

Moral Tribes on Wikipedia: the Mental Representations of Social Norms

Bradi Heaberlin^{1,2} and Simon DeDeo^{1,3,4}

¹Cognitive Science Program, Indiana University

²Department of Political Science, Indiana University

³Santa Fe Institute

⁴School of Informatics and Computing, Indiana University

June 20, 2016

Abstract

In a complex social system, behavior is often prescribed and regulated by hundreds, or even thousands, of interacting norms. Yet it is individuals that must learn norms—through interaction, exploration heuristics, and local peer influence—and individuals that decide when and how to use them. Little is known about this crucial process. How do individuals selectively, and collectively, navigate and utilize the norms of their social system, and what impact does this process have on that system’s development? We provide new answers to this question by tracking how people use and invoke norms in a real world social system: the English-language Wikipedia. As a community-managed knowledge commons, Wikipedia relies on its norms for governance; they form shared expectations for content creation and conflict resolution. Norms are invoked both directly and indirectly in discussions and arguments on the encyclopedia’s “talk” pages. We sample approximately 11,000 Wikipedia editors and use the detailed, and dated, log of their talk page edits to track the norms each editor references over the course of 15 years. Combined with prior work on Wikipedia’s norm network, this allows us to infer the coarse-grained mental representations individuals use, to test mechanisms for norm learning, to determine how an individual’s invocation of norms is influenced by context, and to test the “moral tribes” hypothesis: that individuals cluster together in distinct regions of the larger norm network.