

# A Computational Approach to Understanding Empathy Expressed in Text-based Mental Health Support



Ashish Sharma



Adam Miner



Dave Atkins



Tim Althoff

# Content Warning

This presentation contains examples related to suicidal ideation and self-harm

# Mental Health: Need vs. Access

**Access to mental health care is poor across the globe**

- Low-income, middle-income countries
  - **1** psychiatrist per **100k** individuals
- United States
  - **60%** of the counties do not have a single psychiatrist

**Key:** We may **never** have enough mental health professionals to meet the **need**

**Online peer support platforms can help!**

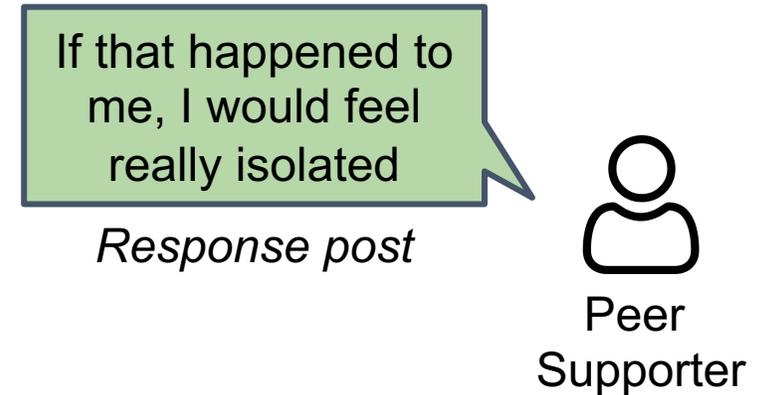
- E.g. TalkLife, 7cups, Mental health subreddits
- Millions of users **seek** and **provide support** through **conversations**



# Text-based, asynchronous conversations on peer-support platforms



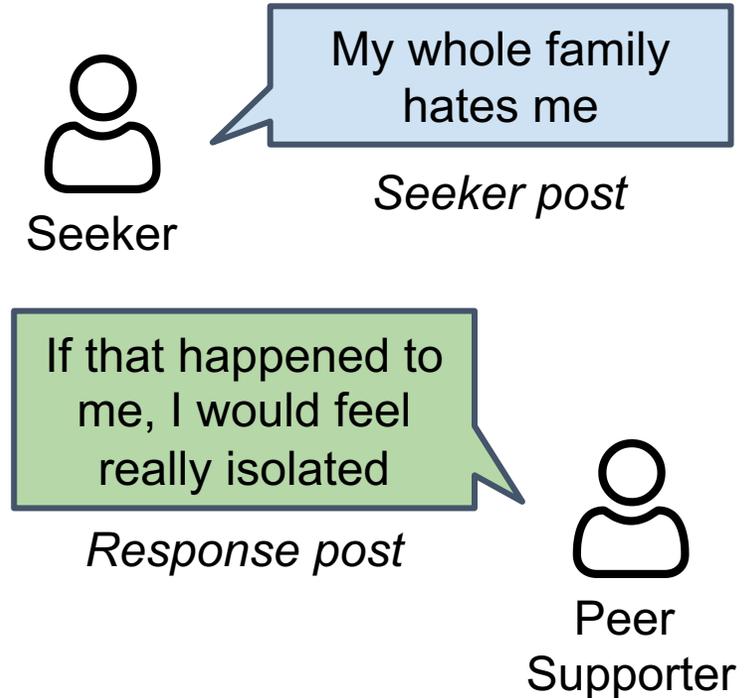
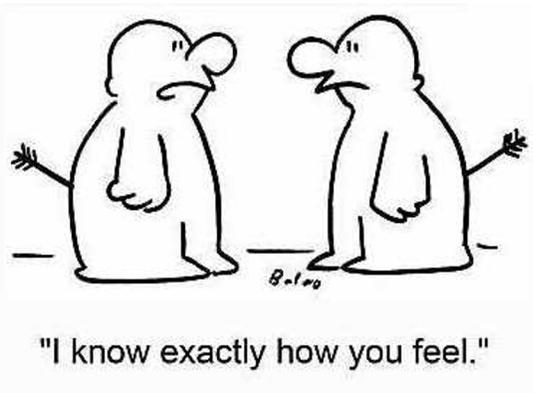
**For online mental health platforms to be helpful, peer-supporters should provide effective support**



**Communication of Empathy  
in Conversation**

# Empathy

- **Core intervention component** of therapeutic counseling
- Quantitative evidence shows **strong associations with positive counseling outcomes** ([Bohart et al., 2002](#); [Elliot et al., 2011](#))
  - Symptom improvement
  - Alliance and rapport



# Empathy: Current limitations and challenges

- Computational methods are **limited to face-to-face, speech-based therapy** ([Gibson et al., 2016](#), [Perez-Rosas et al., 2017](#))



Audio-video signals: pitch, prosody

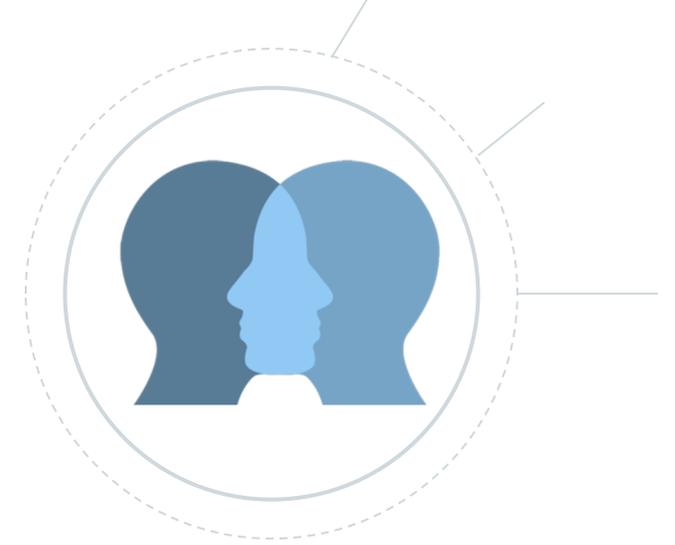
- Previous NLP research focuses on **empathy as reacting with emotions of warmth and compassion** ([Buechel et al., 2018](#)) or as **emotionally-grounded conversations** ([Rashkin et al., 2019](#)).

I feel sorry for you

- **Communicating cognitive understanding of feelings and experiences** of others is key in mental-health support ([Selman, 1980](#))

If that happened to me, I would feel really isolated

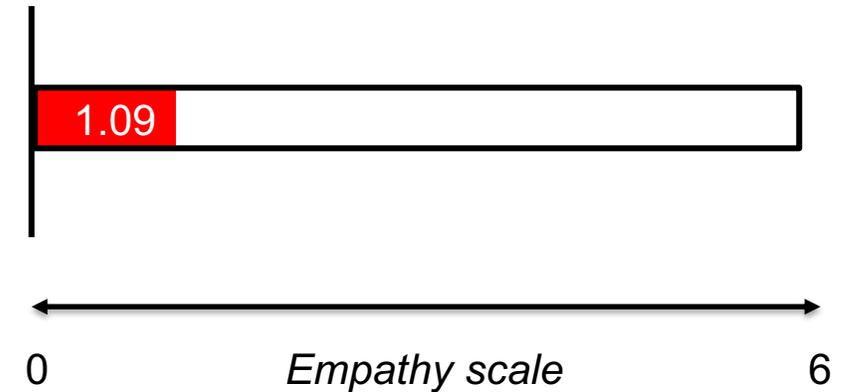
# Empathy expressed in Text-based Mental Health Support



A new  
empathy  
framework and  
dataset

Computational  
model for  
identifying  
empathy

Model-based  
insights into  
mental health  
platforms



Our analysis suggests that **highly empathic conversations are rare** in text-based mental health support

# **A new framework and dataset of empathy expressed in text-based mental health support**

Adapted most prominent empathy scales  
from psychology/psychotherapy research to  
text-based mental health support in  
collaboration with clinical psychologists

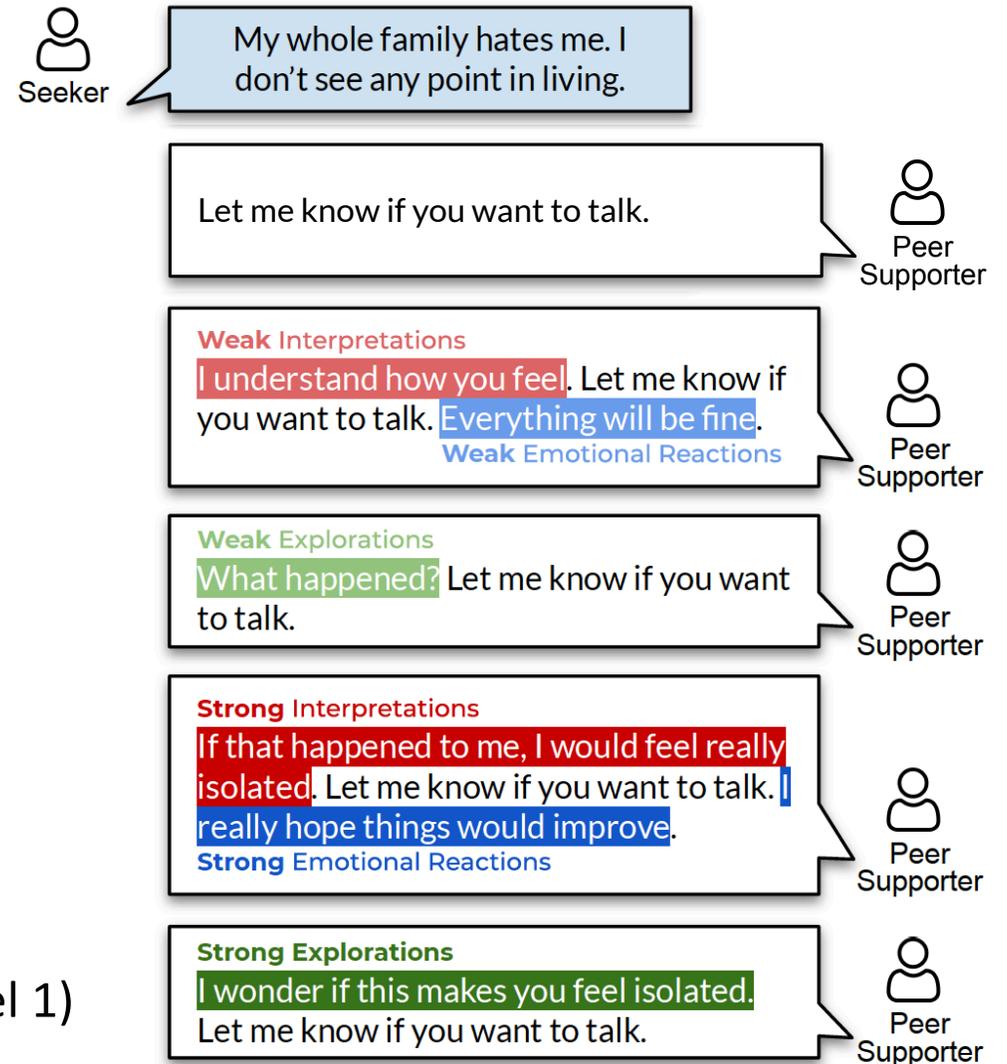
# Framework of empathy expressed in conversations

## Three communication mechanisms of empathy

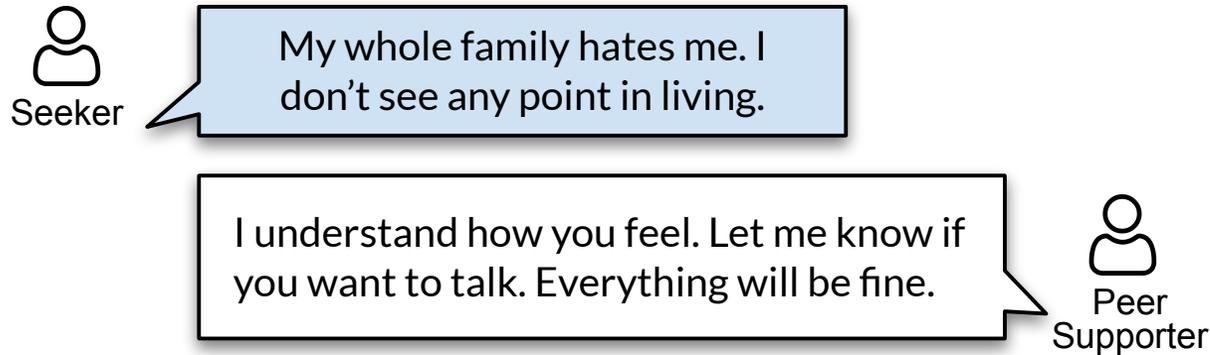
- **Emotional Reactions:** communicating the emotions experienced after reading a post
- **Interpretations:** communicating understanding of the inferred feelings / experiences
- **Explorations:** improving one's understanding by exploring feelings / experiences

We differentiate between

- peers **not** expressing them at all (level 0)
- peers expressing them to some **weak** degree (level 1)
- peers expressing them **strongly** (level 2)



# Tasks and Dataset



## Task 1: Empathy Identification

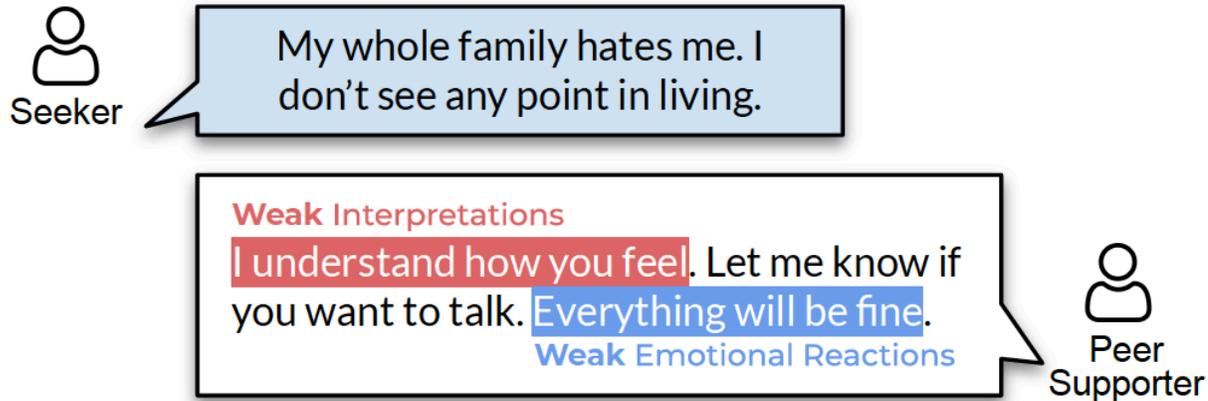
How empathic is response post in the context of seeker post?

Emotional Reactions – 1 out of 2  
Interpretations – 1 out of 2  
Explorations – 0 out of 2

## Task 2: Rationale Extraction

What is the supporting rationale for the identified empathy levels?

# Tasks and Dataset



## Task 1: Empathy Identification

How empathic is response post in the context of seeker post?

Emotional Reactions – 1 out of 2  
Interpretations – 1 out of 2  
Explorations – 0 out of 2

## Task 2: Rationale Extraction

What is the supporting rationale for the identified empathy levels?

# Tasks and Dataset

## Task 1: Empathy Identification

How empathic is response post in the context of seeker post?

## Task 2: Rationale Extraction

What is the supporting rationale for the identified empathy levels?

- **Dataset of 10k (post, response) pairs annotated on our framework of empathy with supportive evidences (*rationales*)**
  - 7k from TalkLife, 3k from mental health subreddits
  - Hired and trained freelancers on Upwork
  - Series of phone calls and manual/automated feedback on sample posts
  - Kappa = **0.6865**

# Model for identifying empathy with supportive rationales

Multi-task, RoBERTa-based bi-encoder  
model

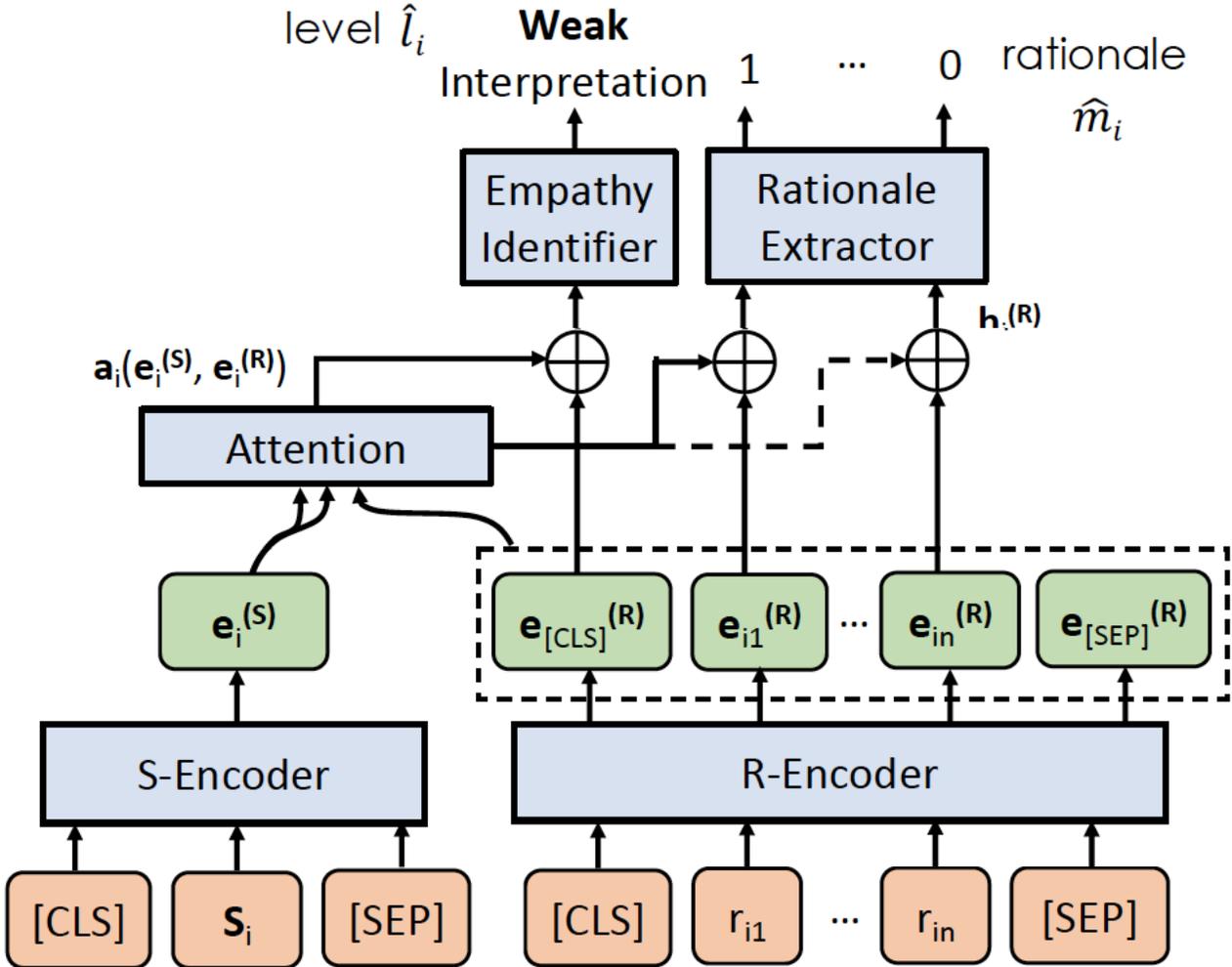
# Our Computational Model

Linear layer for input-level and token-level predictions

Attention between the two encodings

Two independently pretrained RoBERTa-based encoders

Tokenize seeker post and response post



- (1) Empathy identification
- (2) Rationale extraction

(Seeker post, response post)

# Results

- Our approach identifies empathy with **~80% acc., ~70% f1**
  - Random baseline would be 33% accurate

Model	Emotional Reactions		Interpretations		Explorations	
	acc.	f1	acc.	f1	acc.	f1
Log. Reg.	58.02	51.58	55.53	41.19	63.23	51.97
RNN	69.09	54.02	82.25	47.94	73.40	28.22
HRED	78.91	48.70	79.26	29.48	73.40	28.22
BERT	76.98	70.31	85.06	62.24	85.87	71.56
GPT-2	76.89	70.76	80.00	58.43	83.25	65.65
DialoGPT	76.71	70.42	85.67	66.60	83.95	66.34
RoBERTa	78.28	71.06	86.25	62.69	85.79	71.83
<b>Our Model</b>	<b>79.93</b>	<b>74.29</b>	<b>87.50</b>	<b>67.46</b>	<b>86.92</b>	<b>73.47</b>

Low f1-scores, unable to distinguish between the three empathy levels

Bi-encoder architecture works better than concatenating seeker post and response post

## Empathy Identification Task

# Results

- Both token-level (T-f1) and span-level (IOU-f1) evaluation for the rationale extraction task

Model	Emotional Reactions		Interpretations		Explorations	
	T-f1	IOU	T-f1	IOU	T-f1	IOU
Log. Reg.	47.44	63.27	46.92	32.97	47.18	62.25
RNN	62.80	58.22	67.26	57.31	63.29	64.65
HRED	60.56	55.01	64.26	70.92	61.54	70.85
BERT	61.29	51.20	61.06	67.33	62.50	64.80
GPT-2	47.39	51.27	64.06	81.12	<b>66.71</b>	78.21
DialoGPT	66.24	61.24	64.05	79.64	57.95	76.95
RoBERTa	59.12	63.82	60.08	84.85	60.05	78.21
<b>Our Model</b>	<b>68.49</b>	<b>66.82</b>	<b>67.81</b>	<b>85.76</b>	64.56	<b>83.19</b>

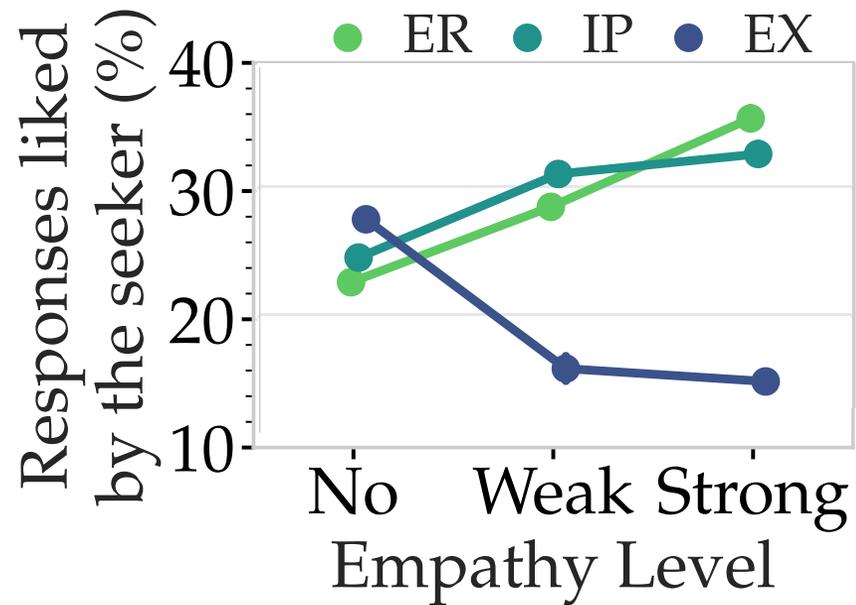
} Baselines with appropriateness to the related task of generating free-text rationales

**Rationale Extraction Task**

# Model-based Insights into Mental Health Platforms

Applied our model to a carefully filtered dataset of 235k interactions of significant mental health challenges on TalkLife(based on categories and triggering posts)

# Validation: Positive feedback from seekers

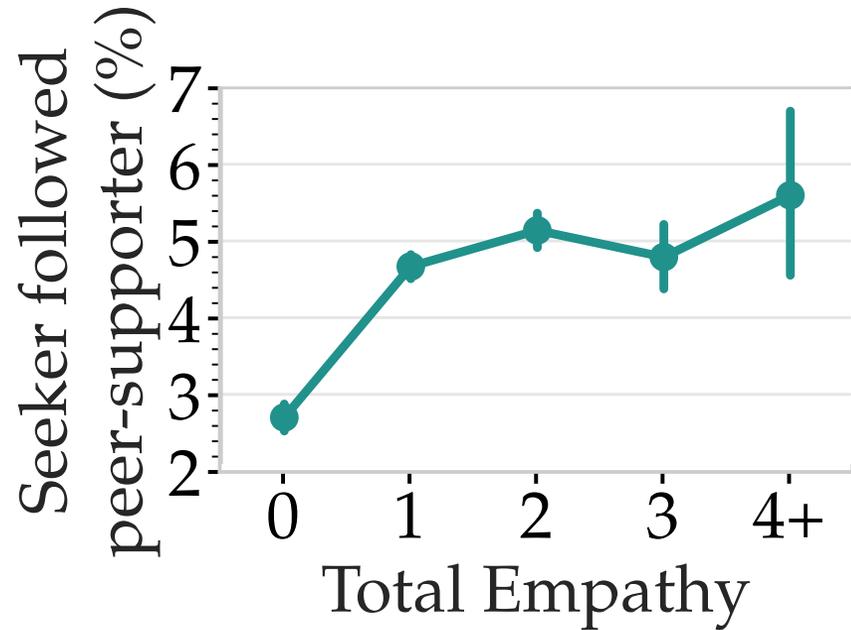


Strong communications of emotional reactions and interpretations receive **45% more likes** than their no communication

Stronger explorations get **47% more replies**

High empathy interactions are received positively by seekers

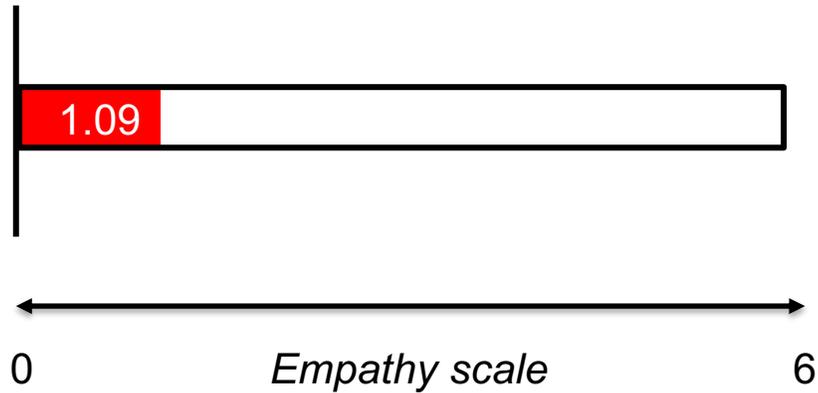
# Validation: Forming of relationships



Seekers are 79% more likely to “follow” peer supporters after an empathic interaction than after a non-empathic one

**Relationship forming more likely after empathic interactions**

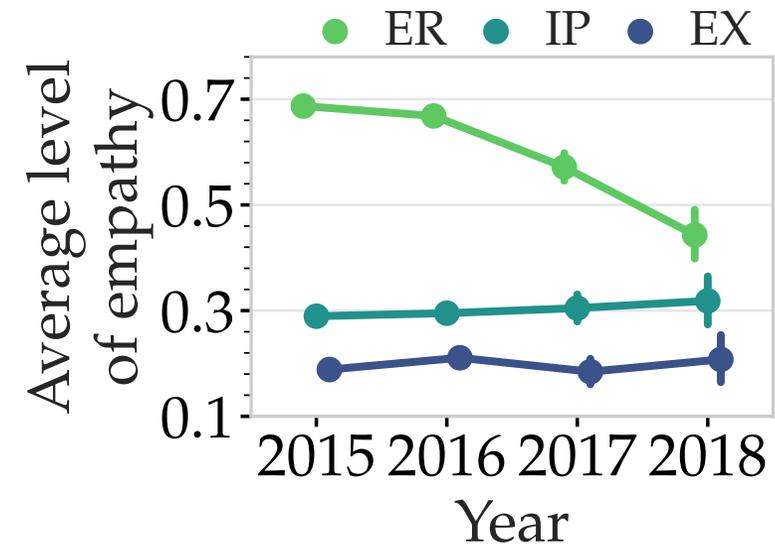
# How empathic are peer supporters?



Empathy expressed is low

- Does it improve over time?

Peer-supporters do not self-learn empathy over time

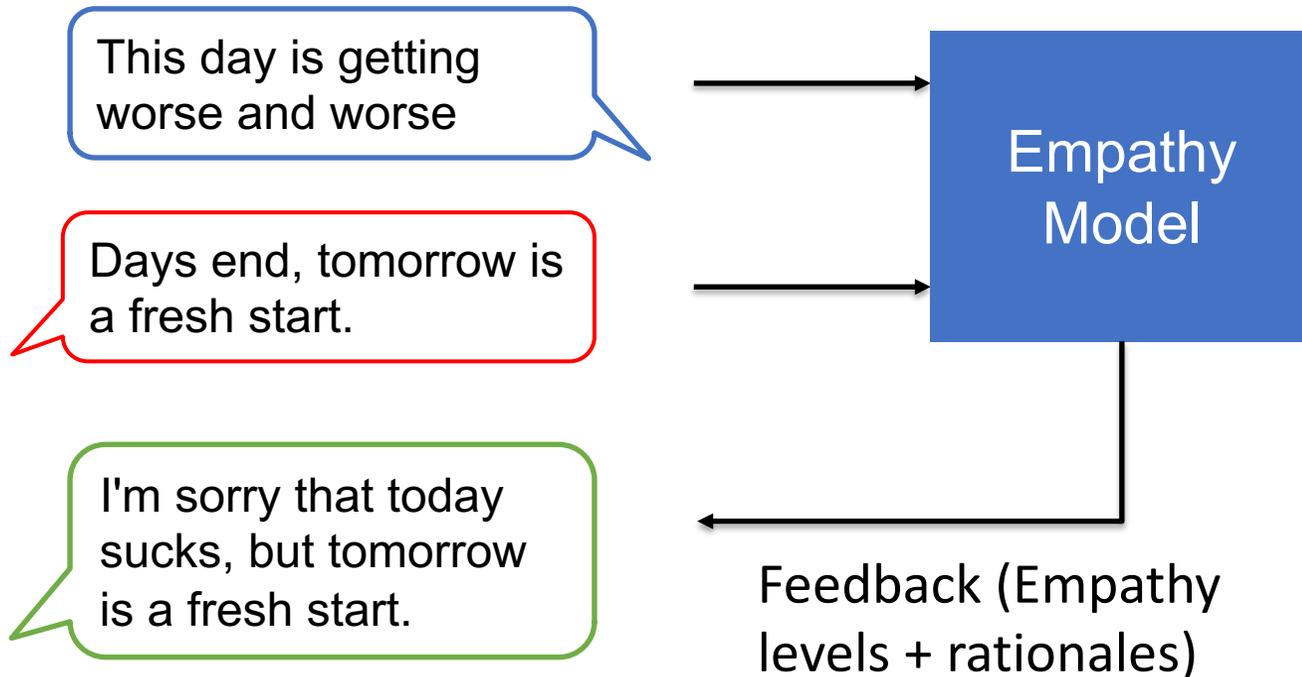


- This is also true for therapists!
  - Without **deliberate practice** and **specific feedback**, even trained therapists often diminish in skills over time ([Goldberg et al., 2016](#))

# Implications for empathy-based feedback

- We can measure empathy successfully, and the measured components are important to mental health platforms
- However, highly empathic conversations are rare
- How can you help people express more empathy?

## Simple Proof-of-concept using model-based feedback



- Three participants were asked to rewrite responses using model-based feedback
- Empathy increased from 0.8 to 3

# Thank You!

- Codes and dataset available at:

<https://github.com/behavioral-data/Empathy-Mental-Health>



@sharma\_ashish\_2



[ash-shar.github.io](https://ash-shar.github.io)