

Author(s)

Jess Hohenstein (Cornell University) - jch378@cornell.edu
Malte Jung (Cornell University) - mfj28@cornell.edu
Rene Kizilcec (Cornell University) - kizilcec@cornell.edu
Dominic DiFranzo (Lehigh University) - djd219@lehigh.edu
Zhila Aghajari (Lehigh University) - zha219@lehigh.edu

1) Have any data been collected for this study already?

No, no data have been collected for this study yet.

2) What's the main question being asked or hypothesis being tested in this study?

RQ: How do smart replies (i.e., a set of algorithmically-generated suggested responses) influence computer-mediated conversations and interpersonal relationships?

H1: People use smart replies ("sr_use") in conversation when given the opportunity.

H2: Ego's smart reply use ("sr_use") increases the rate of messages sent by the ego ("n_messages"/"time").

H3: Ego's smart reply use ("sr_use") and partner's smart reply use ("sr_use_partner") increase perceived smart reply use of the conversation partner ("perceived_sr_use").

H4a: Perceived smart reply use of the conversation partner ("perceived_sr_use"), accounting for their actual use, decreases the ego's rating of cooperative communication ("coop_comm").

H4b: Perceived smart reply use of the conversation partner ("perceived_sr_use"), accounting for their actual use, decreases the ego's rating of affiliation ("affiliation").

H4c: Perceived smart reply use of the conversation partner ("perceived_sr_use"), accounting for their actual use, decreases the ego's rating of trust ("trust").

3) Describe the key dependent variable(s) specifying how they will be measured.

Behavioral measures obtained from conversation log:

sr_use = proportion of messages sent by the ego that are smart replies

sr_use_partner = proportion of messages sent by the partner that are smart replies

time = number of seconds that the conversation lasts

n_messages = number of messages sent by the ego

Self-report measures obtained from survey after conversation (ratings by ego about partner):

perceived_sr_use = "How often do you believe your partner used smart replies?" on a 5-point scale (1=Never; 5=Always)

trust = 5 items from Wheelless' 7-point trust scale, summed to obtain index

coop_comm = 7 items from the 7-point perceived cooperative communication scale (1=Strongly disagree; 7=Strongly agree), summed to obtain index (4 of 7 items are reverse coded)

affiliation = 16-item IAS-R 7-point scale (1=Extremely inaccurate; 7=Extremely accurate); the two items that make up each of the octants of the Wiggins'

Interpersonal Circumplex (arrogant/calculating (BC), warm-agreeable (LM), etc.) are summed to yield a score for each octant, and affiliation is calculated as $LM - DE + .707 * (NO - BC - FG + JK)$

4) How many and which conditions will participants be assigned to?

4 conditions; between subjects design: In the first condition, both participants will use a messenger that features smart replies. In the second condition, one participant will have smart replies while their partner does not have smart replies. In the third condition, one participant will not have smart replies while their partner has smart replies. In the fourth condition, neither participant will have smart replies.

5) Specify exactly which analyses you will conduct to examine the main question/hypothesis.

We use an instrumental variable analysis with robust and conversation-clustered standard errors.

```
library(AER)
```

```
library(clubSandwich)
```

```
library(estimatr)
```

```
# H1: People use smart replies ("sr_use") in conversation when given the opportunity.
```

```
lm_robust(sr_use ~ has_sr, d, clusters = convold)
```

```
# H2: Ego's smart reply use ("sr_use") increases the rate of messages sent by the ego ("n_messages"/"time").
```

```
iv_inter = ivreg(l(n_messages/time) ~ sr_use | cond, data = d)
```

```
summary(iv_inter, diagnostics = T)
coef_test(iv_inter, vcov = "CR2", cluster = d$convoID)
```

H3: Ego's smart reply use ("sr_use") and partner's smart reply use ("sr_use_partner") increase perceived smart reply use of the conversation partner ("perceived_sr_use").

```
iv_med = ivreg(perceived_sr_use ~ sr_use + sr_use_partner | cond, data = d)
summary(iv_med, diagnostics = T)
coef_test(iv_med, vcov = "CR2", cluster = d$convoID)
```

H4a: Perceived smart reply use of the conversation partner ("perceived_sr_use"), accounting for their actual use, decreases the ego's rating of cooperative communication ("coop_comm").

```
iv_out_cc = ivreg(cooperative_comm ~ perceived_sr_use + sr_use_partner | cond, data = d)
summary(iv_out_cc, diagnostics = T)
coef_test(iv_out_cc, vcov = "CR2", cluster = d$convoID)
```

H4b: Perceived smart reply use of the conversation partner ("perceived_sr_use"), accounting for their actual use, decreases the ego's rating of affiliation ("affiliation").

```
iv_out_af = ivreg(affiliation ~ perceived_sr_use + sr_use_partner | cond, data = d)
summary(iv_out_af, diagnostics = T)
coef_test(iv_out_af, vcov = "CR2", cluster = d$convoID)
```

H4c: Perceived smart reply use of the conversation partner ("perceived_sr_use"), accounting for their actual use, decreases the ego's rating of trust ("trust").

```
iv_out_tr = ivreg(trust ~ perceived_sr_use + sr_use_partner | cond, data = d)
summary(iv_out_tr, diagnostics = T)
coef_test(iv_out_tr, vcov = "CR2", cluster = d$convoID)
```

6) Describe exactly how outliers will be defined and handled, and your precise rule(s) for excluding observations.

We will exclude conversations from the analysis that have <10 messages exchanged overall and where one participant sent <3 messages. For the analysis of post-conversation self-report outcomes, we exclude participants who do not complete the full survey.

```
sub = dat %>%
  group_by(convoID) %>%
  filter(sum(n_messages) >= 10 & min(n_messages) > 2) %>%
  ungroup
```

7) How many observations will be collected or what will determine sample size? No need to justify decision, but be precise about exactly how the number will be determined.

We will collect 400 individual responses, 100 participants (50 conversations) per condition.

8) Anything else you would like to pre-register? (e.g., secondary analyses, variables collected for exploratory purposes, unusual analyses planned?)

We additionally assess and analyze conversation sentiment which we expect to go up slightly with smart reply use simply because smart reply messages tend to have more positive valence: Conversations will be recorded and the sentiment of each message will be analyzed using the polarity score from VADER (Valence Aware Dictionary and sEntiment Reasoner). Message sentiment will be calculated from these polarity scores.