Tracking opinion convergence online: the effect of facial attractiveness

Marina Terkourafi[a] and Sarah Lord[b]
[a]Leiden University Centre for Linguistics and [b]University of Illinois at Urbana-Champaign

Accepted in *Internet Pragmatics*

Abstract
We investigate whether facial attractiveness, as one source of positive/negative attitudes towards one's conversational partner, affects the degree and type of opinion convergence online, even in the absence of physical co-presence. Our hypothesis is that when you interact with someone you find attractive, opinion convergence will occur even if you are not physically co-present with them. Additionally, we tracked different types of opinion convergence (one-sided or mutual) and how convergence is linguistically negotiated in these different circumstances. Our hypothesis was confirmed, to a point. Opinion convergence was most frequent among Attracted pairs; however, opinion convergence was greatest among Neutral pairs. Opinion convergence was qualitatively different in the 3 conditions. This research adds to previous studies which highlighted aspects of communication unique to online environments (anonymity, invisibility) to explain the heightened tendency for face-threatening behaviours to occur online, by showing how implicit biases (operationalized here as facial attractiveness) can be an additional factor influencing online behaviour.
(155 words)

Keywords: anonymity, facial attractiveness, opinion convergence, implicit bias

1. Introduction: online rudeness and the invisible other

The age of the internet is also widely acknowledged to be a rude age.[1] Flaming, trolling and cyberbullying[2] have become part of our everyday vocabulary in a short period of time and their wide-ranging effects are being felt in multiple domains of our public and private lives, from our political cultures (Nithyanand et al. 2017) to our emotional well-being (Satz et al. 2017). While much of this research focuses on the influence of online speech on offline behaviour, our focus in this article is in the opposite direction. Specifically, we are interested in how attitudes we hold prior to engaging in online interaction may affect the outcome of that interaction. That may seem like an obvious point to make, yet previous research has focused by and large on aspects of internet-mediated communication that are unique to this medium, ignoring aspects of the communicative situation that may be shared between the online and offline channels. The small-scale study we present in this article aims to begin to redress this balance.

In a special report by TIME magazine (Stein 2016), Jessica Moreno, a former Reddit[3] employee targeted for her attempts to curtail bad behaviour online, provides the following disconcerting profile of the typical troll:

---

[1] http://www.huffingtonpost.com/janet-kinosian/ours-is-a-rude-age-but-ha_b_194895.html; last accessed August 16, 2017.

[2] Despite some overlap between these terms, flaming is typically used to refer to impoliteness in response to a perceived threat, while trolling is reserved for impoliteness as an end in its own right and cyberbullying, which can encompass both, tends to be preferred for online harassment involving teenagers.

[3] Founded in 2005, Reddit is a U.S.-based platform for aggregating social news, rating web content and providing anonymous commentary on multiple topics (subreddits), that defines itself as "the frontpage of the

> "The idea of the basement dweller drinking Mountain Dew and eating Doritos isn't accurate," she says. "They would be a doctor, a lawyer, an inspirational speaker, a kindergarten teacher. They'd send lovely gifts and be a normal person." These are real people you might know, Moreno says. There's no real-life indicator. "It's more complex than just being good or bad. It's not all men either; women do take part in it." (Stein 2016: 30)

Whitney Phillips, a University professor cited in the same report, echoes Moreno's views:

> "Trolls are portrayed as aberrational and antithetical to how normal people converse with each other. And that could not be further from the truth," […]. "These are mostly normal people who do things that seem fun at the time that have huge implications. You want to say this is the bad guys, but it's a problem of us." (Stein 2016: 28)

In an early attempt to explain what might drive otherwise "normal" people to misbehave to various extents online, psychologist John Suler (2004) described the online disinhibition effect. According to this, six factors intersect in complex ways to produce the amplifying effect on acting out behaviour that is frequently observed online. These are: dissociative anonymity, invisibility, asynchronicity, solipsistic introjection (feeling as if when communicating online one is conversing with oneself), dissociative imagination (treating online communication as more game-like than real), and minimization of (real-life) authority and status (another version of the internet as "the great equalizer"; Truong 2015). Subsequent research has elaborated on these factors by showing that lowered social control can be the result of momentarily heightened self-esteem, occurring when one is essentially 'cheered on' by a crowd of agreeing onlookers (Wilcox & Stephen 2013), a situation which can build up to a more enduring phenomenon of subversive identity formation as has been observed, for instance, among Korean teenagers (Jang et al. 2014, Park et al. 2014).

Prominent among the six factors underlying the online disinhibition effect is anonymity, which can be defined as a condition where the originator of the information is absent or not identifiable (Scott 2004). The effects of this lack of possible identification operate on multiple levels, from the absence of paralinguistic and other interactional cues (visual anonymity), to the lack of accountability that follows when future interaction is unlikely (discursive anonymity).[4] As Misoch (2015: 536) points out, "both kinds of anonymity can easily be realised on the internet through the use of pseudonyms, or by using services that have reduced communication channels (e.g., text only)." These different levels mean that anonymity is not an all-or-nothing property of online communication but is best conceptualized as a matter of degree.

internet". As of April 2017, it ranked fourth among most visited websites in the U.S. and ninth in the world, with 542 million monthly visitors.

[4] In this respect, internet communication may be thought to have succeeded in creating the ultimate situation of face-less communication: if the option of borrowing your neighbour's lawn mower isn't there (cf. Leech 1983: 82), there is little reason to worry that the channel of communication between you will break down and therefore less need to be polite to them in the first place.

2. Previous research on the effects of lack of physical co-presence

In an inspiring study, Lipinski-Harten & Tafarodi (2013) set out to investigate to what extent disagreeing strangers may end up converging in their opinions about a topic if they interact face to face compared to online. Attitude moderation is an important part of negotiating disagreement. In the absence of such moderation, disagreement can escalate to wholesale conflict, as is often observed online with the detrimental effects noted above. Their hypothesis was that, since text-based online communication offers less possibilities for developing affinity among strangers, opinion convergence would occur less in online compared with face-to-face communication. To test this hypothesis, they asked 142 undergraduates from the University of Toronto to rate on a scale from 1 to 10 their degree of agreement with four statements about controversial issues (free higher education, taking another's life, becoming a Canadian citizen, and voting in federal elections). Between one and three months later, they invited the subjects to converse in pairs about one of these statements, about which they held either consistent or opposing opinions. During testing, subjects were presented with the critical statement -- the one they unknowingly agreed or disagreed about – and asked to discuss their thoughts and feelings about it for 20 minutes. Subjects interacted in one of two conditions, either face-to-face (FTF) or online (OC). After the conversation, they were asked to rate on a scale from 1 to 10 their degree of agreement with 12 statements, including the critical one. The variable of interest for the experimenters was the change in reported agreement with the critical statement from the pre-test to post-conversation. This was taken to indicate the effect the conversation had, if any, on the opinions held by the subjects.

The experimenters found that "students matched with partners who clearly disagreed with them on the issue under discussion showed greater attitude moderation after 20 min[utes] of face-to-face conversation than after the same length of online chat" (Lipinski-Harten & Tafarodi 2013: 2492). Specifically, those in the FTF condition showed almost twice as much attitude moderation (that is, their opinions converged almost twice as much) after the conversation as those in the OC condition. To explain these results, Lipinski-Harten & Tafarodi (2013: 2492) proposed that

> "the invisibility and inaudibility of the other in online chat would reduce the felt mental and ethical presence of one's partner and promote greater focus on one's own thoughts and feelings at the expense of understanding and acknowledging the validity of the partner's sentiments and position […] Insofar as participants in online chat remain more attuned to themselves than to those who manifest as mere text on a screen, their susceptibility to dissuasion away from strongly held beliefs and commitments would be low. The visual, auditory, and physical confrontation that defines face-to-face conversation forces each speaker into a deeper engagement with the subjective world of the other…"

These results are consonant with the Social Model of Deindividuation Phenomena (SIDE; Reicher et al. 1995) that predicts that reduced awareness of others' feelings can lead to a tendency to be more aggressive and less considerate of others on social media, as has also been shown in numerous linguistic studies (e.g., Lorenzo-Dus et al. 2011, Moor et al. 2010, *inter alios*).

3. Motivation for the present study and research questions

While we find these results and the conclusions based on them intuitively appealing, we also believe that this earlier research glosses over some important differences between face-to-face and online communication, closer attention to which could reveal additional sources for the observed convergence in opinions that go beyond the specificities of the online medium. Specifically, this earlier study conflates the effects of invisibility and inaudibility with the effects of privately-held attitudes and implicit biases which can also be triggered by physical co-presence and are known to be a powerful factor in utterance interpretation.

Implicit biases are favourable and unfavourable assessments of other people that we develop from an early age based on characteristics such as gender, age, race, ethnicity, and appearance. These assessments do not necessarily align with our explicitly held beliefs, yet they affect our understanding, decisions and actions in an unconscious manner.[5] Research in this vein suggests that face-to-face conversation affords us with a lot more than the possibility of direct engagement with another's paralinguistic cues. Physical co-presence during face-to-face conversation can also trigger implicit biases, that is, positive as well as negative attitudes toward the other on account of visible physical traits, religious symbols etc., before any interaction occurs, and these can variably impact attitude moderation above and beyond any communicative cues directly associated with the delivery of one's message in real time (tone of voice, facial expression, body language etc.). In other words, it is impossible to know, from the Lipinski-Harten & Tafarodi study alone, to what extent physical co-presence in the FTF condition may have triggered positive attitudes due to physical appearance or other visual cues that facilitated opinion convergence independently of – or, cumulatively with – the "subjective engagement with the world of the other" that these authors consider to be the hallmark of face-to-face conversation but also "more taxing" inasmuch as it involves close attention to physical and auditory cues and provides less possibility for expressive and editorial control (Lipinski-Harten & Tafarodi 2013: 2490).

Another aspect that this earlier study sheds little light on concerns how opinion convergence was linguistically realized. Specifically, researchers did not distinguish one-sided opinion convergence (where only one party shifts closer to the opinion of the other) from mutual opinion convergence (where both parties shift toward each other). They also did not track potential qualitative differences between these different types of opinion convergence.

Taking our cue from these observations, we designed a follow-up study which aimed to disentangle the effect of physical co-presence from the effect of implicit biases triggered by visual information and to observe the effect on opinion convergence online of the latter alone. Since we were only interested in the extent to which opinion convergence would occur online in these two different sets of circumstances, our experiment was conducted solely online (i.e. we did not compare FTF and OC conditions). This type of experimental design has a couple of advantages. It is closer to the reality of our daily interactions with others online, where clues to our interlocutor's identity (from actual names, to pseudonyms, which are often meaningful, or even dialectal turns of phrase in their speech) can trigger implicit biases toward them, even in the absence of prior acquaintance. And it operationalizes anonymity as a matter of degree (Misoch 2015), in that discursive anonymity can be attenuated along the lines described above (availability of (first) names, pseudonyms, dialectal cues), at the same time as visual anonymity is maintained.

[5] For a recent overview of implicit bias research, see Brownstein 2017.

To do this, we chose to focus on facial attractiveness, which can trigger positive assessment on multiple levels, from increased liking to higher grades in educational settings and higher salaries in professional ones (Kwan & Trautner 2009) based on visual cues alone. Facial attractiveness is, of course, a different kind of source for implicit bias effects from the discourse-based ones listed above, which are more frequently available in online communication between strangers. Nevertheless, what we were interested in was the possibility of triggering implicit biases in the absence of actual physical co-presence. Inasmuch as facial attractiveness can do that, it is as good a source as any of implicit bias effects. Facial attractiveness also has another advantage in this respect: while affiliative reactions based on first names, pseudonyms or dialectal background can be highly subjective, research on facial attractiveness has shown that this is assessed based on both universal and culture-specific criteria (Langlois et al. 2000; Coetzee et al. 2014). This means that consensus about facial attractiveness among members of a group can be expected to be more easily forthcoming compared with consensus about discourse-based sources of implicit biases. Facial attractiveness thus offers a fertile ground for testing the effect of implicit biases and attitudes held prior to the interaction on the outcome of the interaction itself above and beyond other visual information which is provided through the interlocutors' physical co-presence in real time.

Our research hypothesis was that when interacting with someone you find attractive, this will positively influence whether (and potentially how) you converge with them even if you are not physically co-present with them. To find out what opinion convergence looks like in these different circumstances, we tracked the direction of opinion convergence (whether this was one-sided or mutual), as well as how opinion convergence was linguistically constructed through various markers (see 5.2 below).

4. Methodology

Our methodology was inspired by that of Lipinski-Harten and Tafarodi (2013), adapted to answer the research questions above, and, like theirs, it included three stages. Stage I was a screening survey, stage II involved an online chat, and stage III consisted in a post-survey that subjects filled in individually. The three stages were run in parallel until a total of 30 online chats (ten in each of three conditions, see 4.3 below) were collected.

4.1 Participants

Sixty undergraduate students at a large Midwestern University (22 M; 38 F) were selected for participation based on their answers to the screening survey from a larger pool of applicants recruited through posting on various online platforms. All were native speakers of English, who had spent at least 10 years living in the US (participants who answered negatively to this question were filtered out after the screening survey). Survey respondents selected to participate in the study were compensated $10 for their participation. The study was approved by the University of Illinois Institutional Review Board and run according to the provisions for ethical online research.

4.2 Screening survey

The aim of the screening survey was to gather information about potential participants that would later be used to match them into disagreeing pairs in one of three conditions: Attracted, Unattracted, and Neutral. To gather this information, potential participants were

first instructed to complete an online survey posted on Qualtrics (https://www.qualtrics.com/) which consisted of a variety of general questions pertaining to:[6]

(a) Basic demographic information (race, major, country of origin, gender identity)
(b) Their engagement with different online activities (social media, gaming, email, etc.) reported in hours per week
(c) Who they found attractive (Male, Female, both, open)
(d) Distractor questions about other preferences (favourite food / music / sport)

Additionally, the survey was used to gauge initial (pre-test) participant opinions about a variety of commonly debated topics that could generate conflict in conversation. To accomplish this, participants were asked to indicate on a scale from 1 to 10, 1 meaning "strongly disagree" and 10 being "strongly agree," whether they agree or disagree with the twelve controversial statements shown in Fig. 1.

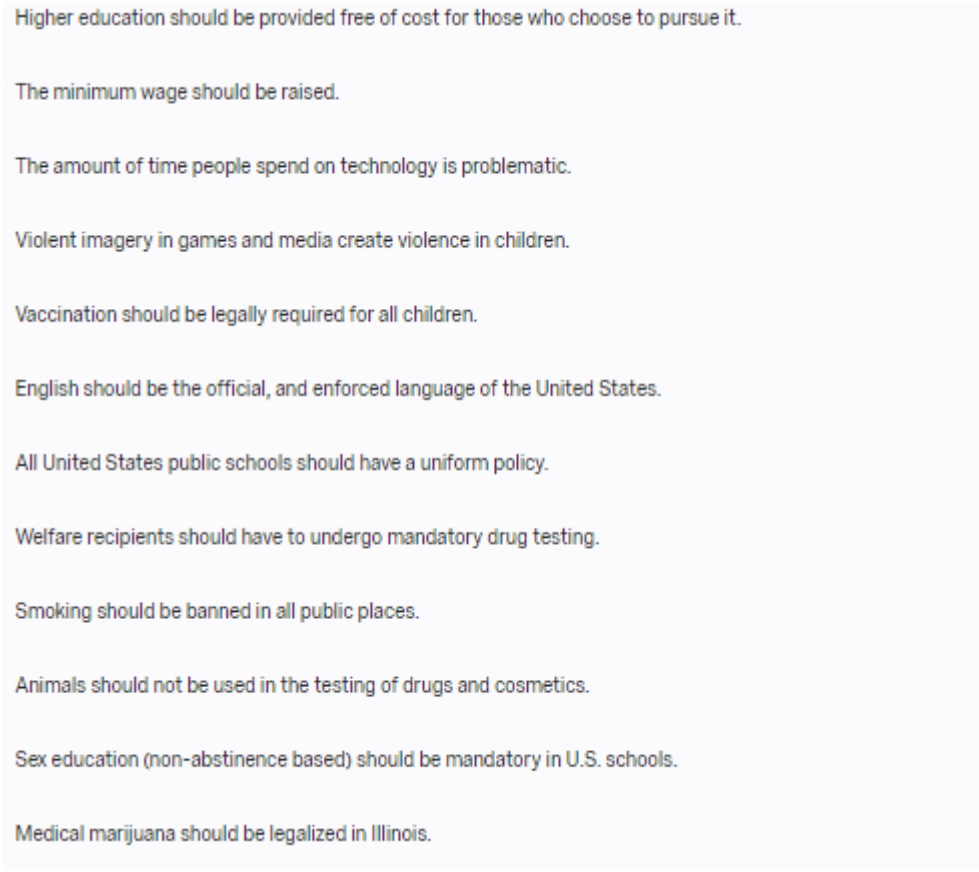

Fig. 1: Potentially controversial statements used in the screening survey

Finally, potential participants were shown photos of varying ethnicities from the Chicago Face Database (http://faculty.chicagobooth.edu/bernd.wittenbrink/cfd/index.html) of the gender identity they had indicated they found attractive in response to the initial set of questions. This information was used to determine consensus on the male and female faces users found most and least attractive. The male and female faces rated as most and least

[6] For the full text of the screening survey, see Appendix I.

attractive by participants (see Fig. 2) provided the visual cues for the Attracted and Unattracted conditions respectively during the second stage of the study (online chat).

Fig. 2: Female and male faces from the Chicago Face Database rated more often as attractive (left) and more often as not attractive (right) by participants in the screening survey

4.3 Online chat

The second stage of our study was the online chat itself. Chats were scheduled for a twenty-minute time slot where both participants could attend and be monitored. The goal of this stage was to observe a possibly argumentative chat under different conditions of facial attractiveness while also finding out the linguistic correlates of opinion convergence under these circumstances.

Once the screening survey was completed, participants were selected from the pool of fully completed surveys to participate based on scheduling, disagreement ratings on at least one of the twelve controversial statements, and gender. Participants eligible to participate in the rest of the study from the initial survey were sent an Informed Consent form via Qualtrics, a requirement for our human subject research. This form stated the conditions of the rest of the study after the initial survey, including risks, rights, confidentiality and procedures. A participant's university ID and date of signing were a participant's recognition

that they have read these terms and agree with them. Only those who read this and accepted the agreement would be scheduled to participate further.

In order to be considered to be disagreeing about a statement, one participant had to have rated the statement 1-3 on the scale, meaning they strongly disagreed with it, while the other had to have rated it between 8-10, meaning they strongly agreed with it. Once this disagreement gap between participants was secured, the resulting pairs were placed into one of three attractiveness conditions: mutually Attracted, mutually Unattracted, or Neutral. Those in the Attracted group were to receive a photo from the Chicago Face Database that they had rated as attractive in the initial survey and told this photo was of the person they were communicating with during the chat. Those in the Unattracted group were to receive a photo from the Chicago Face Database that they had rated as unattractive in the initial survey and told the photo was of their conversation partner. Those in the Neutral condition received no photo and no information whatsoever about their conversation partner before the chat. They were given a randomly generated username consisting of a string of letters and numbers to use in the chat to avoid any predispositions during the chat.

The online appointments were set up for a 20-minute chat window. This could take place anywhere from a week to several weeks after taking the initial survey. On the day of the appointment, a couple of hours before the agreed time, the participants were emailed the information about the chat, including the photo and a pseudonym to use with their partner depending on gender ('Sophia' or 'Anna' for photos of female faces and 'Chris' for photos of male faces),[7] and the link to the one-time use chat room in the secure chat interface Discord (https://discordapp.com/). They were instructed to enter the chatroom at the time given and to keep the photo of their putative conversational partner visible on their desktop during the chat. Once both participants entered the chatroom, they were instructed by the second author, who was also present in the chatroom during the first few minutes of the chat, to chat for 20 minutes about the specific topic about which they (unknowingly) disagreed. This was also the participants' chance to ask any questions they may have had about the task. After that, the experimenter left the chatroom open in case of any questions from participants, but did not directly watch or participate, and the chat began upon the first message being sent. Once 20 minutes were up, the experimenter returned to thank the participants and give them a link to fill in the post-survey.

4.4 Post survey

The third and final stage of our study was the post-survey. This final stage aimed to find out whether the opinion of the participants on the selected topic changed as a result of the chat they just had, while also obtaining information about the perceptions they had about the chat and about their chat partner. For this second survey, the participants were asked to rate the same twelve topics as in the screening survey, including the one they just talked about. Then, they were asked general questions about rudeness online, such as:[8]

(a) From the websites you mentioned visiting regularly in the first survey, do you often see people being rude or aggressive towards each other in these platforms?

[7] Pseudonyms were provided to allow subjects to address each other while protecting their privacy and to increase naturalness. At the same time, to avoid them generating their own implicit biases separately from the facial attractiveness condition we controlled for, they were selected to be as race and ethnicity neutral as possible.

[8] For the full text of the post-survey, see Appendix 2.

(b) Can you think of any examples of how people are rude or aggressive in online platforms?
(c) Do you feel people are more rude online versus face-to-face?

Participants were also asked specific questions about the chat they had just had:

(a) How did you overall feel about the person you spoke with today?
(b) Would you speak with this person again if given the chance? Why or why not?
(c) Were there any instances during this chat when you felt offended, felt they were being impolite, or just generally rude towards you? If so, examples?
(d) Were there any instances during this chat when you felt you were being rude or impolite towards your partner? Examples?

The results of this survey were used to gather quantitative results about opinion convergence following the chat, as well as to find out whether participants' perceptions about the chat itself and their conversational partner in terms of impoliteness/rudeness correlated with facial attractiveness and with whether or not their opinion had changed following the chat.

5. Results

The two research questions we set out to answer in this study were:

RQ1) When interacting with someone you find attractive, to what extent does that influence whether you end up agreeing with them?
RQ2) How is opinion convergence linguistically negotiated online and is it negotiated differently depending on whether you are interacting with someone you find attractive or not?

5.1 Degree, frequency and type of opinion convergence as a function of facial attractiveness
In response to the first of these questions, a comparison of ratings of controversial statements from stage I (the screening survey) and stage III (the post-survey) in the three attractiveness conditions reveals some degree of convergence in all conditions (see Fig. 3).

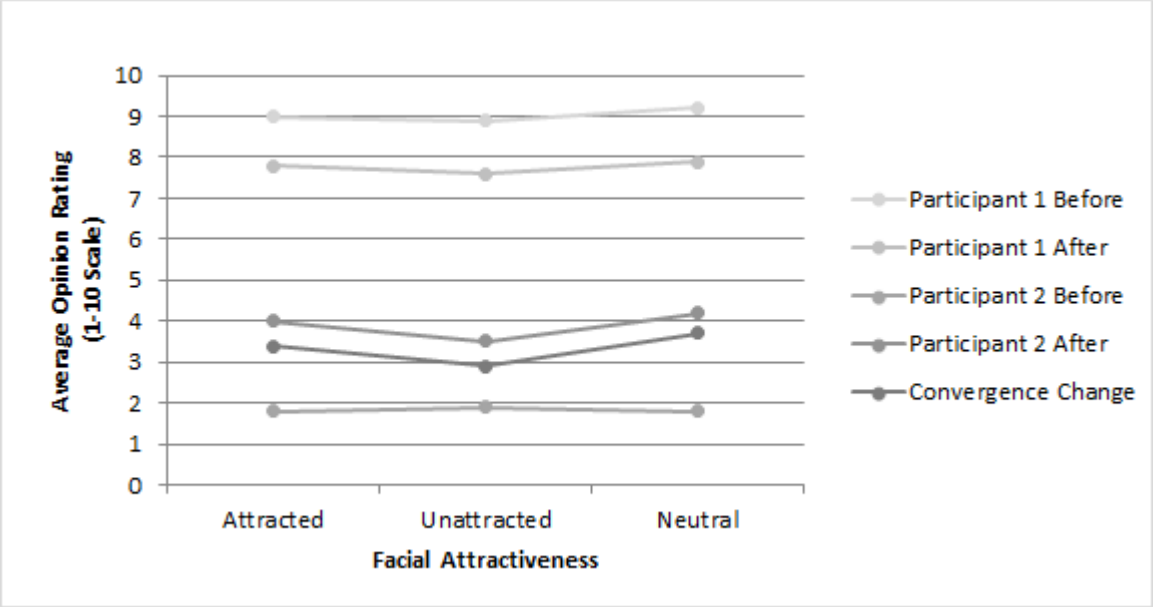

Fig. 3: Degree of opinion convergence as a function of facial attractiveness

	Attracted	Unattracted	Neutral
Participant 1 Before	9	8.9	9.2
Participant 1 After	7.8	7.6	7.9
Participant 2 Before	1.8	1.9	1.8
Participant 2 After	4	3.5	4.2
Convergence Change	3.4	2.9	3.7

Table 1: Average participant ratings and change in convergence

In other words, overall, all of our disagreeing pairs ended up converging in their opinions somewhat, irrespective of condition. However, in terms of magnitude, opinion convergence was greatest among Neutral pairs, who converged on average by 3.7 points, followed by Attracted pairs, who converged by 3.4 points, and finally Unattracted pairs, who converged by 2.9 points. This finding provides partial support for our hypothesis that opinion convergence is modulated by facial attractiveness, as one source of implicit biases that can affect opinion convergence even in the absence of physical co-presence.[9] We discuss limitations in our sampling that may have affected this outcome in section 6 below.

Another measure of opinion convergence that is possible to assess based on our data is the frequency and type of opinion convergence that took place, that is, how frequently opinion convergence occurred in each condition and whether this was one-sided, with just one party changing their opinion to approximate that of their conversational partner, or mutual, with both parties exhibiting this type of behaviour. In terms of frequency, opinion

[9] Since this was an exploratory study on a small sample size, we use descriptive statistical measures to describe our findings. These show that these elements are related in the sample studied but do not allow us to establish statistical significance for these findings beyond this sample. Future studies with larger sample sizes, formal control groups and data from previous studies or an explicit control group for comparison can yield results for which statistical significance can be reported.

convergence was most frequent in the Attracted condition, with 9 out of 10 conversations converging as a result of the chat, while in one conversation in this condition opinions actually diverged. Opinion convergence was only marginally less frequent in the Unattracted condition, where 8 out of 10 conversations converged as a result of the chat, while for the remaining two conversations, in one instance there was divergence and in the other no change in opinion was noted (the pre- and post-survey ratings were identical). Of the three conditions, opinion convergence was least frequent in the Neutral condition, where only 7 out of 10 conversations converged, while in the remaining three instances there was no opinion change.

More telling is the direction of convergence in each of the three conditions. In the Attracted condition, in 6 out of 9 converging conversations, convergence was one-sided, with males changing to approximate the opinion of the female twice as often as females (4 males changed their opinion, while only 2 females did). This pattern was repeated in the Unattracted condition, where in 5 out of 8 converging conversations, convergence was again one-sided with males taking the lead (3 males changed opinion, as opposed to 2 females). The pattern was reversed in the neutral condition, where opinion convergence was mutual in the majority of cases where it occurred (5 out of 7), while in the remaining two conversations where convergence was one-sided, the gender balance was equally split (1 male, 1 female). These results again provide preliminary support for our hypothesis that facial attractiveness can encourage opinion convergence even in the absence of physical co-presence. They furthermore suggest that facial attractiveness is a stronger source of implicit bias for males than it is for females, as in 8 out of 13 cases where opinion convergence was one-sided, it was the male that changed their opinion to approximate that of the female. This result is in line with prior research, which has shown that a "beauty premium" exists with more salient effects on males, recently confirmed by fMRI and ERP results (Cloutier et al. 2008, Ma et al. 2017).

5.2 The linguistic negotiation of opinion convergence in different attractiveness conditions

To answer our second research question, pertaining to the linguistic negotiation of opinion convergence under different attractiveness conditions, the 30 conversations (10 in each condition) obtained during the second stage (online chat) of our study were annotated by two independent annotators, who coded for verbal markers and other devices of dis/agreement using a specially devised coding scheme. This scheme was intended to encompass as wide as possible a range of markers of agreement or disagreement realized through different formal and pragmatic means and did not distinguish between agreement and disagreement in this regard. This is because our goal was to provide a qualitative overview of how opinion convergence was negotiated overall in our data and whether there was any difference between the three attractiveness conditions in this regard, rather than to track potential differences between how agreement and disagreement were linguistically signalled in these data.

To this end, markers of dis/agreement were broadly distinguished into form-based and function-based ones. Under the former, we categorized instances of lexical agreement (*yeah, yes, of course, true, I agree, exactly, good idea, good point*), agreement-seeking tag questions (*If that makes sense?, You know?, Perhaps?, Right?*), elaborating phrases (*I think that…, For instance…, I mean…, For example…*), evaluative terms indicating high emotional involvement (*ridiculous, upset, annoying, unfair, sooo stupid*), CMC cues (abbreviations such as *lol, haha,* emojis such as :D, :o, and hashtags such as *#budgetcutsatuiuc*), lexical hedges (*Sort of, I guess, Yeah but, However*), thanks (*Thanks, Thanks for chatting*), and

conversational openings and closings (*Nice to meet you, Nice talking to you, Good chatting with you on this topic*), as well as the use of first names (that is, the pseudonyms provided to participants for this purpose: *Sophia, Anna, Chris*). Function-based markers of dis/agreement included expressions of empathy/understanding (*I understand where you're coming from, that's a good point, I hear you out, That makes sense*), expressing agreement on an unrelated topic (*That's something we both can agree on, ….*), requests to elaborate (*What makes you say that? Do you mean…?*), asserting lack of opinion or knowledge about a topic (*I had no idea, I assume, I'm not an economist, I am not sure about…, I'm not familiar with…*), meta-comments about the conversation (*We've gotten far away from the topic*), rhetorical questions (*Why not set the minimum wage to $100 an hour?*), direct statement of the opposite opinion (*High school isn't free, Well I wouldn't agree to that*), disagreeing by omission or changing the topic (*Let's talk numbers*), challenging the veracity of the other's statement (*Do they not exist today?*), responding to a challenge (*I firmly believe so, On the contrary though…*), asserting humorous intent (*I'm kidding*) and deferring discussion (*That's a topic for another day/time*). Finally, we also estimated turn length by establishing the difference between the longest turn (in syllables) in the first half of the conversation and the longest turn (in syllables) in the second half of the conversation, and averaging the differences of all conversations in each group. We normed these three averages with the lowest value across the three groups, thus a turn length ratio of one means the least difference while values higher than one have a greater difference in turn length over the conversation in comparison to the others. Our hypothesis in this regard was that shorter turns in the second half of the conversation may indicate a quicker pace, which would be indicative of a rising, conflicting tone.

Our results showed that, overall, lexical agreement, lexical hedges, evaluative terms indicating emotional involvement, use of first names, expressions of empathy and understanding, and requests to elaborate as well as asserting lack of opinion or knowledge about a topic are preferred means of negotiating opinion convergence among Attracted and Unattracted pairs alike (see Fig. 4 and 5). All of these indicate a wish to maximize agreement and minimize disagreement, as predicted by, e.g., politeness frameworks such as Leech (2014). Despite this qualitative similarity, Attracted pairs used these dis/agreement markers almost twice as much as Unattracted pairs, suggesting that they are much more invested in the kind of relational work that negotiating dis/agreement entails.

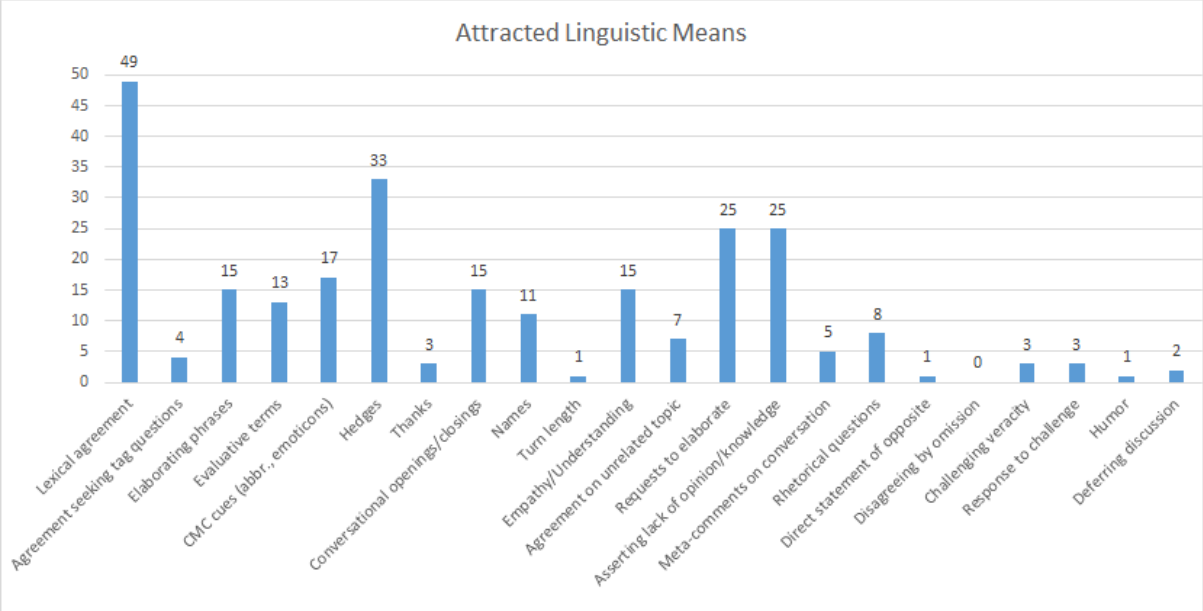

Fig. 4: The linguistic negotiation of opinion convergence between Attracted pairs

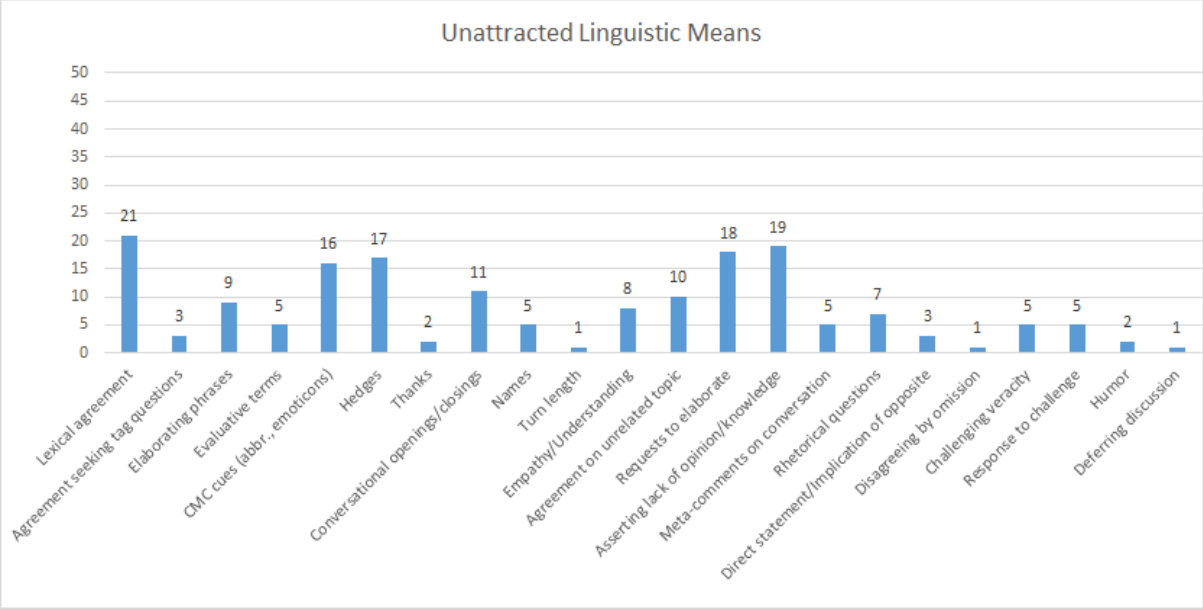

Fig. 5: The linguistic negotiation of opinion convergence between Unattracted pairs

Conversations between Neutral pairs look somewhat different in this regard. While lexical markers of agreement are still widely used, CMC cues as well as emotionally involved words are now also quite frequent (see Fig. 6).

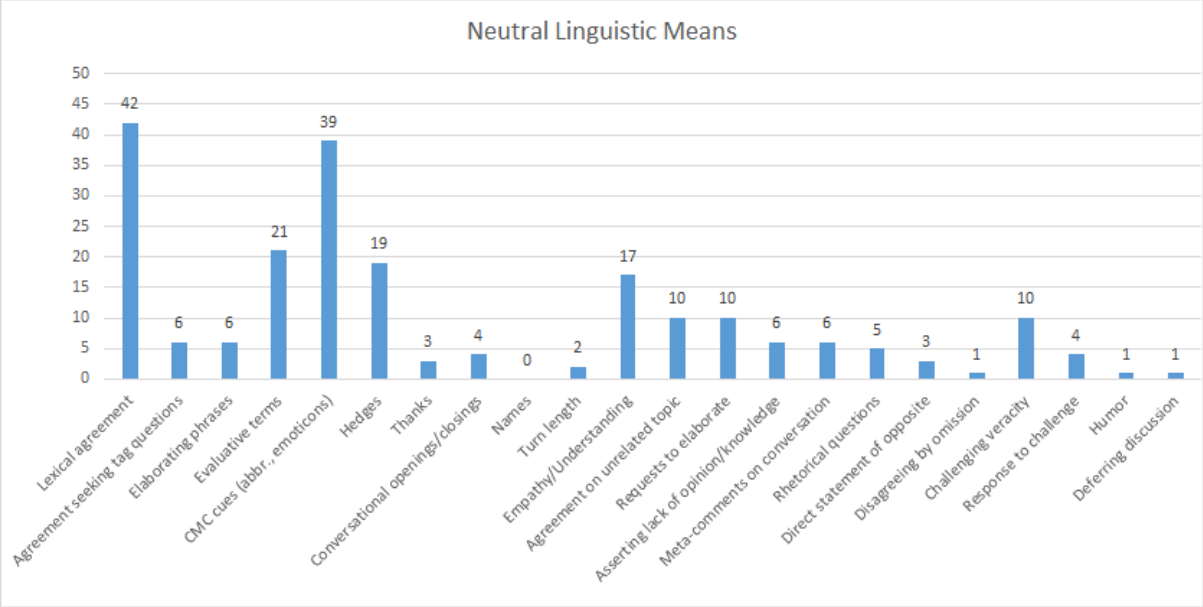

Fig. 6: The linguistic negotiation of opinion convergence between Neutral pairs

However, these results should be taken with a sizable grain of salt, as it is possible that they reflect limitations in our sampling rather than real differences between the three conditions. We briefly discuss these limitations below.

6. Limitations

Unlike the online world of the internet and especially platforms like Reddit cited at the outset, where people from all walks of life meet and interact, our study was conducted among a sample of University undergraduates in a University setting, at a University campus where diversity issues and respect for different viewpoints had been the topic of a recent campus-wide campaign, and among participants who knew the other participants were also young students. These ecological considerations may well have limited our findings in a number of ways. Most importantly, they may be behind the fact that opinion convergence was observed in the majority of conversations in all three attractiveness conditions investigated. This high frequency of opinion convergence is quite unlike that found in the Lipinski-Harten & Tafarodi (2013) study which inspired our own, in which opinion convergence in the OC condition was half of that found in the FTF condition (although their study was conducted with University undergraduates also). Based on their findings, we would have expected opinion convergence in our chat data to occur less frequently overall. However, that is not what we found.

This discrepancy between our experimental findings on the one hand and previous research as well as people's general experience of interacting online on the other, is supported by participants' comments in the post-survey administered after the chat. With regard to general rudeness online, participants were overall very aware of the frequency with which people argue online and the rudeness that goes along with it in their opinion. Most participants said they have witnessed rudeness on online platforms they use, examples ranging from opinions on political and social issues, criticism and negativity, to more complex and deeply rooted aggression, such as racism. They also overwhelmingly agreed that people are more rude online compared with face-to-face (59 out of 60 participants either agreed with this statement or stated that this is sometimes the case). With regard to the

particular chat they had, however, most participants said that their conversation was pleasant, with 51 out of 60 stating that they would speak to the other person again, if given the chance. Only 2 out of 60 participants said they felt the other was rude, because s/he disrespected their opinion or tried to make them feel stupid. However, 5 out of 60 felt they were being rude themselves (none of these five corresponded directly to the two instances where they felt the other person was rude). These comments suggest that our survey may have created a sanitized environment which artificially encouraged participants to be on their 'best behaviour', contrary to the realities of (unmoderated) communication online. Seen from this perspective, that facial attractiveness was shown to have an effect at all under these circumstances of potential self-censorship is noteworthy – yet precisely what we would expect of an implicit bias – and promising for future explorations of this notion.

Moreover, we should note a gender imbalance between conditions which may have biased the degree and type of convergence we observed. Specifically, based on their responses about who they find attractive and their own gender identity in the screening survey, 6 out of 10 conversations in the Neutral condition were between females, whereas only 2 out of 10 conversations in the Unattracted condition and none in the Attracted condition were between females. This over-representation of females in the Neutral condition may be responsible for the greater degree of convergence observed in this condition (3.7 points), which, contrary to prediction, exceeded that observed in the Attracted condition (3.4 points). The same gender imbalance may also underlie the higher use of emotionally-loaded words and CMC cues in that condition, consonant with recent research on Twitter, which has found that women are more likely to use emotion terms and emoticons as well as CMC words (*LOL, omg*) and hesitation markers on that medium (Banman et al. 2014).

7. Conclusions

Rudeness and aggressive behaviour are widely acknowledged to be a problem in online communication and the anonymity that results from lack of physical co-presence is often blamed for that. In this study, we attempted to disentangle the effects of lack of physical co-presence from those of implicit biases that can be triggered also in the absence of physical co-presence and can be a powerful factor in opinion convergence that can help prevent conflict online.

Taking our lead from a study by Lipinski-Harten and Tafarodi (2013), who found that online communication was about half as conducive to opinion convergence between disagreeing parties as face to face interaction, we designed a study in which participants interacted online in three different conditions. In the Attracted condition, participants chatted for 20 minutes about a topic they held opposite opinions about with someone who they were told looked like an individual they had rated as attractive in a prior screening survey. In the Unattracted condition, they did the same but this time they were led to believe their interlocutor looked like someone they had rated as unattractive in the screening survey. Finally, in the Neutral condition, participants were not given any information, visual or otherwise, about their interlocutor. Our goal was to gauge the effects of facial attractiveness on the magnitude, frequency and direction of opinion convergence online under these different conditions.

Our results showed that facial attractiveness, as one source of implicit (positive) bias toward our fellow communicators, indeed makes a difference, inasmuch as it produces opinion convergence of a greater magnitude and with increased frequency when interactants find each other attractive compared to when they do not. It also tends to predispose

participants to heightened engagement in relational work through the use of a variety of devices to maximize agreement and minimize disagreement. These results are not without their limitations, however. Given our participant pool was limited to University undergraduates and the gender imbalance noted in the Neutral condition, future work with more diverse and more balanced samples of participants is needed before these findings can be generalized to a wider population of internet users.

Nevertheless, we think these results are important for a couple of reasons. By focusing on facial attractiveness, we were able to implement anonymity as a matter of degree, which is a more realistic representation of online communication that often contains partial clues to others' identity. First names are often used in online communication and these, like pseudonyms, can be telling of users' social identities (gender, age, ethnic background; e.g., Liu & Ruths 2013). Similarly, dialectal words and turns of phrase, and even lexico-semantic preferences can betray someone's social or regional background (Banman et al. 2014, Rustagi et al. 2009), equally triggering implicit biases. Stereotypes about all of these can trigger positive or negative assessments of others online as much as they do offline, such that lack of physical co-presence leading to reduced engagement with the subjectivity of the other is only one among several factors affecting opinion convergence online. Thus, unlike previous studies which have tended to explain the heightened frequency of face-threatening behaviours online by highlighting aspects of communication unique to online environments (anonymity, invisibility, lack of engagement with subjectivity of other), we have been able to show that implicit biases (which operate offline as well as online) are an additional factor influencing such behaviour. Although such biases can be triggered in a multitude of ways by information provided online, these are currently under-explored by focusing on anonymity as an over-arching determinant of online communication. Considering a lot of our online communication happens with people we know offline, or at least, know something about, it is important that we also start to pay attention to aspects of communication that are shared between the offline and online modes. As our understanding of online interactions is becoming increasingly sophisticated, it is time to re-examine not just their differences but also how the two modes of communication are similar to each other. Only through a more realistic portraying of online discourse can we hope to understand both its limitations as well as the possibilities it offers.

Acknowledgements

The research reported in this article was carried out with support by an SLCL Undergraduate Research award to the first author from the University of Illinois. We thank Chris Fraley for discussion of aspects of our experimental design relating to facial attractiveness.

References

Bamman, David, Jacob Eisenstein and Tyler Schnoebelen (2014) "Gender identity and lexical variation in social media." *Journal of Sociolinguistics* 18(2): 135–160.

Brownstein, Michael (2017) "Implicit bias." In *The Stanford Encyclopedia of Philosophy,* ed. by Edward N. Zalta. Available online at: https://plato.stanford.edu/archives/spr2017/entries/implicit-bias/

Cloutier J, Heatherton TF, Whalen PJ, Kelley WM. (2008) "Are attractive people rewarding? Sex differences in the neural substrates of facial attractiveness." *Journal of Cognitive Neuroscience* 20(6): 941–951.

Coetzee V, Greeff JM, Stephen ID, Perrett DI (2014) “Cross-cultural agreement in facial attractiveness preferences: The role of ethnicity and gender.” *PLoS ONE* 9(7): e99629. Available online at: https://doi.org/10.1371/journal.pone.0099629

Jang, H., Song, J., & Kim, R. (2014). “Does the offline bully-victimization influence cyberbullying behavior among youths? Application of general strain theory.” *Computers in Human Behavior* 31: 85–93.

Kwan, Samantha, and Mary Nell Trautner (2009) “Beauty work: Individual and institutional rewards, the reproduction of gender, and questions of agency.” *Sociology Compass* 3(1): 49–71.

Langlois JH, Kalakanis L, Rubenstein AJ, Larson A, Hallam M, et al. (2000) “Maxims or myths of beauty? A meta-analytic and theoretical review.” *Psychological Bulletin* 126: 390–423.

Leech, Geoffrey (1983) *Principles of Pragmatics*. London: Longman.

Leech, Geoffrey (2014) *The Pragmatics of Politeness*. Oxford: Oxford University Press.

Lipinski-Harten, Maciek & Tafarodi, Romin (2013) “Attitude moderation: A comparison of online chat and Face-to-face conversation.” *Computers in Human Behavior* 29, 2490 –2493.

Liu, Wendy and Derek Ruths (2013) “What’s in a name? Using first names as features for gender inference in Twitter.” *Analyzing Microtext: Papers from the 2013 AAAI Spring Symposium*. Available online at: http://www.ruthsresearch.org/static/publication_files/LiuRuths_AM2013.pdf

Lorenzo-Dus, Nuria, Blitvich, Pilar and Patricia Bou-Franch (2011) “On-line polylogues and impoliteness: the case of postings sent in response to the Obama Reggaeton YouTube video.” *Journal of Pragmatics* 43: 2578-2593.

Ma, Qingguo, Da Qian, Linfeng Hu, and Lei Wang. 2017. “Hello handsome! Male’s facial attractiveness gives rise to female’s fairness bias in Ultimatum Game scenarios – An ERP study.” PLoS ONE 12(7): e0180459. https://doi.org/10.1371/journal. pone.0180459

Misoch, Sabina (2015) “Stranger on the internet: Online self-disclosure and the role of visual anonymity.” *Computers in Human Behavior* 48: 535–541.

Moor, Peter, Heuvelman, Ard and Ria Verleur (2010) “Flaming on YouTube.” *Computers in Human Behavior* 26: 1536-1546.

Park, S., Na, E. Y., and Kim, E. M. (2014) “The relationship between online activities, netiquette and cyberbullying.” *Children and Youth Services Review* 42: 74-81.

Reicher, Steve, Spears, Russell and Tom Postmes (1995) “A social identity model of deindividuation phenomena.” *European Review of Social Psychology* 6: 161-198.

Rishab Nithyanand, Brian Schaffner, and Phillipa Gill (2017) “Measuring offensive speech in online political discourse.” In *USENIX Workshop on Free and Open Communications on the Internet (FOCI)*. Vancouver, Canada. August 2017.

Rustagi, Mayur, Prasath, R. Rajendra, Goswami, Sumit, Sarkar, Sudeshna, Chaudhury, Santanu, Mitra, Sushmita, Murthy, C. A., Sastry, P. S. and Sankar K. Pal (2009) “Learning age and gender of blogger from stylistic variation.” In *Pattern Recognition and Machine Intelligence: Third International Conference, PReMI 2009*, ed. by Chaudhury, Santanu, Mitra, Sushmita, Murthy, C. A., Sastry, P. S. and Sankar K. Pal. Lecture Notes in Computer Science 5909, 205–212. New York: Springer.

Satz, Samantha, David L. Pogge, and Philip D. Harvey (2017) “Cyberbullying and adolescent mental health: A study of adolescents on an acute inpatient psychiatric unit.”

Presented at: the *American Psychiatric Association Annual Meeting*, May 23, 2017, San Diego, CA.
Scott, Craig R. (2004) "Benefits and drawbacks of anonymous online communication: Legal challenges and communicative recommendations." In *Free Speech Yearbook*, Vol. 41, ed. by S. Drucker, 127–141. Washington, DC: National Communication Association.
Stein, Joel (2016) "How trolls are ruining the internet." TIME, August 29, 2016.
Suler, John (2004) "The online disinhibition effect." *Cyberpsychology and Behavior* 7(3): 321–326.
Truong, Dan (2015) "The great equalizer 2.0". Available online at: http://www.huffingtonpost.com/dan-truong/the-great-equalizer-20_b_7157662.html
Wilcox, Keith & Stephen, Andrew (2013) "Are close friends the enemy? Online social networks, self-esteem, and self-control." *Journal of Consumer Research* 40: 90-103.

Appendix I: Screening Survey

Fields with an asterisk (*) were required by participants. Those with no options were an open-ended text box for individual user input.

Page 1

The first page contained an agreement for participants to e-sign verifying they understood the terms of the participation, compensation, and confidentiality, including a statement of informed consent required by the Institutional Review Board from the University of Illinois. They had to agree to all terms before completing the survey. The agreement for this particular form was purely for collection of this first survey. Participants selected for participation also completed a full confidentiality agreement for the rest of the study, as previously described.

Page 2

1. * Netid
2. * E-mail
3. * What is your current affiliation with the University of Illinois?
   a. *I am an undergraduate student*
   b. *I am a graduate student*
   c. *I am an exchange student on study abroad*
   d. *I am a faculty/staff member*
   e. *I am not affiliated with the University of Illinois*
4. * What is your major?
5. * In what country were you born?
6. At what age did you come to the U.S.?
7. * How would you racially or ethnically define yourself? (Ex. African-American, Filipino, Mixed (Caucasian/Latina), etc.)
8. * I personally identify as...
   a. *Male*
   b. *Female*

c. *Not listed - please specify (open ended text box)*

9. * I personally am attracted to...
   a. *Men*
   b. *Women*
   c. *Men and Women*
   d. *Not listed- please specify (open ended text box)*
10. Please signify, in hours, the amount of time in a day you spend online doing the following activities: (open ended number box for each)
    a. *Social Networking (Facebook, Twitter, etc.)*
    b. *Video Games (Steam, Xbox, etc.)*
    c. *Sending/Checking E-mail*
    d. *Chatting/IMing*
    e. *Blogging (LiveJournal, Tumblr, etc.)*
    f. *Commenting or reading comments (Youtube, articles, etc.)*

11. * Please rate how much you agree with the following statements on a scale of 1-10, with 10 being strongly agree and 1 being strongly disagree, 5 being neutral/no opinion.
    a. *Higher education should be provided free of cost for those who choose to pursue it.*
    b. *The minimum wage should be raised.*
    c. *The amount of time people spend on technology is problematic.*
    d. *Violent imagery in games and media create violence in children.*
    e. *Vaccination should be required of all children.*
    f. *English should be the official language of the United States.*
    g. *All United States public schools should have a uniform policy.*
    h. *Welfare recipients should have to undergo mandatory drug testing.*
    i. *Smoking should be banned in all public places.*
    j. *Animals should not be used in the testing of drugs and cosmetics.*
    k. *Sex education (non-abstinence based) should be mandatory in U.S. schools.*
    l. *Medical marijuana should be legalized in Illinois.*
12. * What is your favorite kind of food/cuisine?
13. * Are you interested in sports? If so, which?
14. * What is your favorite music genre?

Page 3

Below you will see some photos of different people. For each photo, please indicate whether you think the person is attractive or unattractive. Don't spend too much time thinking about it, just go with your initial thoughts about the person. Also note that this does not have to be purely sexual attractiveness either. In the event that maybe you do not personally experience sexual attraction, you can think of this as a perception of likability - if you saw this person on the street, how likely is it you would want to talk to them?

At this point, the participant sees photos from the Chicago face database. If on question 9 they said 'Men', they would only see 50 randomly selected photos of young men of various ethnicities from the Chicago face database. If on question 9 they said 'Women', they would only see 50 randomly selected photos of young women of various ethnicities from the Chicago face database. If on question 9 they said 'Men and Women' or 'Not listed', they saw all 100 randomly selected photos of young men and women from various ethnicities from the Chicago face database. They could then drag these photos to either a box labeled 'Attractive' or 'Unattractive' and submit the survey answers. Only participants who completed every required question up until this last page were considered for participation.

Appendix II: Post-Survey

All initial questions on this survey were required. Those with no options were an open-ended text box for individual user input.

1. Please rate how much you agree with the following statements on a scale from 1-10, 10 being strongly agree and 1 being strongly disagree, 5 being neutral/no opinion.
    a. *Higher education should be provided free of cost for those who choose to pursue it.*
    b. *The minimum wage should be raised.*
    c. *The amount of time people spend on technology is problematic.*
    d. *Violent imagery in games and media create violence in children.*
    e. *Vaccination should be required of all children.*
    f. *English should be the official language of the United States.*
    g. *All United States public schools should have a uniform policy.*
    h. *Welfare recipients should have to undergo mandatory drug testing.*
    i. *Smoking should be banned in all public places.*
    j. *Animals should not be used in the testing of drugs and cosmetics.*
    k. *Sex education (non-abstinence based) should be mandatory in U.S. schools.*
    l. *Medical marijuana should be legalized in Illinois.*
2. How do you feel overall about the person you spoke with today?
3. Would you speak with this person again if given the chance?
    a. if *Yes*: Why? *(open-ended text box)*
    b. if *No*: Why not? *(open-ended text box)*
4. Were there any instances during this chat where you felt offended, felt they were being impolite, or just generally rude towards you?
    a. If *Yes*: Can you think of any examples of their rudeness? Why do you think they acted in this way? *(open-ended text box)*
    b. *No*
5. Were there any instances during this chat where you felt you were being rude or impolite towards your partner?
    a. If *Yes*: Can you think of any examples of your rudeness? Why do you think you acted in this way? *(open-ended text box)*
    b. No
6. In the websites you mentioned using often on the first survey, do you often see people being rude or aggressive towards each other in these platforms?
    a. If *Yes*: Can you think of any examples of how people are rude or aggressive in online platforms? *(open-ended text box)*
    b. No
7. Do you think people are more rude online versus face-to-face?
    a. Yes
    b. Sometimes
    c. No
8. Do you feel your experience today contradicts or agrees with what you thought in the previous question? Clarify.
9. When do you feel people feel the need to be aggressive and why?

At this point the participants filled out their personal information for compensation purposes.