

The EVOLUTION of TROLLING

Understanding the scientific underpinnings of toxic online behavior.

By CYNTHIA ELLER



What causes people to display more toxic behavior when interacting with others online rather than in person? Caltech's Dean Mobbs, professor of cognitive neuroscience, along with postdocs Swati Oandity and Ketika Garg, and former research assistant Jianjin Zhang, have constructed a theoretical model to answer that question.

Mobbs, who is also the director and Allen V. C. Davis and Lenabelle Davis Leadership Chair of the Caltech Brain Imaging Center, and an affiliated faculty member of the Tianqiao and Chrissy Chen Institute for Neuroscience at Caltech, calls the model he and his team created the DAD framework (Disembodiment, lack of Accountability, and Disinhibition). The factors that comprise the framework's name make it more likely that social media interactions skew toward a level of nastiness and misinformation not seen in face-to-face communication.

First, disembodiment: In most social media exchanges, people have no direct sensory experience of one another. "When I'm speaking to somebody online, the conversation I'm having is all in my head; it's completely disembodied," Mobbs explains. Without cues from another's facial expressions or body position, these interactions, though they may feel external, are transacted entirely within one's internal world. It is here, Mobbs says, where the problem lies. "Your internal world is your playground not just for ideas or for what you want to say, but sometimes for things you shouldn't say. When you are communicating in this disembodied state, having a conversation in your mind, you begin to forget that you're having a conversation with a real person."

Lack of accountability occurs when social and cultural norms that operate in person to provide checks on interpersonal communication are absent and, with anonymity, even the fear of criminal punishment is gone. "I can say something via text on social media, and, often, I don't have to pay the consequences of saying it," Mobbs says.

"I don't get social disapproval at the same level I would as if I were with someone in person, and, frequently, I either am or believe myself to be completely anonymous."

The first two factors lead to the third: disinhibition. "It is disinhibition that allows you to say whatever you think, whatever you want," Mobbs says. "All of those nasty thoughts that you have in your head can just come out of your fingertips without interference."

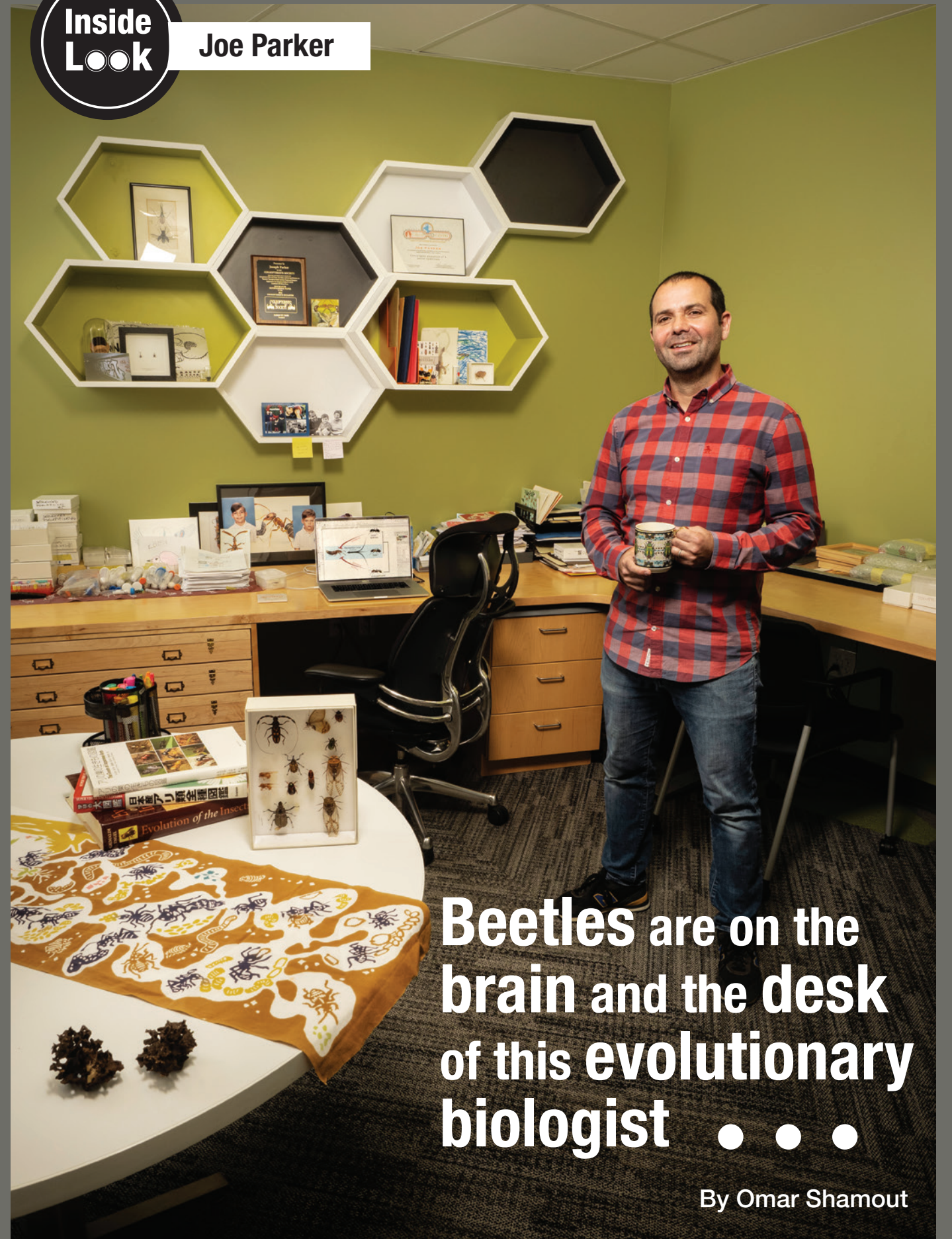
The DAD framework is grounded in an evolutionary perspective on emotion that is at the core of Mobbs's research. "Other animals have different strategies, such as camouflage or enhanced senses that detect threats, but we typically rely on avoiding predators before actually encountering them. We have not evolved for a social media environment," Mobbs says. "The sensory systems and theory of mind systems we have evolved in previous millennia do not translate well into an online domain," which leads to the "impaired interactions" that create online toxicity.

Mobbs and his co-authors suggest accountability can be strengthened by forcing users to register their social media accounts under their legal names, by slowing down the rate of interactions, or by introducing AI content moderators to provide more time for users to think about the consequences of their actions. Even the simple use of emoticons and avatars can help to mitigate the effects of disembodiment on our online behavior as they help to make others appear more real and their feelings more apparent.

The paper, "Three roots of online toxicity: disembodiment, accountability, and disinhibition," appears in the September 2024 issue of *Trends in Cognitive Sciences*. The work was funded by a gift from Sonja and William Davidow.

Inside
Look

Joe Parker



Beetles are on the brain and the desk of this evolutionary biologist ● ● ●

By Omar Shamout