

# Understanding Identity Signalling in Persuasive Online Text

Melanie Subbiah, Kathleen McKeown

Department of Computer Science, Columbia University, New York, NY  
m.subbiah@columbia.edu, kathy@cs.columbia.edu

## Abstract

Shared identity can be a powerful tool for persuasion. In online communities, people can recognize shared identity through perceiving identity signals in writing. Analyzing these types of signals in text could therefore indicate the targets of persuasive online strategies, such as misinformation campaigns. In this vision paper, we first elaborate on why identity signalling is an important area to study in persuasive online text, and then outline some initial steps to facilitate research in this area.

## Introduction

As humans, our social nature allows us to work together in groups and thereby benefit from this group membership. As a result, we are skilled at identifying which groups we are considered a part of and which ones we may be excluded from (ingroup and outgroup status). When we see someone as similar to ourselves, we are more likely to trust and feel empathy for them (Anders et al. 2020). We are also more likely to form social groups based on these similarities, and then more likely to share beliefs or assume shared beliefs within these groups (McPherson, Smith-Lovin, and Cook 2001). These groups can influence belief change through social pressure if someone values the social reward of fitting in or when group status is an important part of their identity (Kelman 1961).

This ingroup connection can therefore be manipulated as a method of persuasion when a communicator uses signals of ingroup status to make a message or viewpoint more persuasive to a target audience, such as for targeted disinformation campaigns. Nkonde et al. describe a phenomenon, *disinformation creep*, whereby a campaign manipulates messaging that signals one belief system and set of values to push a contrary message (Nkonde et al. 2021). This work shows that understanding how we signal commonality in online discourse could be an important part of identifying targets of online persuasion, which in turn helps with identifying the intent of that persuasion.

There are several ways people identify and communicate similarities and *ingroup status* with someone else. The way that someone projects these attributes is called signalling

Copyright © 2021, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

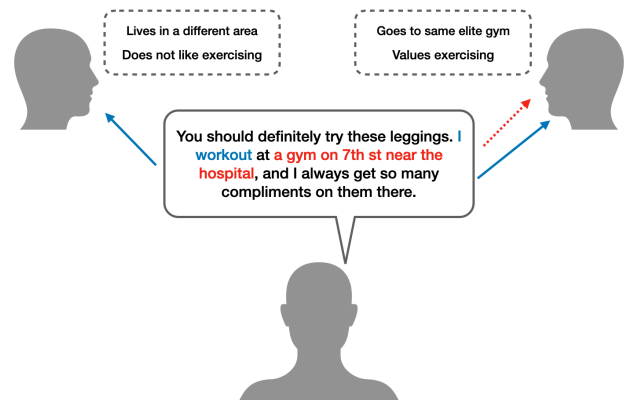


Figure 1: The speaker here is overtly signalling their value of exercising, but is covertly signalling wealth only to the second listener who goes to the same gym and therefore, knows a membership there is expensive.

which has been studied in the social science literature. Signalling can be overt or covert (Smaldino and Turner 2021). Overt signals are ones that can be interpreted by anyone, whereas covert signals are designed to only be interpretable by other members of the signaller’s ingroup. Covert signalling can be used to either maintain some sort of elite or exclusive status or, in cases where the signaller could experience a negative consequence from being identified by an outgroup member. For example, gay men have historically used details of their dress to communicate their sexual orientation only to other gay men (Fischer 2015).

We believe that understanding the ways signalling is used in persuasive online discussions is an important unexplored area at the intersection of several disciplines.

## Background

Previous NLP work on persuasion has primarily focused on identifying whether or not a particular piece of text accomplished the goal of persuading someone. Work in this area has used a variety of approaches and has looked at news editorials (El Baff, Wachsmuth, and Al-Khatib 2020), reddit forums (Hidey and McKeown 2018), and online debate forums

(Durmus and Cardie 2019a), among other sources (Pryzant, Chung, and Jurafsky 2017). However, much of this work has focused on the content and style of the writing and not on the interpersonal dynamics between the person persuading and the person being persuaded.

Some previous work has begun to explore this area. As some examples, Durmus and Cardie (2019b) demonstrated how including a feature that represented the similarity of two users' views on some core issues improved the prediction of whether an argument on an unrelated issue was considered persuasive or not. Shmueli-Scheuer et al. showed that the interaction between the personality types of an author and reader affects persuasion outcome (Shmueli-Scheuer et al. 2019). And Tan et al. observed that the interplay in content and style used between the participants in a debate was a significant predictor of persuasive outcome (Tan et al. 2016).

## Towards Detecting Identity Signalling in Persuasive Research

Previous work demonstrates that there are clear interactions between participants' prior beliefs and values and what is persuasive for them. However, there is a need for a broader study that looks at a wide range of prior beliefs and values someone might hold, how they signal these pieces of their identity, and how they adapt the persuasive strategies they employ when communicating with someone more similar to or more different from them. For example, for someone who centers religion in their life, persuasive big picture arguments would likely need to leverage this framework. So how could someone identify this attribute of the person they are trying to persuade and signal a shared understanding? Does providing an ingroup signal advantage the persuader?

These questions are challenging from a computational perspective in part because it is difficult to find data where we know the beliefs and value systems of the individuals involved, both in terms of the producer and the consumer of the text. It could be hard to produce this data even with unlimited resources because individuals may not be aware themselves of all the subtle cues they are signalling or understanding in persuasive text. Perhaps for this reason, the previous work described above has focused on predicting persuasive outcome. When prior beliefs and values do come into play, they are mostly focused on political ideology or discourse style as these qualities can be easier for people to self-identify or to identify in others.

We believe creating a dataset is an important next step to enable research on this problem. This dataset would need annotated spans indicating signalling intent, annotated spans for persuasive techniques used, and information about annotators' and writers' identities and their ratings of the persuasiveness of the different texts. The annotators would likely need to participate in the debates themselves so that we would have full identity information on the writers. Such a dataset would allow us to further understand how signalling interacts with persuasive techniques. We may also find that signalling attempts are often unsuccessful. This dataset would allow us to answer questions such as: Is signalling performed primarily through content selection or dis-

course style? How do writers adapt their signalling and persuasive techniques based on the intended audience?

We believe the successful creation of a dataset like this would require inter-disciplinary collaboration ideally between computer scientists, linguists and social scientists due to the challenges involved with data collection and annotation. For example, particularly when it comes to covert signals, many annotators would miss signalling attempts in text. We would therefore require a diverse pool of annotators, and could use the feedback from different annotators on the same passage to identify which signalling attempts may be covert vs. overt. Another potential concern is that there is a fine line between successful signalling and biased assumptions. For example, an annotator may think that someone is signalling their gender by talking about shopping for clothes, but this thought could actually reflect a bias from the annotator that enjoying shopping must be associated with a specific gender.

## Conclusion

Understanding how identity signalling interacts with persuasion in online text is an important problem which could facilitate the identification of targeted online campaigns. However, making progress in this area will require expertise from social science, linguistics, and natural language processing.

## References

- Anders, S.; Beck, C.; Domin, M.; and Lotze, M. 2020. Empathic responses to unknown others are modulated by shared behavioural traits. *Scientific Reports* 10: 1938.
- Durmus, E.; and Cardie, C. 2019a. A Corpus for Modeling User and Language Effects in Argumentation on Online Debating. *arXiv: 1906.11310v1 [cs.CL]*.
- Durmus, E.; and Cardie, C. 2019b. Exploring the Role of Prior Beliefs for Argument Persuasion. *arXiv: 1906.11301v1 [cs.CL]*.
- El Baff, R.; Wachsmuth, H.; and Al-Khatib, K. Stein, B. 2020. Analyzing the Persuasive Effect of Style in News Editorial Argumentation. In *Analyzing the Persuasive Effect of Style in News Editorial Argumentation*, 3154–3160.
- Fischer, H. 2015. *Gay Semiotics: A Photographic Study of Visual Coding Among Homosexual Men*. Cherry and Martin.
- Hidey, C.; and McKeown, K. 2018. Persuasive Influence Detection: The Role of Argument Sequencing. In *The Thirty-Second AAAI Conference on Artificial Intelligence*.
- Kelman, H. 1961. Processes of Opinion Change. *Public Opinion Quarterly* 25(1): 57.
- McPherson, M.; Smith-Lovin, L.; and Cook, J. 2001. Birds of a Feather: Homophily in Social Networks. *Annual Review of Sociology* 415–444.
- Nkonde, M.; Rodriguez, M.; Cortana, L.; Mukogosi, J.; King, S.; Serrato, R.; Martinez, N.; Drummer, M.; Lewis, A.; and Malik, M. 2021. Disinformation creep: ADOS and the strategic weaponization of breaking news. *Harvard Kennedy School (HKS) Misinformation Review* 1: 7.

Pryzant, R.; Chung, Y.; and Jurafsky, D. 2017. Predicting Sales from the Language of Product Descriptions. In *Proceedings of the SIGIR 2017 eCom workshop*.

Shmueli-Scheuer, M.; Herzig, J.; Konopnicki, D.; and Sandbank, T. 2019. Detecting Persuasive Arguments based on Author-Reader Personality Traits and their Interaction. In *Proceedings for ACM UMAP 2019 Main Track*.

Smaldino, P.; and Turner, M. 2021. Covert signaling is an adaptive communication strategy in diverse populations.

Tan, C.; Niculae, V.; Danescu-Niculescu-Mizil, C.; and L., L. 2016. Winning arguments: Interaction dynamics and persuasion strategies in good-faith online discussions. In *In Proceedings of the 25th International Conference on World Wide Web*, 613–624.