

In this paper, I will propose a systematic philosophical approach to explain and describe instances of hostile interaction in online environments, specifically polarization and trolling on social networking sites. Specifically, I will propose a recent approach in philosophy of mind, known as “mindshaping” that sees social cognition as fundamentally relying on norms of behavioral expectations. I intend to show that this approach to social cognition can be particularly resourceful for understanding the social and psychological mechanics of hostile online interaction.

After explaining the notion of mindshaping in the context of philosophical theories of social cognition, I will describe some forms of online hostility, with a special focus on trolling, as an instance of mindshaping - specifically, as a way of making others predictable through the imposition of behavioral expectations by taking advantage of some intrinsic features of social networking sites. The most important characteristics of social networking platforms that are at play in hostile online behavior are the somewhat disembodied character of computer-mediated communication considered one of the main causes for online disinhibition; and what is known as context collapse, the overlapping of audiences for generated content on social networking sites. I will argue that the involvement (and, in the case of trolling, the exploitation) of these two factors are some of the most important causes of hostile behavior and incivility in social networking sites.