Sympathy, Empathy, and Twitter

Herzog, Lisa

Published in:
Proceedings of the Aristotelian Society

DOI:
10.1093/arisoc/aoad004

Publication date:
2023

Citation for published version (APA):
Herzog, L. (2023). Sympathy, Empathy, and Twitter: Reflections on Social Media Inspired by an Eighteenth-Century Debate. *Proceedings of the Aristotelian Society*, *123*(1), 51-72. https://doi.org/10.1093/arisoc/aoad004

Meeting of the Aristotelian Society, held at Senate House, University of London, on 14 November 2022 at 6:00 p.m.

III—Sympathy, Empathy, and Twitter: Reflections on Social Media Inspired by an Eighteenth-Century Debate

Lisa Herzog

How can the harm caused by waves of fake news or derogatory speech on social media be minimized without unduly limiting freedom of expression? I draw on an eighteenth-century debate for thinking about this problem: Hume and Smith present two different models of the transmission of emotions and ideas. Empathetic processes are causal, almost automatic processes; sympathy, in contrast, means putting oneself into the other person's position and critically evaluating how one *should* react. I use this distinction to argue that the architectural logic of social media should be improved to prevent cumulative harms and to facilitate sympathetic processes.

I

Introduction. Social media have massively increased the amount and speed of human communication. Alongside many positive effects, this has also allowed the spread of falsehoods and negative emotions, sometimes in veritable 'hate waves' that can also have repercussions in the offline world, such as outbreaks of violence (Robb 2017). But one regularly finds reports from individuals who decided to contact their 'haters' and found that in one-to-one interactions, these individuals were respectful and held much more nuanced positions (Pearl 2011). There seems to be something specific about social media that makes it particularly conducive to negative outbursts.[1]

In this paper, I suggest shifting attention away from online *content*, towards *mechanisms of social media communication*, to address this problem. I take my cue from what may appear an unexpected

[1] However, some studies also find that the same individuals are aggressive in both online and offline communication (though it is more visible online): see, for example, Sest and March (2017), Bor and Betersen (2022). I come back to that point in note 10 below.

Proceedings of the Aristotelian Society, Vol. 123, *Part* 1
https://doi.org/10.1093/arisoc/aoad004

source: an eighteenth-century debate about sympathy and empathy between David Hume and Adam Smith. Hume and Smith reflected on the human ability to share ideas and emotions with others but conceptualized these phenomena differently. This debate, and some of the scholarly discussion about it (Darwall 1998; Sugden 2002; Rick 2007; Fleischacker 2012; Sayre-McCord 2013; Khalil 2015; Sagar 2017; McHugh 2018), can provide inspiration for thinking about the ways in which ideas and emotions spread online.

The core difference between their accounts is this. Hume uses a model of contagion: when seeing another person's emotions, the spectator is infected by them as well. For Smith, in contrast, sympathy, 'our fellow-feeling with any passion whatever' (Smith [1776] 1976, I.I.I.5 cited hereafter as *TMS*), arises because we put ourselves into the situation of the other person and observe our own reaction. We may initially react in the same way, but we can also correct our reaction, and conclude that the other person reacted to the situation in an inappropriate way. It is this critical wedge between actor and spectator that enables Smith to develop his moral philosophy, which leads to the figure of the 'impartial spectator' (*TMS*, especially Part III). For the sake of simplicity (sacrificing historical terminological accuracy), I will refer to the Humean account as empathy, and the Smithian one as sympathy.[2]

Hume and Smith could not of course have anticipated the way social media work. But their nuanced observations of human interactions are of enduring interest. If my argument is correct, they are also relevant for thinking about the phenomenology of social media. The assumption that underlies my discussion is that certain features of human psychology and sociability are relatively constant over time, and that different media of communication can lead to different interpersonal processes. On social media, certain human propensities, for example, the joy of sharing emotions with others, can misfire, such as when it makes us unthinkingly share a piece of fake news. In response, we should ask how the architecture of online communication can be better designed.

In the next section (§II) I will present Smith's and Hume's accounts in more detail. I then discuss the applicability of these accounts to social media (§III). I argue that certain architectural design decisions

[2] I will not discuss the broader role of sympathy or empathy for morality; see, for example, Darwall (1998, pp. 271–9), or, more critically, Bloom (2016).

Proceedings of the Aristotelian Society, Vol. 123, *Part* 1
https://doi.org/10.1093/arisoc/aoad004

can make it more or less likely that users will operate in the mode of either 'sympathy' or 'empathy'. This suggestion is in line with recent proposals to design the architecture of social media in ways that reduce the spreading of misinformation or hate messages (for instance, Lorenz-Spree et al. 2020). This perspective avoids questions about content moderation—and thus difficult questions about censorship, free speech, and so on—and focuses, instead, on the interplay between media of communication and human propensities.

II

Smith and Hume on Sympathy and Empathy. In the Scottish Enlightenment, discussions about the nature of morality were closely tied to psychological observations. Smith and Hume aimed at explaining the emotions and behaviours we describe as 'moral' as growing out of more basic psychological mechanisms. Both were keen observers of human psychology, working with various vignettes and examples. It is these psychological accounts (rather than their normative theories) that I draw on. As already mentioned, I use 'sympathy' for Smith's model and 'empathy' for Hume's. Historically, both authors used the term 'sympathy' for 'any case in which one person participates in another's feelings' (Fleischacker 2012, p. 273), but they describe different mechanisms underlying this phenomenon.

Hume introduces 'sympathy' as 'that propensity we have to sympathize with others, and to receive by communication their inclinations and sentiments, however different from, or even contrary to our own' (Hume [1739–40] 1978, II.I.XI, p. 316; cited hereafter as *Treatise*).[3] He adds that in addition to emotions, 'opinions' are also transmitted: even 'men of the greatest judgment and understanding' are prone to adopt the opinions of their 'friends and daily companions' (*Treatise* II.I.XI, p. 316). Hume builds on his distinction between 'ideas' and 'impressions' to explain the mechanism behind this sharing (in what appears to be an explanation for the transmission of 'opinions' and 'passions' alike). We pick up the 'affections' of others by 'external signs in the countenance of conversation' that 'convey an idea of it'; this idea, however, quickly turns into an

[3] I focus here on the passage in the *Treatise*; on differences in the *Enquiry* (Hume [1751] 1975), see, for example, Van Holthoon (1993, pp. 36–42) and Morrow (1923, pp. 66–7).

https://doi.org/10.1093/arisoc/aoad004

impression,[4] 'and acquires such a degree of force and vivacity, as to become the very passion itself, and produce an equal emotion, as any original affection' (*Treatise* II.I.XI, p. 317; see also pp. 319–20; for discussion, see McHugh 2018, pp. 685–6). This mechanism is 'instantaneous' and hardly perceptible to the person herself, though an observing philosopher can distinguish its different steps (*Treatise* II.I.XI, p. 317).

Hume also speaks of the 'minds of men' as 'mirrors to one another, not only because they reflect each others [sic] emotions, but also because those rays of passions, sentiments and opinions may be often reverberated, and may decay away by insensible degrees' (*Treatise* II.II.V, p. 365). Thus the immediate communication seems to preserve the exact same thing (or something very close to it). Moreover, the metaphor of mirroring suggests passivity; in fact, a mirror cannot prevent mirroring, so a literal reading of the metaphor suggests that humans *cannot but* share the 'passions, sentiments and opinions' of others. The metaphor of resonating 'strings equally wound up' (*Treatise* III.III.I, p. 575) is, in this respect, similar. Hume also uses verbs such as 'infusing' (*Treatise* II.I.XI, p. 317), again implying susceptibility vis-à-vis external affections. Another quote makes this explicit:

> So close and intimate is the correspondence of human souls, that no sooner any person approaches me, than he diffuses on me all his opinions, and draws along my judgment in a greater or less degree. And though, on many occasions, my sympathy with him goes not so far as entirely to change my sentiments and way of thinking; yet it seldom is so weak as not to disturb the easy course of my thought, and give an authority to that opinion, which is recommended to me by his assent and approbation. (Hume, *Treatise* III.III.II, p. 592)

Several features of the Humean account stand out. First, it is a *causal* process by which a sentiment, opinion or passion is translated into the same or a similar sentiment, opinion or passion in another person (see also Rick 2007, p. 137). Second, the process seems *automatic*, with a strong likelihood that an individual will take over a sentiment, opinion or passion that is of almost the same intensity as the original. Third, human beings are *passive* in this process.

[4] Impressions, for Hume, are sense impressions, emotional states or desires, while ideas are 'the faint images of these in thinking and reasoning' (*Treatise*, I.I.I.1).

https://doi.org/10.1093/arisoc/aoad004

Commentators disagree about the appropriateness of the metaphor of 'contagion' for this process, but agree that it happens largely on a non-cognitive, sub-conscious level (for discussion, see Fleischacker 2012, pp. 290–1).[5]

This is only a brief sketch of Hume's account, but it suffices to see that Smith's account is different. Admittedly, Smith also describes situations in which the process seems very similar to Hume's description. For example, he writes that 'A smiling face is, to everybody who sees it, a cheerful object …' (*TMS* I.I.I.6), which sounds as if a direct transmission of sentiments is taking place. But not all of Smith's vignettes can be explained by Hume's account, and there is scholarly agreement that Smith's and Hume's account differ (Fleischacker 2012, pp. 278–9). For example, Smith notes that 'When we see a stroke aimed, and just ready to fall upon the leg or arm of another person, we naturally shrink and draw back our own leg or our own arm' (*TMS* I.I.I.3). We react instinctively to a danger even if the potential victim is *not* aware of it, because we imagine ourselves in his or her position. So, importantly, something other than a direct transmission of sentiments, passions or opinions must take place. Instead, what here elicits our reaction is that we imagine ourselves in another person's physical situation and react to the threat to which she is exposed.

Smith adduces the case of sympathy with the dead to underline the point that sympathy is not merely a transmission of feelings (*TMS* I.I.I.10–13, II.I.II.4; see also McHugh 2018, p. 685; Sayre-McCord 2013, p. 215). It is the imagination of the other person's situation, and then our own reaction to it, that creates the 'fellow-feeling': 'sympathy … does not arise so much from the view of [another person's] passion, as from that of the situation which excites it' (*TMS* I.I.I.10). Smith even says that in this imaginative process, one 'become[s] in some measure the same person with' the other person (*TMS* I.I.I.2), which is why one can come to feel 'something which, though weaker in degree, is not altogether unlike' the original person's sensation (*TMS* I.I.I.2; see also Darwall 1998, p. 267, who speaks about an 'imagined surrogate'; see also Rick 2007, p. 138).

Crucially, however, sympathy arises out of a *situation* and not a *transmission* of sentiments (or opinions, or passions) as such. This

[5] This raises some challenges for the notion of the 'common point of view' that Hume uses for moral evaluations; for discussion, see Cohon (1997).

Proceedings of the Aristotelian Society, Vol. 123, *Part* 1
https://doi.org/10.1093/arisoc/aoad004

means that there can be a difference between what the first person feels and what the observer feels. Smith holds that we sympathize less strongly if we do not agree with the person's reaction to a situation, noting that

> our sympathy with the grief or joy of another, before we are informed of the cause of either, is always extremely imperfect. … The first question which we ask is, What has befallen you? Till this be answered, though we are uneasy both from the vague idea of his misfortune, and still more from torturing ourselves with conjectures about what it may be, yet our fellow-feeling is not very considerable. (*TMS* 1.1.1.9; see also Darwall 1998, pp. 269–70, Sayre-McCord 2013, p. 215)

This is an important difference to Hume, and it adds a temporal dimension: we may have a first spontaneous reaction but then come to question its appropriateness when we receive more information about the other person's situation. This allows criticizing the first person's reaction—and starting a conversation about how *different* individuals would react to the situation. For Smith, this is the starting point for his normative project, in which he develops the perspective of an 'impartial spectator', whose reactions are not distorted by any partial feelings or one-sided preferences.

Neither Smith nor Hume draws a systematic distinction between the *emotional* and the *cognitive* side of this process. As shown above, Hume treats 'opinions' and 'passions' on a par when it comes to the influence other people have on us. For Smith, the whole process of emotional sympathizing has a stronger cognitive component from the start (see also McHugh 2018, p. 684). We must *imagine* ourselves in another person's situations, and while some of his examples suggest that this at first happens almost automatically, we then need to grasp the situation in more detail and think about how we would react to it. We can also extrapolate from Smith's writings what a version of sympathy that refers purely to opinions would look like: if someone tells us their opinion, we imagine ourselves in their situation, with the same evidence available to us. As with emotions, we might also arrive at a different judgement: we might, for example, suspect that the other person has given too much weight to a specific piece of evidence, and begin a dialogue about this question.

One way of understanding the relation between Hume and Smith would be to see them as giving competing explanations of the same phenomena. Another possibility, however, is that both mechanisms

https://doi.org/10.1093/arisoc/aoad004

can occur, in different situations. There is the level of immediate spontaneous reactions, for example by mirroring a smile (Darwall 1998, p. 264), and there is the more complex process of reacting to someone's anger or joy by imagining what it must be like to be in their situation, which sometimes begins with an almost automatic reaction, but which then takes us onto a more reflective level. The second mechanism might even lead us to rejecting the sentiments or opinions of the first person (and our own first spontaneous reaction) because we conclude that their reaction was inappropriate.

Admitting that both phenomena exist aligns with the fact that Smith's text includes a few passages that sound very much like 'mirroring', even though he is keen to emphasize that there must be more to sympathy (see also Khalil 2015, pp. 656, 678). It is also a possible way of explaining some of the vagueness in Hume's account, as when he uses phrases such as 'in a greater or less degree' (*Treatise* III.III.II, p. 592). We can, for example, imagine that we are immediately attracted to the opinion held by another person but, reflecting on their epistemic situation, realize that we do not fully agree with their judgement.

Hume's and Smith's discussions include several additional features that are useful for thinking about these phenomena. One concerns the *pleasure* that humans derive from sympathetic and empathetic experiences. If our mood is lightened by observing the smile of another person, this is a pleasurable experience. The experience of being infected with negative feelings, however, is more complex, because the feeling we are infected with is negative, while the fact *that* we are infected by it induces a positive feeling. For Hume, the outcome in such cases depends on several factors, such as the closeness of the individuals and the intensity of the feelings (*Treatise* II.II.IX, esp. p. 387). Smith, in contrast,[6] draws a clear distinction between two kinds of sensations: the primary sensation that is being shared and a second-level sensation that arises *from the phenomenon of sympathy* (*TMS* I.III.I.9 n.14).

For recipients, the second-level sensation is always positive: 'nothing pleases us more than to observe in other men a fellow-feeling with all the emotions of our own breast' (*TMS* I.I.II.1). And it seems

[6] And in reaction to discussions about the first edition of the book, including a letter from Hume; see Sagar (2017, p. 687); see also Raynor (1984) on the interaction between Hume and Smith.

https://doi.org/10.1093/arisoc/aoad004

that for spectators, this holds as well, at least in cases in which the spectator ends up sharing the first person's sentiment: 'the emotion which arises from his [the spectator's] observing the perfect coincidence between this sympathetic passion in himself, and the original passion in the person principally concerned' is 'always agreeable and delightful' (*TMS* I.III.I.9 n.14; see also Khalil 2015, pp. 663–9; Sugden 2002, pp. 70–3; Fleischacker 2012, p. 300).[7]

It thus seems that there is a certain 'pull' towards agreement with the feelings (and probably, in parallel, the opinions) of others, because this allows us to feel the pleasure that arises from the awareness of sympathizing with them (and, one might add, the anticipation that the other person will in turn feel positive about being sympathized with; cf. *TMS* I.I.I.2). This might even stop us from engaging in the reflexive process that Smith describes and ask about the appropriateness of a reaction, because we would risk leaving the space of pleasant mutual sympathy. It may take a certain degree of will-power, or the virtuous traits of a trained moral character, to nonetheless ask questions about appropriateness that move us in the direction of an impartial view.

Another feature that both Smith and Hume emphasize is proximity: the closer a person is to us, or the more we have in common with them, the stronger the sympathetic effect. Hume writes, 'The stronger the relation is betwixt ourselves and any object [here, other people], the more easily does the imagination make the transition, and convey to the related idea the vivacity of conception, with which we always form the idea of our own person' (*Treatise* II.I.XI, p. 318). As a result, 'we sympathize more with persons contiguous to us, than with persons remote from us' (*Treatise* III.III.I, p. 581; see also McHugh 2018, pp. 686–7). As Rick notes, there is an epistemic dimension here: the closer we are to a person, the better we can understand them (2007, p. 146). If people are more distant, in contrast, the transmitted sentiment or opinion leaves less of an impression—and it may even reverse its sign if the person is a 'rival', with whose pleasure we feel pain, and vice versa (*Treatise* II.II.IX, p. 384).

Smith agrees with this point, but with a twist. He describes a picture that Forman-Barzilai (2010) has characterized as 'circles of sympathy', in which human sympathy reaches out, in concentric

[7] Of course, this does not hold if the spectator comes to the conclusion that the sentiment is inappropriate.

https://doi.org/10.1093/arisoc/aoad004

circles, to those around us, and gets weaker the greater the distance to them (*TMS* VII.II.I.44). We therefore care more about ourselves and people around us than about some strangers in China who are threatened by an earthquake (*TMS* III.III.4). But this tendency is problematic: it contradicts the impartiality that is often required for *moral* judgement. That is why for Smith, the *lack* of sympathy that more distant individuals have for our plight is an important corrective to our tendency to give too much weight to ourselves and our loved ones, because it reminds us that, seen from a greater distance, our problems are not so different from those of other people (*TMS* III.III.22).

Very often, this critical distance to ourselves and those close to us is necessary in human interactions. Our spontaneous reactions may lead to moral or epistemic mistakes, which we make in the heat of the moment, when going along with someone else's emotions or opinions. The virtuous individual needs to learn not to give in to his or her immediate reactions, but rather to ask the question of appropriateness, whether of emotions or of opinions. For Smith, such self-control is an important component of morality (*TMS* VI.III).

Of course, the ability to maintain a distance from the emotions and opinions of others and not to be swayed by them is not, per se, what makes human behaviour moral. A potential murderer might be affected, through empathy or initial sympathy, by the fear of her potential victim, and an inability to control her affective states would make a better outcome more likely (no murder would occur). But arguably, such constellations are rare; the more likely scenario is one in which we would, in a cool mode, know the right thing to do, but are swayed by spontaneous emotions or opinions that we inappropriately take over from others, and act in ways that we later regret. If this assumption is correct, then *on average*, it is better to critically reflect on someone else's situation and the emotions or opinions that *should* follow from it, instead of unthinkingly going along with them.

III

Sympathy and Empathy on Social Media. How, then, can the distinction between sympathy and empathy help us understand communicative processes on social media? The phenomena I am interested in are the sharing of information or other items (pictures, memes,

https://doi.org/10.1093/arisoc/aoad004

and so on) in ways that suggest that some transmission of 'passions, sentiments and opinions' takes place.[8] Of course, on social media the conditions for such transmissions are rather specific. There is no reciprocity: users' followers need not be the ones they themselves follow (Marwick and boyd 2010, p. 116). Moreover, lacking face-to-face communication, users often do not know who sees their posts and therefore *imagine* an audience; typically, they imagine it to be similar to themselves (Marwick and boyd 2010, p. 120). In Smithian language, one could say that they imagine a 'circle of sympathy'.

As an example of such transmission processes and the ways in which they can take unexpected turns, take a case reported by New York journalist P. E. Moskowitz (2021). Having parked their car in a very narrow parking space and bragging about it on Twitter, they caused a wave of outrage, with other users accusing them of being a bad person (for example for making it harder for other car owners to leave, but also for bragging) and threatening to damage their car or even beat them up. What makes this a useful example is that it seems a relatively 'pure' case of internet outrage, rather than one in which contested moral or political issues would be at stake. It also illustrates that the harmfulness of such phenomena is cumulative: one or two derisive comments would hardly cross the threshold for counting as harm, but the mass and intensity of a whole wave of such reactions is likely to leave its mark on the victim.

What is clearly part of the phenomenology of this case, and of many others, is that the users feel pleasure in sharing emotions. Pouring out anger over someone seems to work just as well, for that purpose, as sharing other content. No matter what the underlying emotion is, there is a second-level, positive emotion arising from the sheer fact of learning about the same emotion in others—as described as a feature of empathy or sympathy by both Smith and Hume. Arguably, this pleasure contributes to the attractiveness of social media (see also Nguyen 2021). Given that their commercial model relies on maximizing users' interaction time, their algorithms show individuals the content that is most likely to lead to extended interactions, but not necessarily because the *content itself* would be

[8] I take no stance here on the question of whether the algorithms of social media platforms lead to 'filter bubbles' (Pariser 2011). This is empirically contested: various studies seem to confirm that online networks provide users with a rather diverse news diet; see, for instance, Bruns et al. (2017).

https://doi.org/10.1093/arisoc/aoad004

particularly pleasant or convincing (it might well be 'fake'). Instead, the intensity of the second-order sentiment seems crucial for keeping people engaged.

Another relevant feature of social media is that a feeling of psychological closeness is highly valued there. Marwick and boyd (2010) show that Twitter users, for example, care a lot about appearing 'authentic', through strategies such as sharing personal content, for example, their musical tastes. This relates to the argument by Smith and Hume that empathy and sympathy, and hence also the positive emotions that accompany them, are most intense when the interactions are with those close to us. Apparently, it is possible to create a kind of simulacrum of proximity where followers *think* they are close to a person because they know certain random details about them that one usually only knows of family members and friends.

Internet scholars, especially danah boyd, use the notion of 'context collapse' for describing the way in which communication on social media 'flattens multiple audiences into one' (Marwick and boyd 2010, p. 122). The normal distinctions between different social spheres and their different social norms and expectations break down. A tweet might be read by my second cousin, a childhood sweetheart, a member of my sports club, and so on. This breaks up the offline logic of the 'circles of sympathy', in which there is a rough correspondence between how close we are to a person, how much we know about them, and how much we care about them. We might *think* that we know a person relatively well—say, well enough to react to a tweet with a mean comment that is meant to be funny—when in fact we do not, thus not knowing what harm our comments might do. And because of the lack of reciprocity, we might never learn what their reaction was, so learning processes are inhibited. This contrasts with similar phenomena in the offline world, for example, when a group heaps scorn on an individual. Often, the visibility of the reaction of the victim and its impact, via empathy or sympathy, would stop the attackers, or at least make them think twice. The 'context collapse' of social media undermines this counter-mechanism.

One might react to these points by saying that they only apply to processes in which the transmission between individuals is purely emotional, without any cognitive content. For the example of Moskowitz's parking job, this may be true. But in many other cases, cognitive and emotional content seem to be almost inseparably

https://doi.org/10.1093/arisoc/aoad004

intertwined, and their transmission seems to follow the same logic. For example, there is empirical evidence that individuals often share articles when they find the headlines appealing, but without reading the text as a whole (Dewey 2016). Also, for Twitter it has been established that on average, fake news travel faster than real news (Vosoughi, Roy and Aral 2018), and that morally loaded emotional framing increases the likelihood that posts go viral (for example, Brady et al. 2017, quoted in Steinert 2020). One possible explanation is that the emotional intensity of fake news (in terms of surprise, outrage, and so on) is higher than that of normal news, and that this is one explanation of why they are shared more often.

It seems that in such cases, the Humean rather than the Smithian model prevails—a kind of automatic, semi-conscious contagion, instead of a process of calm and (self-)critical reflection in which individuals might also arrive at the conclusion that certain emotional or cognitive reactions are exaggerated or misguided. One background factor that is likely to contribute to such unthinking behaviour is the 'information overload' that individuals experience in the digital world (Lorenz-Spree et al. 2020, p. 1104), which invites quick processing and decision making. Instead of engaging in a potentially more long-winded process of reflecting how one would react by putting oneself into the other person's shoes, one feels a similar emotion, or spontaneously shares an opinion—and then passes it on to others, in the (maybe semi-conscious) expectation of experiencing the positive meta-emotion that comes from sharing. The offline phenomenon that probably comes closest, and to which it is sometimes compared, is the kind of herd behaviour that one finds in mobs.[9]

There have indeed been some empirical studies on 'emotional contagion' in social media. One was a highly controversial study by researchers from Facebook (Kramer, Guillory and Hancock 2014) who manipulated the posts that users saw (for critical discussions see, for instance, Grohol 2014 and Verma 2014). They showed that if users saw fewer posts with positive words, there was a slight decrease of positive content among their own posts, and vice versa for negative terms (Kramer, Guillory and Hancock 2014, p. 8789). This effect has also been confirmed in a methodologically

[9] Khalil rightly points out that 'herd behavior, mob psychology, and informational cascades' (2015, p. 654), while more easily explained by the Humean model, are phenomena that Smith also acknowledged (2015, p. 676).

https://doi.org/10.1093/arisoc/aoad004

and ethically sounder study (Ferrara and Young 2015). The authors admit, though, that a number of mechanisms might have led to this effect, not only emotional contagion (Ferrara and Young 2015, p. 11).[10] More empirical research will be needed to sort out the different possible explanations.

However, a critic might object that from a normative perspective, the distinction between empathy and sympathy is less important than the question of *who* or *what* users empathize or sympathize with. It is certainly a major problem with mob-like online behaviour that users share feelings or opinions with each other, but without asking about the targets: they care more about sharing a rant over Moskowitz's brag with others than about the effect this might have on Moskowitz. As long as they exclude Moskowitz from their 'circle of sympathy', it might not make a difference whether they follow the Humean or the Smithian model, one might object. But arguably, the Smithian model is better at inviting the question of what a victim might feel. If a user considers whether to share a mean comment, she might imagine what she or other would feel if they were the target, and she might thus be led to also ask what Moskowitz would feel. It is the in-built dynamic towards further questions ('How would yet others feel?' 'How would an impartial spectator feel?') that makes this more likely in the Smithian than in the Humean model.

To be sure, a lot of sharing on social media probably also takes place along Smithian lines, with users reflecting on what they would feel or think, and how others might feel or think, about a situation or a piece of information. Users often add their own perspective when sharing links, which might signal distance from the original content (and of course, many decisions *not* to share certain content might be based on sympathetically induced concerns about appropriateness). The types of behaviours that look more Humean, in contrast, seem to grow out of a kind of superficial sense of shared sentiments or opinions. What is missing in them is a feature of human sociability that Smith, for one, found very important: seeing things from a

[10] Bor and Petersen (2022) have recently shown that online political hostility tends to come from individuals who are status-driven and intentionally use hostile strategies; they do so online and offline, but online, their behaviour is more visible. Expressed in the terms of my paper, one might say that online, a mob stands ready to be stirred up, which is far less likely in the offline world.

https://doi.org/10.1093/arisoc/aoad004

variety of different perspectives, thereby getting a more nuanced and more neutral picture of the matter at hand.[11]

The Smithian process, however, requires emotional and cognitive work. By asking what one would feel or think in a certain situation, one is confronted with the fact that others feel or think differently, which challenges one's own positions. There is an almost inevitable pull—at least if one is willing to follow a train of thought where it takes one—from 'What would I have done in that situation?' to 'What would other people have done?' to 'What would have been the *right* thing to do?' These are questions one can discuss with oneself, but also with others, though typically one does so in circles of trusted others, where one can risk uttering half-developed opinions, or say something that one might want to take back later. Social media can offer such spaces (typically, in small, closed groups), but it is often not what is publicly visible and what is perceived as 'important', as when journalists report about something 'trending on social media'.

Ultimately, what is at stake here are two very different logics of intersubjective engagement. In the first, there is a fusion of sentiments: one passively loses oneself in the crowd, being drawn away by its fluctuations (and potentially contributing to the cumulative harm it causes). In the second, individuals scrutinize their own impulse to go along with others: they want to learn from others, are willing to be challenged, and to and see the world from other perspectives. In the latter scenario, which brings together different views, the 'wisdom of the crowd' can potentially be harvested. Richard Seymour (2019, p. 18) calls these two logics 'hype' and 'hivemind'; Will Davies (2021) connects them to the distinction between 'reputation', based on numbers of reactions, and 'recognition', based on conscious critique according to some external standard. Both logics are part of human life, online and offline—but the question is which one social media allow and encourage.[12]

Cultural pessimists might fear that there is a general crowding out of nuanced forms of interaction. But as long as human beings

[11] Another danger of online communication is that because we do not know how the algorithms work, we do not know whether the perspectives we get are an unbiased sample or pre-selected in some way. For reasons of space, I cannot discuss this issue here.

[12] Behind this issue lurk bigger questions about social polarization (for example, Mutz 2006) that I cannot address here. Social media are of course not the only factor driving it, but there are deep questions about whether they exacerbate it.

https://doi.org/10.1093/arisoc/aoad004

also find other ways for practising more Smithian exchanges, such worries might be exaggerated. However, if it is indeed the case that certain phenomena—especially waves of outrage on social media—follow the Humean model, they should be conceptualized as such: users' decisions to post or share certain content might be opportunities for them to get a warm fuzzy feeling of community, rather than genuine expressions of views or sentiments.[13] If this is the case, such postings should not be misunderstood as saying something deeper about people's views.

Unfortunately, though, this does little to help the victims of hate waves or false allegations; simply telling them that 'people did not mean it' may seem cynical, given the amount of abuse that may wash over them. The question that suggests itself, instead, is whether there might be mechanisms that could reduce the amount of unreflective behaviour that is motivated by the search for the pleasant warm feeling of sharing certain emotions or opinions with others, in favour of more controlled, (self-)critical forms of behaviour. There are at least two normative bases for the regulation of these phenomena (which of course need to be balanced against other considerations, for example, freedom of speech). The first is the imperative to prevent the harm that certain Humean waves can cause for victims, especially in cases in which they incite real violence.[14] The second is a more general argument about making sure that at least in parts of social media, communication can take place along Smithian lines, allowing individuals to enjoy its benefits (mutual learning, exchanges of perspectives, and so on). The current undifferentiated forms of discourse on social media have probably driven away numerous users from participating in online discussions, and while it may be commendable to also leave spaces in the online world for emotional outbursts, it is questionable whether these need to be the major sites used by millions of people.

One can distinguish three broad strategies when thinking about ways to reduce unthinking behaviour. The first is to appeal to human

[13] This argument can be strengthened further, with regard to 'fake news', by taking up an argument by Rigi (2021), who holds that various mechanisms that ensure reliable testimony in real life do not work on social media.

[14] Facebook whistleblower Frances Haugen reported many such problems; see, for example, Haugen (2021).

https://doi.org/10.1093/arisoc/aoad004

virtue.[15] For the issue under discussion, this leads to questions about moral education (maybe as part of 'digital literacy') and other ways of improving and individual's character. There may well be educational strategies that help individuals, and especially children and teenagers, to become better human beings online. For example, there have been promising experiments of 'inoculating' individuals against fake news by teaching them about how it is created (Kozyreva, Lewandowsky and Hertwig 2020). Maybe there could also be pedagogical interventions that let individuals understand what it is like to be the target of an internet mob, and thereby inculcate in them a higher degree of awareness of the potential damage done by the unthinking sharing of problematic content.

A second, related, strategy builds on social norms, based on the assumption that human beings are highly sensitive to approval or disapproval from their peers. But it quickly leads to worries that were famously articulated by John Stuart Mill about pressures to conform and about silencing individuals with dissenting opinions (Mill [1859] 1991, ch. 2). Much more would have to be said here—about the assumptions of the Millian picture, about potential pitfalls, and so on—but it is worth pointing out one feature of the strategies that would suggest itself based on the distinction between Humean and Smithian mechanisms. The social norms in question would not need to refer directly to the content that is being shared, but rather to the meta-sentiment of enjoying the very act of sharing a bit too much. They would aim at introducing a counterweight to the pull of this sentiment, in favour of a more critical attitude.

The third strategy concerns the technical environment and its incentives and disincentives, allowances and restrictions. In the case of social media, this means focusing on the technical architecture of platforms.[16] Here, the problem of undue pressure on individuals takes an even more worrying form: what about the risk of inappropriate censorship and the suppression of unpopular opinions? Again, this is a large and complex area (for example, regarding the

[15] With regard to *cognitive* skills, a term that is used in social psychology for such interventions is 'boosting' (see, for instance, Lorenz-Spree et al. 2020); what would be needed here, however, is *moral* 'boosting'.

[16] Another possibility would be to change the legal framework, for example, by increasing the punishment for certain actions. This may be useful for extreme cases of hate speech, but there are serious challenges concerning enforcement, especially when the harm in question comes about through the *sum* of many speech acts.

https://doi.org/10.1093/arisoc/aoad004

question of whether certain forms of (hate) speech should indeed be banned). But the cues we can take from the focus on sympathy and empathy can avoid many problems by focusing, not on content, but on the psychological mechanisms that lead to either blind sharing or a more reflective stance (see similarly Lorenz-Spree et al. 2020, p. 1103). More specifically, the question is whether there could be features of the architecture of social media—for example, with regard to their temporal dynamics—that make one or the other more likely, and that enable individuals to avoid unthinking behaviour that they would later regret.

We can take some inspiration here from the discussion of Smith and Hume above. One concerns the question of speed: the Humean process happens immediately, while the Smithian one takes more time—at least if it is meant to also include some of the further reflections, up to 'What would an impartial spectator do?' This suggests that decreasing the speed with which individuals can react, or introducing additional warning mechanism (for example, about the reliability of a source) might give individuals more opportunities for second thoughts and reflective behaviour. Another set of considerations concerns the size of the audience and the relationship one has with the people who participate in a discussion, to counteract the simulacrum of closeness that one often finds on social media.[17] For example, individuals might receive information about the chains through which a certain piece of content has reached them, how many people have shared it in what time frame, and how much back-and-forth communication has taken place.

Similar proposals—with a focus on truthfulness and autonomy—come from social psychologists. As in Hume and Smith, one of their starting assumptions is that social contexts matter for individual behaviour, even though it is a probabilistic relation and different individuals will always react differently. One influential recent approach looks at possibilities of designing the architecture of social

[17] There is a broader regulatory question in the background, concerning the relation between 'public' media and 'private' communication. If communication is considered 'public', criticisms and correction mechanisms can counteract fake news or wrong accusations. One recent phenomenon is the move of conspiracy theorists to messenger services where their messages are *not* publicly visible and only like-minded individuals see them, which massively reduces the likelihood of criticism and correction. Maybe there needs to be space for uncontradicted falsehoods in small pockets of society, but in messenger services, the size of the audience is massively increased, raising questions about whether certain forms of 'fact checking' would be appropriate.

https://doi.org/10.1093/arisoc/aoad004

media in ways that make the sharing of fake news less likely. Lorenz-Spree et al. (2020), for example, discuss various ways in which small changes in the choice environment of social media users, such as the provision of meta-information about content or the introduction of additional steps before one can share a piece, might help users to behave in more autonomous ways and to spread fewer fake news items. Various studies found that 'nudges' concerning the reliability of news items have a positive effect on people's behaviour (for example, Momen Bhuiyan 2021). Clearly, more research would be helpful here. If such 'nudges' were implemented on a broader scale, however, questions would also need to be asked about transparency, and about the accountability of those who have the power to 'nudge'.

If my arguments drawn from Smith and Hume are correct, then such interventions should probably not be too far on the cognitive end. Often, what they need to achieve is counterbalancing an emotion: the warm feeling that arises from sharing emotions or opinions. Maybe appealing to other positive emotions, such as pride in taking responsibility for one's social media behaviour, could have at least as strong an effect as providing individuals with additional information. This, at least, is the hypothesis that one would want to pass on to empirical researchers for further exploration.

IV

Conclusion. In this paper, I have drawn on Hume's and Smith's accounts of empathy and sympathy to reflect on the ways in which the transmission of 'passions, sentiments and opinions' takes place on social media. I have argued that we can distinguish two models: a non-cognitive, almost instinctive 'contagion' with the emotions or opinions of others, and a conscious 'sharing place' in which one asks how one would oneself react to a certain situation, which can lead to further questions about how one *should* react to it. Proximity to a person is a key feature in determining the strength of these phenomena, as is the positive emotion that arises from the very act of sharing. But a key difference between the two models is the way in which the possibility of critical distance is built into the second, but not the first.

The Humean model helps explain the way in which emotionally loaded content can lead to whole cascades of unthinking content sharing, because social media users empathize with others who post

https://doi.org/10.1093/arisoc/aoad004

certain content and want to receive empathy in turn. This is harmless when the contents are cute cat videos, but it is deeply harmful if it is, say, resentment or derision of minority groups. This leads to the question of whether social media might enable more Smithian processes through certain interventions, especially in the architecture of social media platforms. Mechanisms that counter the pull of the warm feeling of sharing might play an important role here.

However, the business models of social media platforms thrive on maximum engagement, with no consideration of developing forms of sociability in which individuals can exchange different perspectives, learn from each other, and develop their moral character. It is, therefore, questionable whether one can expect commercial platforms to improve their architecture along the lines that I have discussed.[18] Public regulation might nonetheless impose certain rules on them. The advantage of regulation that focuses on the architecture of platforms is that it avoids the difficult territory of content regulation. Instead, it would better enable citizens to draw on those mechanisms of self-restraint and reflection that they are used to drawing on in the offline world. This could be a step towards making social media more suitable to human sociability of a kind that individuals do not, in retrospect, regret.[19]

University of Groningen
Oude Boteringestraat 52
9712 GL Groningen
Netherlands
l.m.herzog@rug.nl

References

Bloom, Paul 2016: *Against Empathy: The Case for Rational Compassion*. New York: Ecco.

[18] Theoretically, one could imagine that new platforms with a ‘better’ architecture (in this sense) would emerge in the market and that users would switch to them. But because of network effects and because of the market power of the existing platforms, it is not clear whether this could be successful.

[19] I would like to thank audiences at Delft University of Technology and at the meeting of the Aristotelian Society for their comments and questions, and Steffen Steinert and Guy Longworth for detailed written comments.

Proceedings of the Aristotelian Society, Vol. 123, *Part* 1
https://doi.org/10.1093/arisoc/aoad004

Bor, Alexander, and Michael Bang Petersen 2022: 'The Psychology of Online Political Hostility: A Comprehensive, Cross-National Test of the Mismatch Hypothesis'. *American Political Science Review*, 116(1), pp. 1–18.

Brady, William J., Julian A. Wills, John T. Jost, Joshua A. Tucker, and Jay J. Van Bavel 2017: 'Emotion Shapes the Diffusion of Moralized Content in Social Networks'. *Proceedings of the National Academy of Sciences (PNAS)*, 114(28), pp. 7313–18.

Bruns, Alex, Brenda Moon, Felix Münch, and Troy Sadowsky 2017: 'The Australian Twittersphere in 2016: Mapping the Follower/Followee Network'. *Social Media and Society*, 3(4), pp. 1–15.

Cohon, Rachel 1997: 'The Common Point of View in Hume's Ethics'. *Philosophy and Phenomenological Research*, 57(4), pp. 827–50.

Darwall, Stephen 1998: 'Empathy, Sympathy, Care'. *Philosophical Studies*, 89(2/3), pp. 261–82.

Davies, Will 2021: 'The Politics of Recognition in the Age of Social Media'. *New Left Review*, March/April 2021. https://newleftreview.org/issues/ii128/articles/william-davies-the-politics-of-recognition-in-the-age-of-social-media.

Dewey, Caitlin 2016: '6 in 10 of You Will Share this Link without Reading It, a New, Depressing Study Says'. *Washington Post*, 16 June 2016. https://www.washingtonpost.com/news/the-intersect/wp/2016/06/16/six-in-10-of-you-will-share-this-link-without-reading-it-according-to-a-new-and-depressing-study/.

Ferrara, Emilio, and Zeyao Yang 2015: 'Measuring Emotional Contagion in Social Media'. *PLoS ONE*, 10(11), p. e0142390. https://doi.org/10.1371/journal.pone.0142390.

Fleischacker, Sam 2012: 'Sympathy in Hume and Smith: A Contrast, Critique, and Reconstruction'. In Christel Fricke and Dagfinn Føllesdal (eds.), *Intersubjectivity and Objectivity in Adam Smith and Edmund Husserl: A Collection of Essays*, pp. 273–311. Heusenstamm: Ontos.

Forman-Barzilai, Fonna 2010: *Adam Smith and the Circles of Sympathy: Cosmopolitanism and Moral Theory*. Cambridge: Cambridge University Press.

Grohol, John M. 2014: 'Emotional Contagion on Facebook? More Like Bad Research Methods'. *Psych Central* blog, 23 June 2014. https://psychcentral.com/blog/emotional-contagion-on-facebook-more-like-bad-research-methods.

Haugen, Frances 2021: 'A Conversation with Facebook Whistleblower Frances Haugen'. *Your Undivided Attention podcast*, Episode 42, 18 October 2021. https://www.humanetech.com/podcast/42-a-conversation-with-facebook-whistleblower-frances-haugen.

Hume, David [1739–40] 1978: *A Treatise of Human Nature*, 2nd edn. Edited, with an analytical index, by L. A. Selby-Bigge, with text revised

https://doi.org/10.1093/arisoc/aoad004

and variant readings by P. H. Nidditch. Oxford: Clarendon Press. Cited as Treatise.

—[1751] 1975: *An Enquiry Concerning the Principles of Morals. In Enquiries Concerning Human Understanding and Concerning the Principles of Morals*, 3rd edn. Edited by L. A. Selby-Bigge, with text revised and notes by P. H. Nidditch. Oxford: Clarendon Press.

Khalil, Elias L. 2015: 'The Fellow-Feeling Paradox: Hume, Smith and the Moral Order'. *Philosophy*, 90(4), pp. 653–78.

Kozyreva, Anastasia, Stephan Lewandowsky, and Ralph Hertwig 2020: 'Citizens Versus the Internet: Confronting Digital Challenges With Cognitive Tools'. *Psychological Science in the Public Interest*, 21(3), pp. 103–56.

Kramer, Adam D. I., Jamie E. Guillory, and Jeffrey T. Hancock 2014: 'Experimental Evidence of Massive-Scale Emotional Contagion through Social Networks'. *Proceedings of the National Academy of Sciences (PNAS)*, 111(24), pp. 8788–90.

Lorenz-Spree, Philipp, Stephan Lewandowsky, Cass R. Sunstein, and Ralph Hertwig 2020: 'How Behavioural Sciences Can Promote Truth, Autonomy and Democratic Discourse Online'. *Nature Human Behaviour*, 4, pp. 1102–9.

McHugh, John 2018: 'Working Out the Details of Hume and Smith on Sympathy'. *Journal of the History of Philosophy*, 56(4), pp. 683–96.

Marwick, Alice E., and danah boyd 2010: 'I Tweet Honestly, I Tweet Passionately: Twitter Users, Context Collapse, and the Imagined Audience'. *New Media and Society*, 13(1), pp. 114–33.

Mill, John Stuart [1859] 1991: *On Liberty*. Edited by John Gray and G. W. Smith. London and New York: Routledge.

Momen Bhuiyan, Md, Michael Horning, Sang Won Lee, and Tanushree Mitra 2021: 'NudgeCred: Supporting News Credibility Assessment on Social Media Through Nudges'. *Proceedings of the Association for Computing Machinery on Human-Computer Interaction*, 5(CSCW2), Article 427, pp 1–30.

Morrow, Glenn 1923: 'The Significance of the Doctrine of Sympathy in Hume and Adam Smith'. *Philosophical Review*, 32(1), pp. 60–78.

Moskowitz, P. E. 2021: 'The Parallel-Parking Job That Ignited the Internet'. *Curbed*, 3 August 2021. https://www.curbed.com/2021/08/p-e-moskowitz-parallel-parking.html.

Mutz, Diana C. 2006: *Hearing the Other Side: Deliberative versus Participatory Democracy*. Cambridge and New York: Cambridge University Press.

Nguyen, C. Thi 2021: 'Twitter, the Intimacy Machine'. *The Raven*, 1, Fall 2021. https://ravenmagazine.org/magazine/twitter-the-intimacy-machine/.

https://doi.org/10.1093/arisoc/aoad004

Pariser, Eli 2011: *The Filter Bubble: What the Internet Is Hiding from You*. New York: Penguin.

Pearlman, Jeff 2011: 'Tracking Down My Online Haters'. *CNN*, 21 January 2011. https://edition.cnn.com/2011/OPINION/01/21/pearlman.online.civility/index.html.

Raynor, David R. 1984: 'Hume's Abstract of Adam Smith's Theory of Moral Sentiments'. *Journal of the History of Philosophy*, 22(1), pp. 51–79.

Rick, Jon 2007: 'Hume's and Smith's Partial Sympathies and Impartial Stances'. *Journal of Scottish Philosophy*, 5(2), pp. 135–58.

Rigi, Regina 2021: 'Weaponized Skepticism: An Analysis of Social Media Deception as Applied Political Epistemology'. In Elizabeth Edenberg and Michael Hannon (eds.), *Political Epistemology*, pp. 31–48. Oxford: Oxford University Press.

Robb, Amanda 2017: 'Anatomy of a Fake News Scandal'. *Rolling Stone*, 16 November 2017. https://www.rollingstone.com/feature/anatomy-of-a-fake-news-scandal-125877/.

Sagar, Paul 2017: 'Beyond Sympathy: Smith's Rejection of Hume's Moral Theory'. *British Journal for the History of Philosophy*, 25(4), pp. 681–705.

Sayre-McCord, Geoffrey 2013: 'Hume and Smith on Sympathy, Approbation, and Moral Judgment'. *Social Philosophy and Policy*, 30(1–2), pp. 208–36.

Sest, Natalie, and March, Evita 2017: 'Constructing the cyber-troll: Psychopathy, sadism, and empathy'. *Personality and Individual Differences*, 119, pp. 69–72.

Seymour, Richard 2019: *The Twittering Machine: How Capitalism Stole Our Social Life*. London: Indigo Press.

Smith, Adam [1776] 1976: *The Theory of Moral Sentiments*. Edited by D. D. Raphael and A. L. Macfie. Oxford: Clarendon Press. Cited as TMS.

Steinert, Steffen 2020: 'Corona and Value Change: The Role of Social Media and Emotional Contagion'. *Ethics and Information Technology*, 23, Supplement 1, pp. S59–68.

Sugden, Robert 2002: 'Beyond Sympathy and Empathy: Adam Smith's Concept of Fellow-Feeling'. *Economics and Philosophy*, 18(1), pp. 63–87.

Van Holthoon, F. L. 1993: 'Adam Smith and David Hume: With Sympathy'. *Utilitas*, 5(1), pp. 35–48.

Verma, Inder M. 2014: 'Editorial Expression of Concern and Correction'. *Proceedings of the National Academy of Sciences (PNAS)*, 111(29), p. 10779.

Vosoughi, Soroush, Deb Roy, and Sinan Aral 2018: 'The Spread of True and False News Online'. *Science*, 359(6380), pp. 1146–51.

Proceedings of the Aristotelian Society, Vol. 123, *Part* 1
https://doi.org/10.1093/arisoc/aoad004