S^{-1}$.

Forwards- and backwards-ranges (Figure 3D). We take the forwards-range $\overrightarrow{\sigma_w}$ of w to be the average cosine distance from each vector in $\overrightarrow{U_w}$ to the center point $\overrightarrow{u_w}$. Likewise, we take $\overleftarrow{\sigma_w}$ as the average distance from each vector in $\overleftarrow{U_w}$ to $\overleftarrow{u_w}$.

Phrasing-level orientation (Figure 3E). Importantly, since we've computed the forwards- and backwards-ranges $\overrightarrow{\sigma_w}$ and $\overleftarrow{\sigma_w}$ using distances in the same space $\mathbb{T}$, their values are comparable. We then compute the orientation of w as their difference: $\Omega_w = \overleftarrow{\sigma_w} - \overrightarrow{\sigma_w}$.

Utterance-level orientation. To compute the orientation of an utterance c_i, we first compute the orientation of each sentence in c_i as the tf-idf weighted average Ω_w of its constitutent phrasings. Note that a multi-sentence utterance can orient in *both* directions—e.g., a counselor could concatenate c_2 and c_1 from Figure 1 in a single utterance, addressing the texter's previous utterance before moving ahead. To model this heterogeneity, we consider both the minimum and maximum sentence-orientations in an utterance: $\Omega^{\min}$ captures the extent to which the utterance looks backwards, while $\Omega^{\max}$ captures the extent to which it aims to advance forwards.

5 Application to Counseling Data

We apply our method to characterize messages from crisis counselors on the platform. We compute the orientations of the phrasings they use, represented as dependency-parse arcs. We use a training set of 351,935 texter and counselor messages each, from a random sample of conversations omitted in subsequent analyses.[5] Table 1 shows representative phrasings and sentences of different orientations.[6] Around two-thirds of phrasings and sentences have $\Omega < 0$, echoing the importance of addressing the texter's previous remarks.

In what follows, we analyze counselor behaviors in terms of orientation, and illustrate how the measure can be useful for examining conversations. We start by validating our method via two complementary approaches. In a subset of sentences manually annotated with the counseling strategies they exhibit, we show that orientation meaningfully reflects these strategies (Section 5.1). At a larger scale, we show that the orientation of utterances over the course of a conversation aligns with domain knowledge about counseling conversation structure (Section 5.2). We also find that other measures for characterizing utterances are not as rich as orientation in capturing counseling strategies and conversation structure (Section 5.3). Finally, we show that a counselor's orientation in a conversation is tied to indicators of their effectiveness in helping the texter (Section 5.4).

5.1 Validation: Counseling Strategies

Even though it is computed without the guidance of any annotations, we expect orientation to meaningfully reflect strategies for advancing or addressing that crisis counselors are taught. The first author hand-labeled 400 randomly-selected sentences with a set of pre-defined strategies derived from techniques highlighted in the training material. We note example sentences in Table 1 which exemplify each strategy, and provide more extensive descriptions in the appendix.

Figure 4A depicts the distributions of orientations across each label, sorted from most backwards- to most forwards-oriented. We find that the relative orientation of different strategies corroborates their intent as described in the literature. Statements **reflecting** or **affirming** what the texter has said to check understanding or convey empathy (characterized by phrasings like *totally normal*) tend to be backwards-oriented; statements prompting the texter to advance towards **problem-solving** (e.g., *[what] has helped*) are more forwards-oriented. **Exploratory** queries for more information on what the texter has already said (e.g., *happened to make*) tend to have middling orientation (around 0). The standard deviation of orientations over messages within most of the labels is significantly lower than across labels (bootstrapped $p < .05$, solid circles), showing that orientation yields interpretable groupings of messages in terms of important counseling strategies.

The measure also offers complementary information. For instance, we find sentences that aren't accounted for by pre-defined labels, but still map to interpretable orientations, such as backwards-oriented examples assuaging texter concerns about the platform being a safe space to self-disclose.

[5] Further implementation details are listed in the appendix.

[6] Example sentences are derived from real sentences in the data, and modified for readability. The examples were chosen to reflect common situations in the data, and were vetted by the platform to ensure the privacy of counselors and texters.

Orientation	Example phrasings	Example sentences
Backwards-oriented (bottom 25%)	sounds frustrating, totally normal, great ways, on [your] plate, be overwhelming, sometimes feel frightening, on top [of] been struggling, feeling alone	You have a lot of things on your plate, between family and financial problems. **[reflection]** It's totally normal to feel lonely when you have no one to talk to. **[reflection]** Those are great ways to improve the relationship. **[affirmation]**
(middle 25%)	happened [to] make, mean [when you] say, is that, you recognized, source of the moment, are brave	Has anything happened to make you anxious? **[exploration]** It's good you recognized the need to reach out. **[affirmation]** Can you tell me what you mean when you say you're giving up? **[risk assessment]**
Forwards-oriented (top 25%)	plan for, confided [to] anyone, usually do, has helped, been talking, best support have considered, any activities	Can you think of anything that has helped when you've been stressed before? **[problem solving]** I want to be the best support for you today. **[problem solving]** We've been talking for a while now, how do you feel? **[closing]**

Table 1: Example phrasings and sentences with labeled strategies from crisis counselors' messages, at varying orientations: backwards-oriented (from the bottom 25% of Ω), middle, and forwards-oriented (from top 25%).

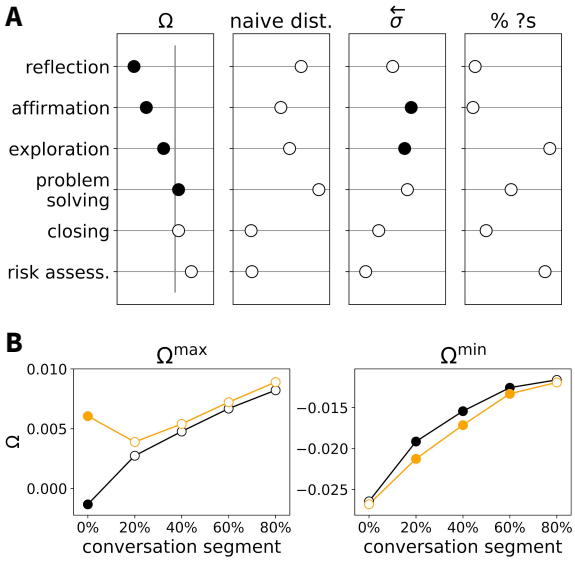

Figure 4: Validating the orientation measure and comparing to alternatives. **A** Leftmost: Mean Ω per counseling strategy label (vertical line denotes $\Omega = 0$). Next three: same for other measures. **B**: Mean Ω^{max} and Ω^{min} per segment for risk-assessed (orange) and non-risk-assessed (black) conversations. Both: Solid circles indicate statistically significant differences (Wilcoxon $p < 0.01$, comparing within-counselor).

5.2 Validation: Conversation Structure

We also show that orientation tracks with the structure of crisis counseling conversations as described in the training material. Following Althoff et al. (2016), we divide each conversation with at least ten counselor messages into five equal-sized segments and average Ω^{max} and Ω^{min} over messages in each segment.

Figure 4B (black lines) shows that over the course of a conversation, messages tend to get more forwards-oriented (higher Ω^{max} and Ω^{min}). This matches a standard conversation structure taught in the training: addressing the texter's existing problems before advancing towards problem-solving. While this correspondence holds in aggregate, orientation also captures complementary information to advancement through stages—e.g., while problem-solving, counselors may still address and affirm a texter's ideas (Table 1, row 3).

We also consider a subset of conversations where we expect a different trajectory: for potentially suicidal texters, the training directs counselors to immediately start a process of **risk assessment** in which actively prompting the texter to disclose their level of suicidal ideation takes precedence over other objectives. As such, we expect more forwards-oriented messages at the starts of conversations involving such texters. Indeed, in the 30% of conversations which are risk-assessed, we find significantly larger Ω^{max} in the first segment (Figure 4B, orange line; Wilcoxon $p < 0.01$ in the first stage, comparing within-counselor). Ω^{min} is *smaller* at each stage, suggesting that counselors balance actively prompting these critical disclosures with addressing them.

5.3 Alternative Operationalizations

We compare orientation to other methods for capturing a counselor's balancing decisions:

Naive distance. We conside the naive approach in Section 4, taking a difference in cosine distances between tf-idf representations of a message and its reply, and a message and its predecessor.

Backwards-range. We consider just the message's backwards-range. For each sentence we take tf-idf weighted averages of component $\overleftarrow{\sigma}_w$ and take minimum $\overleftarrow{\sigma}$ for each message.[7]

[7] We get qualitatively similar results with maximum $\overrightarrow{\sigma}$.

Question-asking. We consider whether the message has a question. This was used in Walker and Whittaker (1990) as a signal of taking control, which could be construed as forwards-oriented.

Within-label standard deviations of each alternative measure are generally not significantly smaller than across-label (Figure 4A), indicating that these measures are poorer reflections of the counseling strategies. Label rankings under the measures are also arguably less intuitive. For instance, reflection statements have relatively large (naive) cosine distance from their predecessors. Indeed, the training encourages counselors to *process* rather than simply restate the texter's words.

These measures also track with the conversation's progress differently—notably, none of them distinguish the initial dynamics of risk-assessed conversations as reflected in $\Omega^{\max}$ (see appendix).

5.4 Relation to Conversational Effectiveness

Past work on counseling has extensively discussed the virtues of addressing a client's situation (Rogers, 1957; Hill and Nakayama, 2000). Some studies also suggest that accounting for *both* addressing and advancing is important, such that effective counselors manage to mix backwards- and forwards-oriented actions (Mishara et al., 2007).

We use orientation to examine how these strategies are tied to conversational effectiveness in crisis counseling at a larger scale, using our measures to provide a unified view of advancing and addressing. To derive simple conversation-level measures, we average $\Omega^{\max}$ and $\Omega^{\min}$ over each counselor message in a conversation.

Adjudicating counseling conversation quality is known to be difficult (Tracey et al., 2014). As a starting point, we relate our conversation-level measures to two complementary indicators of a conversation's effectiveness:[8]

Perceived helpfulness. We consider responses from a post-conversation survey asking the texter whether the conversation was helpful, following Althoff et al. (2016). Out of the 26% of conversations with a response, 89% were rated as helpful.[9]

Conversation length. We consider a conversation's length as a simple indicator of the pace of its progress: short conversations may rush the texter, while prolonged conversations could suggest stalling and could even demoralize the counselor (Landphair and Preddy, 2012).[10]

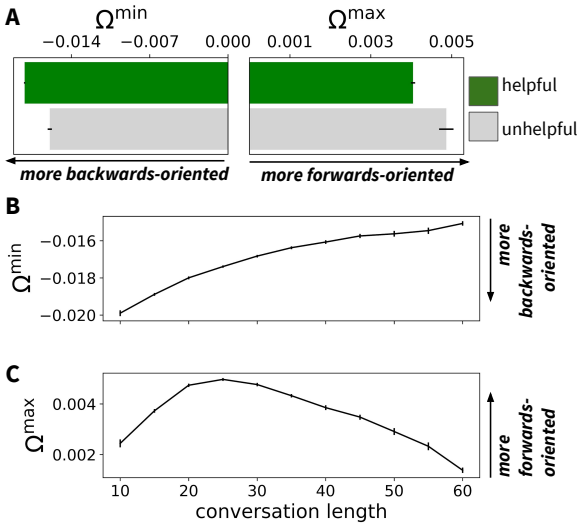

Figure 5: Relation between orientation and conversational effectiveness. **A**: Mean $\Omega^{\min}$ and $\Omega^{\max}$ in conversations rated as helpful (green) or unhelpful (grey) (macroaveraged per conversation). Differences in both measures are significant (Mann Whitney U test $p < 0.001$). **B**, **C**: Mean $\Omega^{\min}$ and $\Omega^{\max}$ of conversations with varying lengths (in # of messages). Both plots: Error bars show 95% bootstrapped confidence intervals.

Figure 5A compares $\Omega^{\min}$ and $\Omega^{\max}$ in conversations rated as helpful and unhelpful by texters. Both measures are significantly *smaller* in conversations perceived as helpful, suggesting that texters have a better impression of relatively *backwards-oriented* interactions where the counselor is inclined towards addressing their situation. As such, this result echoes past findings relating addressing to effectiveness.

Figure 5B compares $\Omega^{\min}$ in conversations of varying lengths, showing that $\Omega^{\min}$ *increases* with length, such that counselors exhibit less propensity for addressing in longer conversations. Anecdotal observations cited in interviews with the platform's staff suggest one interpretation: conversations in which a texter feels their concerns were not satisfactorily addressed may be prolonged when they circle back to revisit these concerns.

Figure 5C relates $\Omega^{\max}$ to conversation length. We find that $\Omega^{\max}$ is smaller in the lengthiest conversations, suggesting that such prolonged in-

[8] We perform all subsequent analyses on a subset of 234,433 conversations, detailed in the appendix.

[9] We note that this indicator is limited by important factors such as the selection bias in respondents.

[10] As the training material notes, conversation length and texter perception may signal complementary or even conflicting information about a texter's experience of a conversation and its effectiveness: "Some texters resist the end of the conversation. They ruminate [...] causing the conversation to drag on without any progress."

teractions may be stalled by a weaker impulse to advance forwards. Extremely short conversations have smaller $\Omega^{\max}$ as well, such that premature endings may also reflect issues in advancing. As such, we add credence to the previously-posited benefits of mixing addressing and advancing: forwards-oriented actions may be tied to making progress, while a weaker propensity to advance may signal a suboptimal pace.

Counselor-level analysis. These findings could reflect various confounds—for instance, a counselor's choice of orientation may have no bearing on the rating they receive from a particularly difficult texter. To address this, we compute similar correspondences between orientation and our effectiveness indicators at the level of counselors rather than conversations; this analysis is detailed in the appendix. Our conversation-level results are replicated under these controls.

6 Discussion and Future Work

In this work, we sought to examine a key balance in crisis counseling conversations between advancing forwards and addressing what has already been said. Realizing this balance is one of the many challenges that crisis counselors must manage, and modeling the actions they take in light of such challenges could point to policies to better support them. For instance, our method could assist human supervisors in monitoring the progress of ongoing conversations to detect instances of rushing or stalling, or enable larger-scale analyses of conversational behaviors to inform how counselors are trained. The unsupervised approach we propose could circumvent difficulties in getting large-scale annotations of such sensitive content.

Future work could bolster the measure's usefulness in several ways. Technical improvements like richer utterance representations could improve the measure's fidelity; more sophisticated analyses could better capture the dynamic ways in which the balance of objectives is negotiated across many turns. The preliminary explorations in Section 5.4 could also be extended to gauge the causal effects of counselors' behaviors (Kazdin, 2007).

We expect balancing problems to recur in conversational settings beyond crisis counseling, such as court proceedings, interviews, debates and other mental health contexts like long-term therapy. In these settings, individuals also make potentially consequential choices that span the backwards-forwards orientation axis, such as addressing previous arguments (Tan et al., 2016; Zhang et al., 2016) or asking leading questions (Leech, 2002). Our measure is designed to be broadly applicable, requiring no domain-specific annotations; we provide exploratory output on justice utterances from the Supreme Court's oral arguments in the appendix and release code implementing our approach at `http://convokit.cornell.edu` to encourage experiments in other domains. However, the method's efficacy in the present setting is likely boosted by the relative uniformity of crisis counseling conversations; and future work could aim to better accomodate settings with less structure and more linguistic variability. With such improvements, it would be interesting to study other domains where interlocutors are faced with conversational challenges.

Acknowledgements. We thank Jonathan P. Chang, Caleb Chiam, Liye Fu, Dan Jurafsky, Jack Hessel, and Lillian Lee for helpful conversations, and the anonymous reviewers for their thoughtful comments. We also thank Ana Smith for collecting and processing the Supreme Court oral argument transcripts we used in the supplementary material. This research, and the counseling service examined herein, would not have been possible without Crisis Text Line. We are particularly grateful to Robert Filbin, Christine Morrison, and Jaclyn Weiser for their valuable insights into the experiences of counselors and for their help with using the data. The research has been supported in part by NSF CAREER Award IIS1750615 and a Microsoft Research PhD Fellowship. The collaboration with Crisis Text Line was supported by the Robert Wood Johnson Foundation; the views expressed here do not necessarily reflect the views of the foundation.

References

Tim Althoff, Kevin Clark, and Jure Leskovec. 2016. Large-scale Analysis of Counseling Conversations: An Application of Natural Language Processing to Mental Health. *Transactions of the Association for Computational Linguistics*.

David C. Atkins, Mark Steyvers, Zac E. Imel, and Padhraic Smyth. 2014. Scaling up the evaluation of psychotherapy: Evaluating motivational interviewing fidelity via statistical text classification. *Implementation Science*.

Graham D. Bodie, Andrea J. Vickery, Kaitlin Cannava, and Susanne M. Jones. 2015. The Role of "Active Listening" in Informal Helping Conversations: Impact on Perceptions of Listener Helpfulness, Sensitivity, and Supportiveness and Discloser Emotional Improvement. *Western Journal of Communication*.

David Bracewell, Marc Tomlinson, and Hui Wang. 2012. Identification of Social Acts in Dialogue. In *Proceedings of COLING*.

Dogan Can, David C. Atkins, and Shrikanth S. Narayanan. 2015. A dialog act tagging approach to behavioral coding: A case study of addiction counseling conversations. In *Proceedings of INTERSPEECH*.

Jie Cao, Michael Tanana, Zac Imel, Eric Poitras, David Atkins, and Vivek Srikumar. 2019. Observing Dialogue in Therapy: Categorizing and Forecasting Behavioral Codes. In *Proceedings of ACL*.

Jonathan P Chang, Caleb Chiam, Liye Fu, Andrew Wang, Justine Zhang, and Cristian Danescu-Niculescu-Mizil. 2020. ConvoKit: A Toolkit for the Analysis of Conversations. In *Proceedings of SIGDIAL*.

Mark G Core and James F Allen. 1997. Coding Dialogs with the DAMSL Annotation Scheme. *AAAI fall symposium on communicative action in humans and machines*.

Jacques Gaume, Nicolas Bertholet, Mohamed Faouzi, Gerhard Gmel, and Jean-Bernard Daeppen. 2010. Counselor motivational interviewing skills and young adult change talk articulation during brief motivational interventions. *Journal of Substance Abuse Treatment*.

Barbara J. Grosz, Aravind K. Joshi, and Scott Weinstein. 1995. Centering: A Framework for Modeling the Local Coherence of Discourse. *Computational Linguistics*.

Clara E. Hill and Emilie Y. Nakayama. 2000. Client-centered therapy: Where has it been and where is it going? A comment on Hathaway (1948). *Journal of Clinical Psychology*.

Jon Houck. 2008. Motivational Interviewing Skill Code (MISC) 2.1.

Neil P. Jones, Alison A. Papadakis, Caitlin M. Hogan, and Timothy J. Strauman. 2009. Over and over again: Rumination, reflection, and promotion goal failure and their interactive effects on depressive symptoms. *Behaviour Research and Therapy*.

Alan E. Kazdin. 2007. Mediators and mechanisms of change in psychotherapy research. *Annual Review of Clinical Psychology*.

Juliette Landphair and Teri Preddy. 2012. More than talk: Co-Rumination among college students. *About Campus*.

Fei-Tzin Lee, Derrick Hull, Jacob Levine, Bonnie Ray, and Kathy McKeown. 2019. Identifying therapist conversational actions across diverse psychotherapeutic approaches. In *Proceedings of the Sixth Workshop on Computational Linguistics and Clinical Psychology*.

Beth L. Leech. 2002. Asking Questions: Techniques for Semistructured Interviews. *Political Science & Politics*.

William C. Mann and Sandra A. Thompson. 1988. Rhetorical Structure Theory: Toward a functional theory of text organization. *Text - Interdisciplinary Journal for the Study of Discourse*.

Brian L. Mishara, François Chagnon, Marc S. Daigle, Bogdan Balan, Sylvaine Raymond, Isabelle Marcoux, Cécile Bardon, Julie K. Campbell, and Alan D. Berman. 2007. Which helper behaviors and intervention styles are related to better short-term outcomes in telephone crisis intervention? Results from a Silent Monitoring Study of Calls to the U.S. 1-800-SUICIDE Network. *Suicide & life-threatening behavior*.

Johanna D. Moore and Cécile Paris. 1993. Planning Text for Advisory Dialogues: Capturing Intentional and Rhetorical Information. *Computational Linguistics*.

Viet-An Nguyen, Jordan Boyd-Graber, Philip Resnik, Deborah A. Cai, Jennifer E. Midberry, and Yuanxin Wang. 2014. Modeling topic control to detect influence in conversations using nonparametric topic models. *Machine Learning*.

Susan Nolen-Hoeksema, Blair E. Wisco, and Sonja Lyubomirsky. 2008. Rethinking Rumination. *Perspectives on Psychological Science*.

Sungjoon Park, Donghyun Kim, and Alice Oh. 2019. Conversation Model Fine-Tuning for Classifying Client Utterances in Counseling Dialogues. In *Proceedings of NAACL*.

Verónica Pérez-Rosas, Rada Mihalcea, Kenneth Resnicow, Satinder Singh, and Lawrence An. 2017. Understanding and Predicting Empathic Behavior in Counseling Therapy. In *Proceedings of ACL*.

Verónica Pérez-Rosas, Xuetong Sun, Christy Li, Yuchen Wang, Kenneth Resnicow, and Rada Mihalcea. 2018. Analyzing the Quality of Counseling Conversations: The Tell-Tale Signs of High-quality Counseling. In *Proceedings of LREC*.

Anthony R. Pisani, Nitya Kanuri, Bob Filbin, Carlos Gallo, Madelyn Gould, Lisa S. Lehmann, Robert Levine, John E. Marcotte, Brian Pascal, David Rousseau, Shairi Turner, Shirley Yen, and Megan L. Ranney. 2019. Protecting User Privacy and Rights in Academic Data-Sharing Partnerships: Principles From a Pilot Program at Crisis Text Line. *Journal of Medical Internet Research*.

Vinodkumar Prabhakaran, Camilla Griffiths, Hang Su, Prateek Verma, Nelson Morgan, Jennifer L. Eberhardt, and Dan Jurafsky. 2018. Detecting Institutional Dialog Acts in Police Traffic Stops. *Transactions of the Association for Computational Linguistics*.

Alan Ritter, Colin Cherry, and Bill Dolan. 2010. Unsupervised Modeling of Twitter Conversations. In *Proceedings of NAACL*.

Carl R. Rogers. 1957. The necessary and sufficient conditions of therapeutic personality change. *Journal of Consulting Psychology*.

Stephen Rollnick and William R. Miller. 1995. What is Motivational Interviewing? *Behavioural and Cognitive Psychotherapy*.

Amanda J. Rose, Wendy Carlson, and Erika M. Waller. 2007. Prospective Associations of Co-Rumination with Friendship and Emotional Adjustment: Considering the Socioemotional Trade-Offs of Co-Rumination. *Developmental Psychology*.

Sara Rosenthal and Kathleen McKeown. 2015. I Couldn't Agree More: The Role of Conversational Structure in Agreement and Disagreement Detection in Online Discussions. In *Proceedings of SIGDIAL*.

Harvey Sacks. 1992. *Lectures on Conversation*. Blackwell.

Jonathan Sandoval, Amy Nicole Scott, and Irene Padilla. 2009. Crisis counseling: An overview. *Psychology in the Schools*.

Emanuel A. Schegloff. 1987. Some sources of misunderstanding in talk-in-interaction. *Linguistics*.

Chenhao Tan, Vlad Niculae, Cristian Danescu-Niculescu, and Lillian Lee. 2016. Winning Arguments: Interaction Dynamics and Persuasion Strategies in Good-faith Online Discussions. In *Proceedings of WWW*.

Michael Tanana, Kevin A. Hallgren, Zac E. Imel, David C. Atkins, and Vivek Srikumar. 2016. A Comparison of Natural Language Processing Methods for Automated Coding of Motivational Interviewing. *Journal of Substance Abuse Treatment*.

Jenny Thomas. 1983. Cross-cultural pragmatic failure. *Applied Linguistics*.

Terence J. G. Tracey, Bruce E. Wampold, James W. Lichtenberg, and Rodney K. Goodyear. 2014. Expertise in psychotherapy: An elusive goal? *The American Psychologist*.

Marilyn Walker and Steve Whittaker. 1990. Mixed Initiative in Dialogue: An Investigation into Discourse Segmentation. In *Proceedings of ACL*.

Xuewei Wang, Weiyan Shi, Richard Kim, Yoojung Oh, Sijia Yang, Jingwen Zhang, and Zhou Yu. 2019. Persuasion for Good: Towards a Personalized Persuasive Dialogue System for Social Good. In *Proceedings of ACL*.

Bonnie Lynn Webber. 2001. Computational Perspectives on Discourse and Dialog. In *The Handbook of Discourse Analysis*. John Wiley & Sons, Ltd.

Harry Weger, Gina R. Castle, and Melissa C. Emmett. 2010. Active Listening in Peer Interviews: The Influence of Message Paraphrasing on Perceptions of Listening Skill. *International Journal of Listening*.

Justine Zhang, Robert Filbin, Christine Morrison, Jaclyn Weiser, and Cristian Danescu-Niculescu-Mizil. 2019. Finding Your Voice: The Linguistic Development of Mental Health Counselors. In *Proceedings of ACL*.

Justine Zhang, Ravi Kumar, Sujith Ravi, and Cristian Danescu-Niculescu-Mizil. 2016. Conversational Flow in Oxford-style Debates. In *Proceedings of NAACL*.

Justine Zhang, Arthur Spirling, and Cristian Danescu-Niculescu-Mizil. 2017. Asking too Much? The Rhetorical Role of Questions in Political Discourse. In *Proceedings of EMNLP*.

A Appendices

A.1 Further Details About Methodology

Here, we provide further details on our methodology for measuring orientation, to supplement the description in Section 4.2 and aid reproducibility.

Our aim in the first part of our methodology is to measure the orientation of phrasings Ω_w. We would like to ensure that our measure is not skewed by the relative frequencies of phrasings, and take two steps to this ends, which empirically produced more interpretable output. First, we scale rows of term-document matrix $\mathcal{X}$ (corresponding to texter phrasings) to unit ℓ_2 norm before deriving their representation in $\mathbb{T}$ via singular value decomposition. Second, we remove the first SVD dimension, i.e., first column of U, and renormalize each row, before proceeding.

A.2 Further Details About Application to Counseling Data

Here, we provide further details on how we applied our methodology to the dataset of counseling conversations in order to measure the orientation of counselor utterances, as described in Section 5. In particular, we list empirical choices

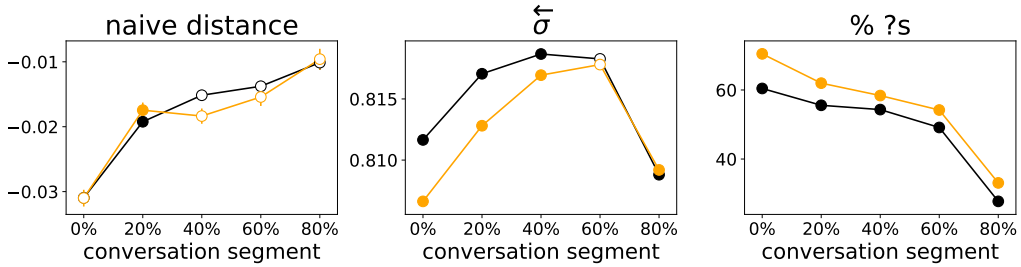

Figure 6: Mean naive distance, backwards-range ($\overleftarrow{\sigma}$), and % of utterances with questions, per segment for risk-assessed (orange) and non-risk-assessed (black) conversations; solid circles indicate statistically significant differences (Wilcoxon $p < 0.01$, comparing conversation types within counselor).

made in extracting and then processing the training set of 351,935 texter and counselor messages used to measure phrasing orientations.

We randomly sampled 20% of *counselors* in the data; all conversations by these counselors were omitted in subsequent analyses. We merged consecutive messages from the same interlocutor. To mitigate potential noise in characterizing phrasings, we considered only counselor and texter message pairs in which each message has between 15 and 45 words. We extracted all messages from the conversations which met these constraints.

We represent counselor phrasings as dependency-parse arcs and texter messages as unigrams, reflecting the comparatively structured language of the counselors versus the texters (counselors are instructed to speak in grammatically well-formed sentences). We consider the 5000 most frequent phrasings for each role, and discard sentences without any such phrasings. Finally, we used 25 SVD dimensions to induce $\mathbb{T}$.

A.3 Full Listing of Counselor Action Labels

Table 2 lists each counseling action derived from the training material and used during the validation procedure (Section 5.1) to label sentences.

A.4 Orientation and Lexical Properties

Here, we supplement our discussion of simple lexical properties that could be used to characterize messages (Section 5.3), discussing how orientation reflects these properties and showing that orientation is not subsumed by them.

Backwards-range. As seen in their weak backwards-range (high, i.e., spread-out $\overleftarrow{\sigma}$), affirmations that highlight the texter's strengths can follow a variety of situations. However, the replies they prompt are yet more diffuse, emphasizing the need to compare both directions.

reflection (113) re-wording to show understanding and validate feelings *It can be overwhelming to go through that on your own.*
affirmation (60) pointing out the texter's positive qualities and actions *You showed a lot of strength in reaching out to us.*
exploration (44) prompting texters to expand on their situation *Is this the first real fight you've had with your boyfriend?*
problem solving (110) identifying the texter's goals and potential coping skills *What do you usually do to help you feel calmer?*
closing (43) reviewing the conversation and transitioning to a close *I think you have a good plan to get some rest tonight.*
risk assessment (19) assessing suicidal ideation or risk of self-harm *Do you have access to the pills right now?*

Table 2: Counseling strategies and representative examples derived from the training material. The number of sentences (out of 400) assigned to each label is shown in parentheses (11 were not labeled as any action).

Question-asking. We see that questions—which nominally prompt the texter for a response—are more forwards-oriented than non-questions; 61% of sentences with '?' have $\Omega > 0$ compared to 21% of sentences without. However, these numbers also show that explicitly-marked questions are inexact proxies of forwards-oriented sentences—as in Table 1, questions can address a past remark by prompting clarifications, while counselors can use non-questions to suggest an intent to advance stages (e.g., to transition to problem-solving).

A.5 Relating Alternate Measures to Conversation Progress

Figure 6 shows averages per conversation segment (i.e., 20% of a conversation) for each alternative measure considered in Section 5.3. Comparing to the average $\Omega^{\max}$ and $\Omega^{\min}$ shown in Figure 4, we see that these measures track with the conversa-

tion's progress differently, and none of them distinguish the initial dynamics of risk-assessed conversations as dramatically as reflected in $\Omega^{\max}$, e.g., simple counts of questions do not distinguish between questions geared towards risk-assessment versus more open-ended problem exploration.

A.6 Further Details About Data Used in Analyses

Here, we provide further details about the subset of data we used to analyze counselors' orientation behavior (Section 5.4). In particular, our aim was to characterize behavior in typical conversations rather than exceptional cases or those that reflected earlier versions of the training curriculum. As such, we only considered the 234,433 conversations which had least five counselor messages, were not risk-assessed or disconnected before completion, and were taken by counselors who joined the platform after January 2017.

A.7 Counselor-Level Analysis

Here, we provide further details about our procedure for analyzing counselor-level correspondences between orientation and effectiveness indicators, as alluded to in Section 5.4.

Recall that our conversation-level findings may be confounded by texter idiosyncracies: for instance, texters with particularly difficult situations might affect a counselor's behaviour, but may also be more likely to give bad ratings, independent of how the counselor behaves. Alternatively, an overly long conversation could arise because the counselor is less forwards-oriented, or because the texter is reluctant to make progress from the outset, making it hard for the counselor to attempt to prompt them forwards.

To separate a counselor's decisions from these situational effects, we take a counselor-level perspective. While counselors cannot selectively talk with different types of texters, they can exhibit cross-conversational inclinations for particular behaviors. We therefore relate these cross-conversational *tendencies* in orienting a conversation to a counselor's long-term propensity for receiving helpful ratings, or having long or short conversations. We proceed to describe our methodology for relating counselor tendencies to perceived helpfulness; an analogous procedure could be applied to conversation length as well.

We characterize a counselor's orienting behavior as the average $\Omega^{\max}$ and $\Omega^{\min}$ over the conversations they take; we likewise take the proportion of their (rated) conversations which were perceived as helpful. We restrict our counselor level analyses to the 20th to 120th conversations of the 1495 counselors who have taken at least 120 conversations (ignoring their initial conversations when they are still acclimatizing to the platform).

To cleanly disentangle counselor tendency and conversational circumstance, we *split* each counselor's conversations into two interleaved subsets (i.e., first, third, fifth ... versus second, fourth ... conversations), measuring orientation on one subset and computing a counselor's propensity for helpful ratings on the other. Here, we draw an analogy to the machine learning paradigm of taking a train-test split: "training" counselor tendencies on one subset and "testing" their relation to rating on the other subset. In general, the directions of the effects we observe hold with stronger effects if we do not take this split.

Echoing conversation-level effects, counselors that tend to be less forwards-oriented and more backwards-oriented (those in the bottom thirds of $\Omega^{\max}$ and $\Omega^{\min}$ respectively) are more likely to be perceived as helpful; this contrast is stronger in terms of $\Omega^{\min}$ (Cohen's $d = 0.30$, $p < 0.001$) than $\Omega^{\max}$ ($d = 0.13$, $p < 0.05$), suggesting that a counselor's tendency for advancing weighs less on their perceived helpfulness than their tendency for addressing. Also in line with the conversation-level findings, counselors with smaller $\Omega^{\max}$ tend to have longer conversations ($d = 0.54, p < 0.001$), as do counselors with larger $\Omega^{\min}$ ($d = 0.17$)—here, a counselor's tendency for advancing is more related to their propensity for shorter conversations than their tendency for addressing.

We note that counselors on the platform cannot selectively take conversations with certain texters; rather, the platform automatically assigns incoming texters to a counselor. As such, the counselor-level effects we observe cannot be explained by counselor self-selection for particular situations.

A.8 Orientation in Multi-Sentence Utterances

Our motivation in characterizing utterances using the minimum and maximum sentence orientation was to reflect potential heterogeneities in utterances which could be both backwards- and forwards-oriented (consider a message where c_2 and c_1 from Figure 1 are concatenated). Examin-

Orientation	Example phrasings	Example sentences
Less forwards-oriented (bottom 25%)	i understand, have been, part of, so you, sentence, talking about might, particular but the, give to	As I understand the facts [...] he had tried to kill the husband, shooting him twice in the head? (Scalia) You started out by talking about what the first jury knew, but [...] we aren't reviewing that determination. (Roberts) I guess the problem is the list of absurdities that they point to, not the least of which is a dry dock. (Sotomayor) So you hedged, because it's very hard to find the right sentence. (Breyer)
More forwards-oriented (top 25%)	hypothetical, would have, agree, difference [between], [do] you think, your position your argument, a question apply, was there	Suppose under this hypothetical [...] the judge doesn't say aggravated murder when he submits it to the jury. (Kennedy) I just want to know your position on the second, the cart before the horse point. (Souter) Do you also agree [...] that if not properly administered there is some risk of excruciating pain? (Stevens) What's the difference between pigment and color [...] ? (Ginsburg)

Table 3: Example phrasings and sentences from utterances of Supreme Court justices, identified in parentheses, which are less or more forwards-oriented (bottom and top 25% of Ω).

ing the 64% of counselor messages with multiple sentences, we see that 52% of these messages have $\Omega^{\min} < 0$ and $\Omega^{\max} > 0$. Our method, which is able to account for this heterogeneity, thus points to one potential strategy for counselors to bridge between both objectives.

A.9 Application to Supreme Court Oral Arguments

Here, we include an exploratory study of how our approach could be adapted to analyze domains beyond crisis counseling conversations, as alluded to in Section 6. We apply the method to measure the orientation of utterances by Supreme Court justices during oral arguments, when they engage in exchanges with lawyers (so justices and lawyers play the roles of counselor and texter, respectively, in our method). We used transcripts of 668 cases, taken from the Oyez project (https://www.oyez.org/), averaging 120 justice utterances per case.[11]

Oral arguments contain more linguistic and topical heterogeneiety than counseling conversations, since they cover a wide variety of cases, and because the language used by each justice is more differentiated. In addition, the dataset is much smaller. As such, this represents a more challenging setting than the counseling context, requiring changes to the precise procedure used to measure orientation, and pointing to the need for further technical improvements, discussed in Section 6.

Nonetheless, our present methodology is able to produce interpretable output. Table 3 shows representative phrasings and (paraphrased) sentences with different orientations. In contrast to the counseling domain, 70% of phrasings and 93% of sentences have $\Omega > 0$, perhaps reflecting the particular power dynamic in the Supreme Court, where justices are tasked with scrutinizing the arguments made by lawyers. We find that highly forwards-oriented phrasings and utterances tend to reflect justices pressing on the lawyers to address a point (e.g., do you *agree*, what's the *difference between*); the least forwards-oriented phrases involve the justice rehashing and reframing (not always in complimentary terms) a lawyer's prior utterances (e.g., *so you* [...], [as] *i understand*).

We used a training set of 15,862 justice and lawyer messages, where each utterance had between 10 and 100 words. Both lawyer and justice utterances were represented as dependency-parse arcs. Empirically, we found that the methodology was sensitive to idiosyncracies of particular cases and justices. To minimize this effect, we restricted the size of the justice's vocabulary by only considering the 398 justice phrasings which occurred in at least 200 utterances.

[11] The data used can be found at http://analysmith.com/research/scotus/data.