The Clinical Panel:
Leveraging Psychological Expertise During NLP Research

Glen Coppersmith
Qntfy
glen@qntfy.com

Kristy Hollingshead
IHMC
kseitz@ihmc.us

H. Andrew Schwartz
Stony Brook University
has@cs.stonybrook.edu

Molly E. Ireland
Texas Tech University
molly.ireland@ttu.edu

Rebecca Resnik
Rebecca Resnik & Assoc., LLC
drrebeccaresnik@gmail.com

Kate Loveys
Qntfy
kate@qntfy.com

April Foreman
Veterans Affairs
acf@docforeman.com

Loring Ingraham
George Washington University
ingraham@gwu.edu

Abstract

Computational social science is, at its core, a blending of disciplines—the best of human experience, judgement, and anecdotal case studies fused with novel computational methods to extract subtle patterns from immense data. Jointly leveraging such diverse approaches effectively is decidedly nontrivial, but with tremendous potential benefits. We provide frank assessments from our work bridging the computational linguistics and psychology communities during a range of short and long-term engagements, in the hope that these assessments might provide a foundation upon which those embarking on novel computational social science projects might structure their interactions.

1 Introduction

Cross-discipline collaboration is critical to computational social science (CSS), amplified by the complexity of the behaviors measured and the data used to draw conclusions. Academic tradition divides courses, researchers, and departments into quantitative ($\mathbb{Q}$s) and humanities ($\mathbb{H}$s), with collaboration more common within $\mathbb{Q}$ disciplines (e.g., engineering and computer science are required for many pursuits in robotics) than across the $\mathbb{Q}$-$\mathbb{H}$ divide (e.g., computational poetry). Ideally, long term collaborations across the $\mathbb{Q}$-$\mathbb{H}$ divide will serve CSS best, but establishing such relationships is challenging and the success of any pairing is hard to predict. How does one find the most technologically-forward $\mathbb{H}$s? Which are the most patient-centered $\mathbb{Q}$s?

Any cross-discipline collaboration requires bridging a gap with some level of familiarization and adaptation, as well as establishment of common ground, common semantics, and common language (Snow, 1959). With intra-$\mathbb{Q}$ endeavors like robotics, many of these commonalities exist (e.g., everyone involved in the endeavor has likely taken calculus and basic programming classes). CSS, however, draws techniques and deep understanding from both $\mathbb{Q}$ and $\mathbb{H}$ disciplines, which makes establishing such commonalities an even larger task. This paper outlines the various ways in which the Computational Linguistics and Clinical Psychology (CLPsych) community has bridged the semantic chasm between the required $\mathbb{Q}$ and $\mathbb{H}$ partners, in the hopes that some of the successful techniques and lessons learned can be adapted for other CSS collaborations. We highlight the actions taken by researchers from both sides to cross the $\mathbb{Q}$ and $\mathbb{H}$ divide. Briefly, we believe in the Gestalt of these approaches—they mutually reinforce and serve to establish a community and maintain commonality, even as research progresses. Concretely, we focus on three categories of approaches: integrated conferences (Section 2), a channel for continual exchange (Section 3), and intensive research workshops (Section 4).

Forging a new successful collaboration is tricky, with expectations on both sides often proving to be a mismatch to reality. For example, due to a lack of understanding of how language analyses are accomplished, $\mathbb{H}$ may expect *feats of magic* from $\mathbb{Q}$s, or for $\mathbb{Q}$s to provide an unrealistic amount of tedious data grunt work. On the other side, $\mathbb{Q}$s may expect

132

diagnostic criteria to be concrete or may not fully appreciate the need to control for sample biases. From an $\mathbb{H}$ perspective, social and cultural barriers may prevent engagement with $\mathbb{Q}$s: $\mathbb{H}$ researchers may be more sensitive to prejudices about research methods in the so-called soft sciences, and misunderstandings may emerge from stereotypes about expressive $\mathbb{H}$s and cold $\mathbb{Q}$s (as well as their underlying kernels of truth). $\mathbb{Q}$s may design tools and methods without proper consideration for making them accessible to $\mathbb{H}$ colleagues, or without a patient-centered design for the patients that $\mathbb{H}$s work with. At a minimum, there is a general ignorance to some of the findings in the other field, and in the extreme, there is complete dismissal of others' concerns or research.

In any $\mathbb{Q}$-$\mathbb{H}$ collaboration, the tendency to lapse into using specific semantically-laden terminology may lead to confusion without recognizing that the other side needs more explanation. For examples of this[1], "self-medicate" is a clinical $\mathbb{H}$ euphemism for destructive behavior involving alcohol or drugs. Similarly, the "suicide zone" is a series of related cognitive phenomena sometimes apparent before a suicide attempt. These terms carry a well-understood and experienced semantic context for the practicing clinicians, but the $\mathbb{Q}$ collaborators lack this nuance. Similarly, $\mathbb{Q}$ researchers are familiar with certain methods in presenting results and graphs, so DET-curves and notched box plots are well-understood to $\mathbb{Q}$s, but require explanation and analysis to be informative to many $\mathbb{H}$s. This effect is amplified when working intensely with a dataset, letting the researchers become intimately (and in cases overly) familiar with it, and the assumptions and limitation of it. This highlights a need to take a step back when presenting graphs or other visual results to collaborators on the other side of the $\mathbb{Q}$-$\mathbb{H}$ divide. Creating clear data and result visualizations was a vital lesson learned to interface successfully between $\mathbb{H}$ and $\mathbb{Q}$ collaborators.

Many of the other lessons learned from our collaborations over the years took us back to basics:

1. Ask whether the analysis really answers the question for which it was motivated.

2. Step through each component of a figure (e.g., explain the axes).

3. Present potential conclusions that might be drawn from these results.

4. Allow for questions and discussion at each step.

In addition to familiarity with the data, familiarity with the statistics and data displays can also impede collaborators' understanding of the results. Clinical $\mathbb{H}$s have typically been exposed to statistics courses within their discipline, which likely cover variance, ANOVAs, MANOVAs, t-tests, χ^2, and standard error of measurement. However, exposure of many machine learning approaches to measurement and analysis is not included, although those with more recent training in computational social science may have more familiarity with these stereotypical $\mathbb{Q}$ approaches. Quite aside from techniques, typical ways to report results differ significantly: F-measure, precision/recall, or true positives/true negatives are common for $\mathbb{Q}$s whereas $\mathbb{H}$s are more familiar with sensitivity/specificity. The strength of a $\mathbb{Q}$-$\mathbb{H}$ collaboration comes largely from learning from one another, of learning to take advantage of an $\mathbb{H}$'s strength in hypothesis testing and a $\mathbb{Q}$'s abilities in advanced predictive modeling, computation, and algorithms.

In CLPsych, each side of these nascent collaborations approached a research problem differently—the $\mathbb{Q}$s often favored bottom-up, data-driven analysis rather than the more traditional and top-down approach generally taken by $\mathbb{H}$s first forming then formally testing a series of hypotheses based on prior knowledge. Though these different approaches have many commonalities and may achieve the same goal, initial discussions in some of the collaborations were needed to overcome the hurdle of different starting assumptions. This co-education across the $\mathbb{Q}$-$\mathbb{H}$ divide was, and continues to be, continual process.

2 Psychologists as Discussants

The CLPsych workshops, co-located at computational linguistic conferences since 2014, have been instrumental in bringing together the computational linguistics and clinical psychology communities (Resnik et al., 2014; Mitchell et al., 2015; Hollingshead and Ungar, 2016). These workshops took care

---

[1]Sometimes referred to as code words (http://rationalwiki.org/wiki/Code_word) or groupspeak (Nguyen et al., 2015).

to have the NLP and Psych constituencies integrated at every sensible step: program committee, reviews, dialog, and co-presentation.

The call for papers made explicit that the papers are to be informative to and understood by both the computer science and the psychology constituencies. Papers that did not meet this standard were harshly reviewed and consistently rejected. All papers were reviewed by both computational linguistics and psychology researchers, and authors were given a chance to edit their submissions in response to the peer-review comments prior to the submission of the camera-ready papers. Concretely, this allowed the authors to incorporate the reviewing psychologists' views, even prior to publication and presentation at the workshop.

Once at the workshop, each presentation was followed by a discussant from the other constituency (i.e., each $\mathbb{Q}$ presentation was followed by an $\mathbb{H}$ discussant and vice versa). This discussant had the paper well ahead of time and was given the chance to prepare a presentation to complement or respond to the paper. Without exception, this enriched the presented material with fresh insight from a novel perspective. The discussants served to expose the researchers and audience alike to the way such work is interpreted by the other constituency. Critically, though, the discussants took care to restate some of the assumptions and findings as how they would expect their constituency to phrase and interpret it – which provided a potent method for establishing and reinforcing common vocabulary and semantics. Together, these effects led to strong semantic foundations and ongoing dialogs between constituencies, ultimately giving rise to increased communication between the workshop participants at the workshop itself and throughout the year.

3   Online Communities & Continual Engagement

Early in this process, CLPsych was fortunate that a group of researchers and clinicians from the suicide prevention community ($\mathbb{H}$s) came upon some popular press coverage of recent research and reached out to the $\mathbb{Q}$ researchers involved (Coppersmith et al., 2014a; Coppersmith et al., 2014b). *#SPSM* (Sui-

cide Prevention and Social Media[2]) is a social media community that focuses on innovation in suicide prevention. They have a weekly broadcast from a topic relevant to suicide prevention, and invited some of the CLPsych work to be presented. Since the first meeting in February 2014, a number of the NLP members ($\mathbb{Q}$s) from CLPsych have been guests on their show, where they have been able to discuss with a primarily $\mathbb{H}$ panel and audience the myriad ways in which research in this space may inform suicide prevention and mental healthcare more generally. *#SPSM* was keen to bring NLP and data science researchers into their community and provided a platform for continual dialog.

Through this platform, the $\mathbb{Q}$-$\mathbb{H}$ dialog was able to extend outside the context of workshops and move to a less-formal conversational style, such that NLP members of the CLPsych community received deeper exposure to clinicians who might eventually benefit from their research. This dialog begat familiarity and lowered the barrier for interaction— common semantics and language were established, which allowed for efficient communication of ideas, preliminary results, and next steps for the $\mathbb{Q}$ researchers who became part of this community.

Beyond the direct effects on research, the *#SPSM* community has also trained the $\mathbb{Q}$ researchers of some of the unwritten rules, cultural norms, and social codes of the mental health community. While mental health might be an extreme case in their sensitivity to language usage, given the discrimination many in the community face, all fields have some equivalent linguistic, political, or historical touchpoints. For example, the colloquial phrase "commit suicide" carries with it a strong negative connotation for the result of a neurological condition, as the term "commit" has a generally negative connotation associated with criminal behavior. Anyone unaware that the suicide prevention community tends to use "die by suicide" in place of "commit suicide" will inadvertently be perceived as crass, discriminating, and out-of-touch with the community that might benefit from their research (Singer and Erreger, 2015).

The *#SPSM* community helped the $\mathbb{Q}$ researchers to understand the important context of their work and the realities of the mental healthcare system. Access

---

to the community also helped to impress upon ℚ researchers the potential impact of the work they are doing, encouraging the work to continue and reshaping it for greater impact and utility. New partnerships have been borne out of online discussions. In turn, the ℚ researchers helped the *#SPSM*'ers to understand the realm of the possible in data science. Informed discussion of data, access, and policy has become a recurring *#SPSM* theme.

From this ℚ-ℍ partnership, the ℍs came to understand what was needed to do successful ℚ research—labeled data—and became advocates for that. The ℍs were able to clearly articulate the barriers to releasing some of the most sensitive data, and collectively the ℚs and ℍs created a method to gather the data necessary to support research (at the data donation site `OurDataHelps.org`) and work with the mental healthcare and lived experience communities to spread the word and collect donations.

4   The Clinical Panel

The CLPsych community was given a chance to work together in a concerted manner at the 2016 Frederick Jelinek memorial workshop, hosted by the Center for Language and Speech Processing at Johns Hopkins University (Hollingshead et al., 2016). Novel datasets were made available for the workshop to advance the analysis of mental health through social media:

1. The Social Mediome project at the University of Pennsylvania provided electronic medical records and paired Facebook data from users who opted in to the study (Padrez et al., 2015);

2. Qntfy provided anonymized data from users who discussed mental health diagnoses or suicide attempts publicly on Twitter (Coppersmith et al., 2015; Coppersmith et al., 2016); and

3. `OurDataHelps.org` provided anonymized Twitter data for users who attempted suicide.

A team of researchers, primarily ℚs and primarily from NLP and data science, came to Johns Hopkins University for 6 weeks to explore temporal patterns of social media language relevant for mental health. In order to make sure the analyses were on the right path and to get some of the benefits of the CLPsych discussants in real time, a *clinical panel* was formed.

This panel was comprised of practicing clinicians, people with lived experience with mental health issues, epidemiologists, and psychology researchers. This was, from the start, an organic non-hierarchical cross-disciplinary experience, as we set out to establish precedent for a mutually respectful and collaborative environment.

During a weekly one hour video conference, the fulltime workshop researchers presented findings from the week's analysis, and were able to raise questions from the data. The ℍs on the panel continuously translated the visual to the clinical. The clinical panel was quick to offer corroboration, counterfactuals and alternate explanations to the presented results, as well as suggesting follow-on analyses. In some cases, these follow-on analyses led to productive lines of research with clear clinical applications. At the same time, it was difficult to maintain a balance between the ℚ-proposed lines of research on changes in language over time and meeting some of the ℍ shorter-term questions on changes in behavior over time, unrelated to language.

Most importantly, this weekly conference provided the panel a real-time and interactive medium to share their clinical experiences with the NLP researchers performing the analyses. For example, clinicians recounted various phenomena that would show up as increased variability over time. This allowed the NLP researchers to quickly adapt and incorporate measures of variability in all analyses going forward. In another example, one of the key findings from the workshop was inspired by an ℍ suggestion that we try to illuminate the "suicide zone"—a period of time before a suicide attempt where one's behavior is markedly different. Critically, the timeliness of this feedback allowed the adjustment to take place early in the workshop, when there was still sufficient time to adjust the immediate research trajectory. The benefit of this might be most stark when examined in contrast to the (perhaps) yearly feedback one might expect from published papers or conference presentations.

Collectively, both ℚs and ℍs involved in these clinical panels had great respect for each other's expertise, knowledge, and willingness to step outside of their discipline. While this healthy respect made for excellent ongoing interaction, it had a tendency to hamper voicing of criticism early on. With some

Benefits/ Successes	*Video-conference clinical panels*: timely interactive feedback from clinicians on novel data findings.
	Clinicians as discussants: immediate interpretation and feedback to presentations, which builds rapport, common semantics, and common vocabulary.
	Clinicians on program committee: fosters findings that are interesting and accessible to all disciplines.
	Continual engagement: ongoing dialog outside of conferences, which serves to refine common semantic picture.
	Problem framing: initial discussions of experimental setups led to framing data-driven, exploratory analysis as hypothesis-driven tests.
Pitfalls/ Challenges	*Publishing in mainstream NLP conferences*: difficult to balance sophistication of method (highly regarded for NLP publications) with general interpretability (necessary for social scientific impact).
	Long-term goals: expectation of new results at regular collaborative check-ins can motivate a team toward short-sighted tasks.
	Fundamental assumptions: understanding, explicitly stating, and challenging fundamental assumptions can create emotionally charged exchanges.

Table 1: Summarized successes and pitfalls of various collaborative interactions between NLP researchers and psychology experts.

frequency, a contrary view to a publicly-expressed viewpoint was harbored by one of the participants, but only shared privately after the panel rather than voicing it publicly and risking damage to these new relationships. While this has merit to building relationships, it does make rapid scientific progress difficult. We feel that finding ways to foster constructive challenging of assumptions would have made the panel even more effective within the limited duration workshop.

To summarize, the clinical panel provided great benefits in their ability to drive the research in more clinically impactful directions than would come from $\mathbb{Q}$s alone. They also were invaluable in keeping the research aligned with the ultimate goal of helping people and provided a regular source of motivation. This approach is not without a significant startup cost to establish common language and semantics, the occasional danger of shortsighted research tasks before the next weekly meeting, and both sides' reluctance to criticize unfamiliar ideas.

5   Conclusion

As we explore the role that computational linguistics and NLP has in psychology, it is important to engage with clinical psychologists and psychology researchers for their insight and complementary knowledge. Our $\mathbb{Q}$-$\mathbb{H}$ collaborations taught us (1) the power of these collaborations comes from diverse experience, which also means diverse needs, (2) establishing common language and semantics is a continual process, and (3) regular engagement keeps one motivated and focused on the important questions. These partnerships are the result of many forms of continual contact and, most importantly, a mutual respect and desire to see progress.

Acknowledgments

References

Glen Coppersmith, Mark Dredze, and Craig Harman. 2014a. Quantifying mental health signals in Twitter. In *Proceedings of the ACL Workshop on Computational Linguistics and Clinical Psychology*.

Glen Coppersmith, Craig Harman, and Mark Dredze. 2014b. Measuring post traumatic stress disorder in Twitter. In *Proceedings of the 8th International AAAI Conference on Weblogs and Social Media (ICWSM)*.

Glen Coppersmith, Mark Dredze, Craig Harman, and Kristy Hollingshead. 2015. From ADHD to SAD: Analyzing the language of mental health on Twitter through self-reported diagnoses. In *Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, Denver, Colorado, USA, June. North American Chapter of the Association for Computational Linguistics.

Glen Coppersmith, Kim Ngo, Ryan Leary, and Anthony Wood. 2016. Exploratory analysis of social media prior to a suicide attempt. In *Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, San Diego, California, USA, June. North American Chapter of the Association for Computational Linguistics.

Kristy Hollingshead and Lyle Ungar, editors. 2016. *Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*. North American Association for Computational Linguistics, San Diego, California, USA, June.

Kristy Hollingshead, H Andrew Schwartz, and Glen Coppersmith. 2016. Detecting risk and protective factors of mental health using social media linked with electronic health records. In *3rd annual Frederick Jelenik Memorial Summer Workshop / 22nd annual Applied Language Technology Workshop at the Center for Language and Speech Processing at Johns Hopkins University*.

Margaret Mitchell, Glen Coppersmith, and Kristy Hollingshead, editors. 2015. *Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*. North American Association for Computational Linguistics, Denver, Colorado, USA, June.

Viet-An Nguyen, Jordan Boyd-Graber, Philip Resnik, and Kristina Miler. 2015. Tea party in the house: A hierarchical ideal point topic model and its application to Republican legislators in the 112th congress. In *Proceedings of ACL*.

Kevin A Padrez, Lyle Ungar, Hansen Andrew Schwartz, Robert J Smith, Shawndra Hill, Tadas Antanavicius, Dana M Brown, Patrick Crutchley, David A Asch, and Raina M Merchant. 2015. Linking social media and medical record data: a study of adults presenting to an academic, urban emergency department. *BMJ quality & safety*, pages bmjqs–2015.

Philip Resnik, Rebecca Resnik, and Margaret Mitchell, editors. 2014. *Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*. Association for Computational Linguistics, Baltimore, Maryland, USA, June.

Jonathan Singer and Sean Erreger. 2015. Let's talk about suicide: #LanguageMatters. http://www.socialworker.com/feature-articles/practice/lets-talk-about-suicide-languagematters/.

Charles Percy Snow. 1959. *The two cultures and the scientific revolution*. University Press.