

Understanding and Decreasing Aversive Behavior in Online Social Contexts

John P. Davis

Social Computing Group
Microsoft Research
One Microsoft Way
Redmond, WA 98052-6399 USA
jpdavis@microsoft.com
+1 425 706-5860

Abstract

Because the Internet is a social space, it has many of the problems associated with social interactions. 'Bad' or inappropriate behavior is an important problem in many online social settings. However, it appears that methods to cope with and correct such behavior do not fully capitalize on some useful social influence mechanisms that are available. In this paper, I explore some of the more social explanations for online aversive behavior and methods to combat such behavior that are based on the explanations.

Introduction

Aversive or 'bad' behavior is a growing problem in many online social environments. There are abundant examples of bad behavior; one has merely to visit a chat room with the word 'teen' in the title to experience it. In a recent survey I conducted, users reported that they had regularly been stalked, harassed, spammed, flamed, scammed, and infected with viruses in the course of their online interactions with others (Davis, 2002). Their experiences of bad behavior took place in diverse online spaces, like chat rooms, email, bulletin boards, discussion groups and online games. Surprisingly, users reported that the source of the bad behavior they experienced was often people they knew, though usually only casually. According to the results of this survey, users think that bad behavior occurs in online settings far more often than in face-to-face interactions.

Explanations for Bad Behavior and Methods for Combating It

Individual Differences (Personality)

Some explanations for aversive behavior on the Internet focus on the personality of the perpetrators of such

behavior. For example, bad behavior can stem from users' ignorance about how to act, lack of skill in acting appropriately, lack of control over their behavior, laziness, malevolence or selfishness (Miller, 2001). 'Treatments' for perpetrators may include personalized interventions, like giving them attention or educating them about the error of their ways. Such interventions may be carried out by human monitors in a particular environment, or perhaps by other users in the environment. Although having monitors police online environments and deal with aversive behavior on a case-by-case basis can be effective, such responses are also very expensive to administer.

Social Presence

Other explanations for aversive behavior focus on the social factors, or qualities of the context, that affect whether and when it will occur. For example, social presence theory (e.g., Short, Williams & Christie, 1977) focused on the qualities of the communication medium that convey the sense that the parties to the communication are 'real'. To the extent that a medium is able to carry social information, such as social cues and tone, it is considered more social and more similar to face-to-face interactions. Presumably, the more social information the medium can transmit, the more likely it is to engage the behavioral norms that accompany face-to-face interactions.

A social presence explanation for aversive behavior suggests that increasing the salience of the interpersonal nature of an online interaction will lead to a decrease in bad behavior. Voice carries a great deal of both semantic and social information, so it is natural to think that the ability to communicate through a voice channel would decrease the tendency to behave badly toward another. In fact, previous research demonstrated that even text-to-speech (TTS) voice communication affected the extent to which people cooperated in a game involving a social dilemma (Jensen, Farnham, Drucker & Kollock, 2000). Voice communication provides rich information about social presence that make interactions satisfying (Connell, Mendelsohn, Robins & Canny, 2001).

Because the effects occurred even though the computer generated voice contained the same semantic content as simple textual communication, they led us to wonder whether participants were inferring additional personality characteristics about their gaming partners when they heard a voice, causing them to behave differently toward them. If so, perhaps access to another's personalized information, the type that may be inferred through a voice, would also lead to increased social presence and decreased aversive behavior. To explore this possibility, we conducted a follow up study that varied mode of communication (i.e., TTS vs. text only) and the presence of personal profile information. In this study, we found that voice communication decreased aversive behavior, but having access to personal profile information did not (Davis, Farnham & Jensen, 2002). The presence of voice and profile information did, however, cause the partners to rate each other more positively before they played the game. Individuating profile information seemingly has effects, but the exact nature of those effects, and their relation to good and bad behavior toward others, remain to be explored (see Swinsh, Farnham & Davis, 2002).

Deindividuation and Social Identity

Some have suggested that disinhibited behavior in computer mediated communication (CMC) is caused by deindividuation, which is the feeling that one has no personal identity or uniqueness. Deindividuation may be caused by several factors, including a reduction in the sense of accountability for one's actions. CMC, compared to face-to-face interactions, may lead to differences in private self-awareness (a focus on internal motives, attitudes and goals) and public self-awareness (a focus on the context or the environment) (Matheson & Zanna, 1988). When people are publicly self aware, they are concerned about being evaluated by others and their status as a member of a group or their social identity.

If deindividuation and its accountability component are factors in online aversive behavior, methods that increase users' accountability for their behavior should also work to decrease aversive behavior. As conceptualized in online settings, a *reputation* is a sort of 'behavioral record' of one's actions, often within a particular context. For example, e-commerce reputation systems (e.g., Ebay, Amazon, Half.com) track the behavior of buyers and sellers; sellers who fill orders quickly and accurately are rated highly by the buyers who purchase from them. Information-based sites may calculate a user's reputation by the number of informative posts she makes, or the number and quality of her answers to others' questions. Users place a higher value on different types or pieces of reputation information, depending on the online context in which they expect to interact with the other person (Jensen, Davis & Farnham, 2002).

Reputation systems work because a reputation is a valuable social commodity, so if a person values her standing within an online community, he should behave in a way that is

appropriate and builds a positive reputation. Good and bad behavior is reflected in one's reputation, and this behavioral record holds one more accountable. An important component of reputation systems is that they provide users with feedback about their behavior from others who have had interactions with them, feedback that is often more direct than that found in typical face-to-face interactions. Reputations may also work because they focus users on the social norms associated with behavior in a particular online context and engage a user's public self awareness.¹

To test whether the ability to give and receive feedback can have an impact on behavior in a social context, we recently built and deployed a simple reputation system (Davis, Ma & Farnham, 2002). In an experimental user study, we randomly assigned users to one of twenty 'social,' email-based distribution lists, each comprising ten people. All groups discussed a range of topics chosen to induce interest and interaction (e.g., "Should men wear jewelry?"). During the two-week group interaction, half of the groups had access to a reputation system through which they could give other users anonymous feedback about their behavior within the list, while the other half did not. We were interested in whether the mere presence of a reputation system would affect both users' behavior within the groups and their perceptions of their own and others' behavior. In addition, we wanted to explore how people used the system itself. We found that both good and bad behavior occurred in this context, and the reputation system itself was used only moderately. In preliminary analyses, we have also found that the groups with access to the reputation system felt that there were fewer messages containing personal attacks and were more satisfied with the interactions than groups that did not have access to the system. It seems that even simple social feedback can affect at least users' perceptions of social interactions in online contexts.

One can imagine additional methods that capitalize on activating behavior-relevant norms to decrease bad behavior in an online context. For instance, asking users to 'join' a community to participate, whether the information they are asked to provide about themselves is detailed or not, may serve to make group membership salient and decrease aversive behavior in the community. Perhaps users could be asked to view a list of community members or some other representation of membership to make the norms for group behavior more salient. More thought is needed to create and test additional ways to identify and activate the critical norms in a particular context.

Additional Methods

Both personal profiles and reputations contain a great deal of information about individuals in online environments, and they can be useful for fostering positive interactions in many contexts. However, they are less useful in curtailing

¹ Note that personal profiles may also work to decrease deindividuation and therefore decrease aversive behavior.

aversive behavior that occurs more immediately (i.e., synchronously). For example, when a user misbehaves in a chat room, he is probably not concerned about the effects of his behavior on his reputation in the chat room. Users (and administrators) need ways to deal with such behavior in a quick, effective, viable manner.

My current interests focus on identifying and testing additional ways use group membership to decrease aversive behavior in online settings. For example, ostracism or 'shunning' is a powerful social force for behavior change. Current tools for 'muting' an offensive user in a chat room allow people to filter his or her content; however, they typically lack feedback to the offender about the action and the reasons for the ostracism. Visible ostracism by a group, particularly a valued group, can put considerable social pressure on the perpetrator of the bad behavior and lead to behavior change. One can imagine a user being given a 'time out' or put in a penalty box as a punishment for behaving inappropriately. The key element is feedback to the perpetrator: the fact that he was ostracized, the reason for the ostracism and the way he can change his behavior to return to the interaction.

Taking Targets Into Account

Mechanisms for dealing with inappropriate behavior in online settings should also take into account the reaction of the target of the bad behavior. The obvious and immediate goal for an intervention is stopping the perpetrator from repeating the aversive behavior, but a secondary goal should be to make the target of the behavior feel better about the situation and stay in the environment. In other words, it may be useful to allow users to 'give voice' to their frustration and anger at perpetrators, whether or not the response decreases the offending behavior. Users may get little satisfaction from 'muting' another person in a chat room, but they may feel better if they know the other user has received feedback about the action. As I mentioned, even when mechanisms for responding to bad behavior do not actually decrease it, they may have a palliative effect on the targets who are able to vent their frustrations, which may have been the case in the 'reputation system' study described earlier. This desirable outcome should not be ignored.

Conclusion

Aversive behavior in online settings is a growing concern, so it has become essential to identify and address its causes and cures. It is also important to recognize the importance of context—whether a behavior is considered aversive depends upon the online environment in which it occurs. Therefore, the fixes must fit the context. Social methods, because they are largely user-driven, are bound to be more effective long-term solutions than those that are laborious to administer or rely too heavily on technology.

References

- Connell, J. B., Mendelsohn, G. A., Robins, R. W., and Canny, J. (2001). Effect of communication medium on interpersonal perceptions: Don't hang up on the telephone yet! In *Proceedings of GROUP '01* (Boulder, CO, USA, Sept. 30-Oct. 3), ACM Press.
- Davis, J. P. (2002). The experience of 'bad' behavior in online social spaces: A survey of online users. *Working paper*.
- Davis, J. P., Ma, M. and Farnham, S. D. (2002). Using a peer rating system to improve online social behavior. *Paper submitted to CHI 2003*.
- Davis, J. P., Farnham, S. D. and Jensen, C. (2002). Decreasing online 'bad' behavior. In *Extended Abstracts of CHI '02* (Minneapolis, MN, April 2002), ACM Press.
- Jensen, C., Davis, J. P. and Farnham, S. (2002). Finding others online: Reputation systems for online social spaces. In *Proceedings of CHI '02* (Minneapolis, MN, April 2002), ACM Press.
- Jensen, C., Farnham, S., Drucker, S. and Kollock, P. (2000) The effect of communication modality on cooperation in online environments. In *Proceedings of CHI '00* (The Hague, Netherlands, March 2000), ACM Press.
- Matheson, K. and Zanna, M.P. (1990). Computer-mediated communications: The focus is on me. Special Issue: Computing: Social and policy issues, *Social Science Computer Review*, 8, p. 1-12.
- Miller, R. S. (2001). Breaches of propriety. In Robin Kowalski (Ed.) *Behaving Badly: Aversive Behaviors in Interpersonal Relationships*. Washington, D.C.: American Psychological Association.
- Short, J., Williams, E. and Christy, B. (1976). *The social psychology of telecommunications*. Chichester: Wiley.
- Swinth, K., Farnham, S. D. and Davis, J. P. (2002). Sharing personal information in online community profiles. *Paper submitted to CHI 2003*.