When is ignorance bliss? Disclosing true information and cascades of norm violation in networks

Michael Mäs [a],*, Karl-Dieter Opp [b,c]

[a] University of Groningen, The Netherlands
[b] University of Leipzig, Germany
[c] University of Washington, USA

ARTICLE INFO

ABSTRACT

It has been hypothesized that disclosing a population's true rate of norm violation increases norm-violating behavior. Withholding such information might, thus, prevent the attenuation of useful norms. Analyzing a classical threshold model with flexible thresholds, we show that disclosing the true rate of norm violation can spark cascades of norm violation but can also have the opposite effect, decreasing norm violation and strengthening norm acceptance. The direction of the cascade depends on the initial rate of norm violation. Furthermore, the disclosure effect depends on whether or not the rate of norm violation is disclosed repeatedly, the structure of the social network, and whether individuals' norm acceptance is inelastic or open to peer-influence.

False beliefs appear to be undesirable. For instance, false beliefs about a bank's allegedly limited solvency can lead to a bank run and induce the destabilization of the bank. On the other hand, generating false beliefs can be socially beneficial and morally acceptable. For example, a physician may downplay the severity of a patient's illness if she expects the patient to commit suicide. Generating this false belief will certainly be regarded as morally acceptable. Likewise, classical philosophers discuss the legitimacy of so-called 'noble lies', famously arguing that lying to the people may be justified when it preserves social order (Allen, 2008; Hoffmann et al., 2013).

In this paper we study the societal effects of specific kinds of false beliefs, viz. beliefs about the amount of norm violation in society. It happens time and again that we read news about unexpectedly frequent norm violations. An example is the large-scale sexual abuse of children by the clergy in 2009/2010, which was kept secret by church officials and took outsiders by surprise. Another example is the surprisingly high extent of plagiarism in German academia. After it became evident that Karl-Theodor zu Guttenberg, Germany's former minister of defense, faked his dissertation, numerous further instances of plagiarism were publicized. It was certainly unexpected, at least by non-academics, that there is such

a large extent of cheating among scholars. These examples have in common that many individuals were not aware of the high degree of norm violation and were surprised when credible information was published.

What are the consequences of disclosing the true rate of norm violation? Here, we study an answer to this question by Heinrich Popitz (1925–2002), who hypothesized that there is a *preventive effect of ignorance* (Popitz, 1968, 2006).[1] The underlying theoretical argument is simple. He argued that disclosing a population's true rate of unexpectedly frequent norm violation leads to more norm violation, and lower norm acceptance. Having learned that there is more norm violation than expected, Popitz explains, norm complying individuals will decide to violate the norm. Furthermore, realizing that many members of the society violate the norm, individuals will decrease their norm acceptance, the extent to which complying with a norm is intrinsically valuable to an individual. This will aggregate to a collective decay of the norm. Popitz concluded that withholding information about the degree of norm violation in a population is socially desirable, as it prevents an increase in norm violation and preserves useful norms. There is

---

* Corresponding author at: Department of Sociology/ICS, University of Groningen, Grote Rozenstraat 31, 9712 TG, The Netherlands.
 E-mail address: m.maes@rug.nl (M. Mäs).

---

[1] The small book from 1968 is part of a larger book from 2006. Quotations refer to the book from 2006. The German expression for "preventive effect of ignorance" (which is our translation) is "Präventivwirkung des Nichtwissens". Recent publication that address Popitz's theory are Diekmann et al. (2015), Groeber and Rauhut (2009), Rauhut (2013), and Opp (2011).

thus a "preventive effect of ignorance," as Popitz puts it. Supporting Popitz's theoretical argument, empirical studies found that norm violation increases when individuals observe deviant behavior of others (Keizer et al., 2008; Keizer et al., 2013; Keuschnigg and Wolbring, 2015; Thaler and Sunstein, 2009).

Popitz's theoretical argument can be criticized for focusing only on those individuals who complied with the norm prior to the disclosure. If the true rate of norm violation is unexpectedly high, however, then there must be individuals who violated the norm before disclosure. How will these norm violators react to the disclosed information? According to Popitz's own theoretical argument, these individuals might have violated the norm because they overestimated the rate of norm violation. Having learned that the true rate of norm violation is lower than they expected, some norm violators will refrain from deviating, an outcome that was also documented in empirical research (Cabinet-Office, 2012; Rauhut, 2013; Schultz et al., 2007; Thaler and Sunstein, 2009). In a recent field experiment (Cabinet-Office, 2012), for instance, official letters were sent to taxpayers calling on them to pay any tax debt. Some of the letters contained the true information that 9 out of 10 citizens pay their tax on time. This information increased the response rate from 67.5% in the control condition to up to 83%. Thus, disclosed information about deviance increased compliance and not norm violation, which is the exact opposite of what Popitz predicted.

This criticism suggests that, even though individuals might react to the disclosed information, the aggregate effects of disclosing the true rate of norm violation might be much weaker than Popitz predicted. In fact, recent empirical research supports this. A laboratory experiment conducted by Rauhut (2013) found strong individual reactions after the disclosure but no macro-effect. Rauhut found decreasing rates of lying amongst dishonest participants after they had been informed about the true rate of lying in the population. Amongst the honest participants, however, lying increased to a similar extent, such that lying remained unchanged on the aggregate level.

Here, we identify so far overlooked, structural conditions under which disclosing the true rate of norm violation leads to substantial changes in collective norm violation, even though individuals might react in opposite ways to the disclosed information. Contrary to the existing literature, which focused on disclosure effects on isolated individuals (Cabinet-Office, 2012; Schultz et al., 2007; Thaler and Sunstein, 2009), we study individuals integrated in networks and analyze the dynamics that emerge on the collective level. Existing research showed, for instance, that individuals react stronger when the disclosed information is accompanied by injunctive messages (Schultz et al., 2007; Thaler and Sunstein, 2009). These findings are interesting and important, but they focus on the reactions of isolated individuals, which neglects that actors who adjusted their behavior after having received information about norm violation might motivate their network contacts, who may not have reacted when the information was disclosed, to also adjust their behavior. These adjustments, in turn, might motivate third actors to reconsider their choices, and so on. Assume, as an example, a population of dog owners who have been informed about the true amount of dog fouling in their town. It is possible that most dog owners accurately estimated the rate of norm violation and, therefore, do not change their behavior. However, there might be a neighborhood where some dog owners had underestimated the rate of norm violation and now decide to no longer clean up behind their dogs. Observing the additional dog mess, also their neighbors might change their behavior, which in turn might motivate others to also violate the norm. Such *behavioral cascades* can profoundly change collective behavior over and above the disclosure effect on isolated individuals that existing studies documented.

Our aim is to develop new and testable hypotheses about the structural conditions under which disclosing the true rate of

norm violation sparks a cascade that aggregates to increased or decreased collective norm violation. To this end, we take as given that individuals adjust their behavior when they are informed about the true rate of norm violation and develop testable hypotheses about the conditions under which cascades are sparked off and norm violation spreads in the network. We study three structural factors, analyzing how changes in aggregate norm violation depend on (i) the degree of clustering in the network, (ii) the magnitude of the disclosed rate of norm violation, and on (iii) whether the true rate of norm violation is disclosed only once or repeatedly.

To this end, we analyze a probabilistic version of the classical threshold models (Granovetter, 1978; Meade and Islam, 2006; Oliver et al., 1985; Valente, 1996) which are standard in the study of diffusion in networks (Barash et al., 2012; Centola and Macy, 2007; Funk et al., 2010; Krassa, 1988; Meade and Islam, 2006; Rogers, 1995; Valente, 1996) and research on political mobilization (Kuran, 1989; Macy, 1991; Oliver et al., 1985; Schelling, 1978). In our model, the probability that an agent violates the norm depends on her beliefs about the behavior of others and her norm acceptance. The latter represents the degree to which complying is intrinsically valuable to the individual and is mathematically equivalent to thresholds in the classical threshold models. Like in the classical threshold models, our agents compare these thresholds with the behavior they observe in others. Before the disclosure, agents only observe the amount of norm violation amongst their network contacts. When the true rate of norm violation is disclosed, however, agents base their decisions on the rate of norm violation in the whole network. We show that this switch from local to global information can fundamentally change dynamics of norm violation and that the structure of the network influences the intensity of this disclosure effect.

Our model goes beyond standard threshold models in that we also study settings with *adaptive thresholds*. Inspired by Popitz's informal reasoning, we include the assumption that thresholds are open to peer influence (Festinger et al., 1950; Wood, 2000), assuming that norm acceptance is weakened when agents observe more norm violation than they expected. Likewise, norm acceptance is intensified when agents observe high degrees of norm abiding behavior in others. We demonstrate that allowing thresholds to adapt in this form has profound effects on model dynamics and predictions concerning the consequence of disclosing the true rate of norm violation.

The new model is very general in that it can be applied to all binary decision problems with so-called positive externalities (Lopez-Pintado and Watts, 2008), cases where actors' tendency to adopt a behavior is intensified when they observe it in others. Typical examples are cases where individuals seek to coordinate with their contacts. For instance, adopting a new communication technology (e.g. fax machine, Facebook) is more attractive when many use it. Also collective-good problems have positive externalities when the production function is convex. Soccer players, for instance, might increase own efforts in order to raise their team's chances of winning when other team members also do their best. Norm violating behavior has positive externalities when individuals believe that norm violation is being punished and infer the chances of punishment from the observed amount of norm violation.

The remainder of this paper is organized as follows. In the next section, we formulate four research questions about the structural conditions of disclosure effects and demonstrate that a formal analysis is needed to validly derive answers to these questions. Subsequently, we describe the formal model that we analyzed to answer the research question and present our analyses. Finally, we summarize our results in a set of new hypotheses, and point to future work.