

THE GREAT ILLUSION: IGNORANCE,  
INFORMATIONAL CASCADES, AND THE  
PERSISTENCE OF UNPOPULAR NORMS

1. INTRODUCTION

Norms of discrimination against women and blacks, norms of revenge still alive in some Mediterranean countries, and norms that everybody dislikes and tries to circumvent, such as the invisible norms of reciprocity that hold among the Iks studied by Turnbull, are all examples of unpopular and inefficient norms that often persist in spite of their being disliked as well as being obviously inefficient from a social or economic viewpoint. The world of business is not immune to this problem. In all those countries in which corruption is endemic, bribing public officials to get lucrative contracts is the norm, but it is often true that such norm is disliked by many, and that it may lead to highly inefficient social outcomes (Bicchieri and Rovelli, 1995).

From a functionalist viewpoint such norms are anomalous, since they do not seem to fulfill any beneficial role for society at large or even for the social groups involved in sustaining the norm. In many cases it would be possible to gain in efficiency by eliminating, say, norms of racial discrimination, in that it would be possible to increase the well-being of a racial minority without harming the rest of society. To social scientists who equate persistence with efficiency, the permanence of inefficient norms thus presents an anomaly. They rest their case on two claims: when a norm is inefficient, sooner or later this fact will become evident. And evidence of inefficiency will induce quick changes in the individual choices that sustain the norm. That is, no opportunity for social improvement remains unexploited for long. Unfortunately, all too often this is not the case, and this is not because people mistakenly believe inefficient norms to be good or efficient.

Should we conclude that individuals often choose irrationally? Is the businessman who freely decides to bribe a public officer though he condemns the practice plainly irrational? For example, evidence gathered through several prosecutions associated to the Italian “Bribesville” scandal shows that often the parties involved in corrupt deals were not endorsing the norms they nevertheless obeyed. It thus seems that a moralization campaign would not be the right weapon in the battle against corruption. Most people’s beliefs and attitudes need not be changed: an ‘ethical therapy’ would just be a waste of time. On the other hand, we believe that individuals are rational at least in the minimal sense of taking those actions that they believe best fulfill their goals. If someone conforms to a norm that she dislikes, there must be some good reason. If we find out what the reason is, and show that it is shared by most participants in that kind of social situation, we have a chance of enacting more effective public policies.

That rational choice may ensue in suboptimal and even disastrous social outcomes is a well known tenet of the collective action literature, exemplified by the somewhat simplistic story of the prisoner’s dilemma. What we are telling here is a different story of mismatch between individual choices and social outcomes. There are psychological factors and systematic cognitive biases that are crucial in shaping decisionmaking: we often hold irrational beliefs and aspirations, or simply have wrong beliefs about our social environment. A robust descriptive theory of choice cannot ignore the extensive empirical evidence gathered by social psychologists about how individuals interpret and model their social reality. Because of systematic biases in interpreting other people’s behavior, individuals may wrongly believe that a certain norm they personally condemn is widely supported in the population or group to which they belong. If they act on their wrong beliefs, and conform to what they take to be the majority’s position, their public behavior will provide further evidence for the validity of their beliefs. Illusory private deviance will be experienced as real as people perceive the norm to have universal support. Misperception of social reality can have dire consequences in terms of persistence of inefficient and unpopular norms.

Norms whose existence depends on a collective illusion can be fragile. Whenever the veil of collective misperception is lifted, such norms may suddenly collapse. In the last part of the paper we show with a simple model how a norm that most people dislike can be established, and under which conditions it will break down. The agents we model are subject to a familiar psychological bias: in certain situations, they have the tendency

to assess other people's reasons for action as different from their own, thus they are likely to draw wrong conclusions from observing others' behavior. Yet our agents are rational, in that they maximize expected utility and update their information using Bayes' rule. Agents may rationally choose to follow the behavior of other agents and ignore their private information (or preferences) because they (wrongly) infer that others' choices are based on information (or preferences) that dominate their own. In this case, we show that there may be quick convergence to an unpopular norm on the basis of very little information. The bright side is that diffusion of even very little new information can upset the established norm.

Since the scope of this kind of model is very general, we will not attempt a more specific application to any particular situation, such as for example the case of norms of corruption.<sup>1</sup> The message we want to convey is that it is possible and useful to model the dynamics of inefficient or unpopular norms as the product of rational choices made by individuals who misperceive their social environment. The policy conclusion one draws from our model is that it might take surprisingly modest public interventions – such as the release of public information – to effect major changes in collective behavior.

## 2. A FEW EXAMPLES AND A FAIRY TALE

Let us consider a few examples of fairly common behaviors that have been extensively studied by social psychologists. What they have in common is that the subjects involved manifest a marked tendency to draw wrong inferences from observing the behavior of others that – in a shared social situation – act exactly like them.

Most of us have been part of (or at least have witnessed) the following classroom dynamics. The teacher pauses during a difficult lecture to ask students if they have any questions. Silence follows. The baffled but outwardly imperturbable students try to figure out their classmates' reactions. Despite widespread confusion, no hands are raised. This feature of the dynamic is not startling, since students obviously fear asking stupid questions and embarrassing themselves in front of the classroom. What is surprising is that students infer from the silence of other students that these individuals grasp the material and that they alone are ignorant and confused (Miller and McFarland, 1987). Accidents or other emergency

situations are less common, but they offer a similarly puzzling picture. Bystanders to emergencies are initially afraid of embarrassing themselves by overreacting. They thus remain cool and poised while they try to figure out if there is cause for concern. Their own composure and inaction they interpret correctly. But they infer from the similar behavior of others that these individuals are genuinely unconcerned and that probably there is nothing to worry about. This reluctance to respond to an emergency when other people are present has been experimentally replicated. Latane and Darley (1968) performed a series of experiments in which they tested the hypothesis that, as the number of bystanders increases, the probability that anyone will help decreases.<sup>2</sup> They concluded that the presence of bystanders inhibits action, and that in such situations a 'diffusion of responsibility' effect is at work.

In all circumstances in which social information is available, inhibition seems to be due to an evaluation of the emergency inferred from others' behavior: observing the inaction of bystanders helps support the conclusion that the victim is not seriously hurt. This must have been the conclusion reached by the neighbors of Kitty Genovese the night she was murdered. Her screams and pleads for help notwithstanding, each of the 38 witnesses who observed the attack must have inferred from the apparent inaction of others that there was nothing to worry about. Instead of a case of generalized indifference, it was probably an example of a common bias in drawing inferences from other people's behavior.

Classroom and emergency situations have something in common: they are ambiguous, in that people usually lack the information and objective criteria to make a judgment. How difficult is the question? How sick is the man lying on the road? The reactions of others provide us with some cognitive clarity, with some information about our environment. What is odd is that most people never seem to suspect that the motive behind others' composure is akin to their own, thus in fact their behavior conveys no information. Other people's nonchalance is not interpreted as a posture: it is rather seen as a genuine outward sign of a superior grasp of the situation. Is it the need to orient ourselves in a complex and largely unknown reality that leads us to impute different motives to others? Or is it because we are among strangers, and thus feel that it would be gratuitous to credit them with our own motives? If so, we would expect this bias to disappear in all those contexts in which there is no cognitive ambiguity and, moreover, individuals interact with each other long enough to dispel the illusion that their motives and those of others are different.

Juvenile gangs, schools, prisons and churches are social environments in which individuals have numerous and protracted interactions with each other. Group members typically share in-group's values that range from the toughness and display of violent behavior common to juvenile gangs to the banning of alcohol, card games and extramarital sex preached by many religious groups. Yet numerous studies show that group members tend to assume that their peers endorse their subculture's values more strongly than they do themselves. Matza (1964) discovered that gang members – when privately interviewed – express considerable discomfort with their antisocial behavior. But since they did not express their criticism publicly, to their peers they appeared fully committed and comfortable with the group violent behavior. Wheeler (1961) and Kauffman (1981) found that prison guards had significantly more liberal private attitudes than those they attributed to their fellow guards. For example, Kauffman found that 78% of the guards approved of an officer defending an accused inmate in a disciplinary board hearing, but only 44% of those guards assumed that their view would be shared by other guards. Most of the school teachers interviewed by Packard and Willower (1972) believed that the majority of their colleagues supported norms enjoining strictness toward students, but actually only a small minority did support such norms. Even children are not immune from misperceiving other children's attitudes. Prentice and Miller (1996) report the results of several studies of gender stereotyping among 3rd and 4th graders; there is ample evidence that children estimate other children's beliefs to be more sex-typed than their own.

All of the above examples have a common feature: individuals systematically underestimate the similarity of their attitudes to those of their peers. This fact, per se, does not amount to misperception. If the individuals in question had no way to observe others' behavior and to constantly compare their own actions to those of others, we might just conclude that they have a tendency to think of themselves as more liberal, sympathetic, humane or whatever than their fellows. In all these studies, however, individuals had plenty of occasions to observe each other in action. The problem is that they typically acted in ways that did not correspond to their private preferences or beliefs, taking public positions that were in line with what they believed to be the majority stance, the norms or values shared by their group. If individuals consciously dissemble, why don't they recognize a similar gap between on-stage and off-stage behavior in their peers? Why take others' behavior at face value?

What is common to students, bystanders to an emergency, prison guards and gang members is that – like most individuals – they are sensitive to the opinions and judgment of those around them and fear expressing views that would put them at a disadvantage, either because they show their ignorance or because they diverge from the perceived public opinion. Furthermore, individuals estimate public opinion on the basis of observable indicators, which in the preceding examples consist of the public behavior of group members. Observability may not be that direct: sometimes media reports will do, and sometimes a few active and vocal individuals suffice to create the illusion that they represent the majority opinion.

In his account of the decline of the French church in mid-eighteenth century, Tocqueville gave a striking example of how people can both publicly misrepresent their private beliefs and assume that the public behavior of others corresponds to their private beliefs. In *L'Ancien Régime*, he maintained that “Those who retained their beliefs in the doctrines of the Church because of being alone in their allegiance and, dreading isolation more than error, professed to share the sentiments of the majority. So what was in reality the opinion of only a part ... of the nation came to be regarded as the will of all and for this reason seemed irresistible, even to those who had given it this false appearance.” (p. 155)

Alexis de Tocqueville and Hans Christian Andersen are an improbable pair. What unites the historian and the novelist is a keen understanding of human psychology and of its social consequences. The covert French Catholics remind of the courtiers in “The Emperor’s new clothes”. Here two impostors manage to convince a rather dull emperor that they have made splendid clothes for him, so beautiful and refined that only smart and discriminating men can see and appreciate them. The unfolding of the tale is well known: until a child shouts “The Emperor is naked!” nobody has the courage to admit that he sees no clothes, and everybody takes others’ admiring murmurs as genuine expressions of superior judgment and refinement. In both Andersen’s tale and Tocqueville’s account there appears to be a perverse sequence that begins with a vocal minority creating the illusion that they are a majority. The members of the silent majority, thinking they are a minority (or even unique), assume that their dissembling peers are acting out of authentic convictions. The illusion of personal deviance persists because everyone misrepresents the conforming behavior of a majority that fears ostracism and ridicule.

### 3. PLURALISTIC IGNORANCE

The individuals in the previous examples are experiencing what social psychologists call *pluralistic ignorance*, a psychological state characterized by the belief that one's private thoughts, attitudes and feelings are different from those of others, even though one's public behavior is identical (Allport, 1924; Miller and McFarland, 1991). Perhaps the term 'ignorance' is not the most appropriate, as the individuals concerned seem to make systematic mistakes in judging the motives, and hence the attitudes and beliefs, of other people. Their judgments are guided by what they observe and, indeed, observability is always a feature of the contexts in which pluralistic ignorance arises. The problem with such judgments is that individuals wrongly infer that – unlike themselves – others must be thinking and feeling the way they are acting. The question naturally arises whether there is anything special or peculiar about the circumstances in which pluralistic ignorance occurs. Are there some contexts in which it is easier for people to interpret the similar behavior of self and others differently?

One feature common to all victims of pluralistic ignorance is that they engage in social comparison. This fact, per se, is not remarkable, given the circumstances in which they find themselves. One of the functions of social comparison is to provide us with information about a new situation, especially when we lack sufficient information or objective criteria to make a judgment, or the environment we face is ambiguous, and open to several possible interpretations. Thus a bystander to an accident will look at fellow bystanders to gauge information about the seriousness of the casualty, and a student will try to guess from the classmates' expressions if the question is a difficult one. Social comparison also helps in self-evaluation, as when we try to assess our abilities, or the goodness of our opinions. Finally, and most important for our topic, social comparison helps to establish one's standing as a member of a valued group. In a study of alcohol use among Princeton students, Prentice and Miller (1996) noted that excess drinking is "central to the social identity of many college students and is an important part of social life in most campuses" (p. 9). Looking at the drinking habits of their peers, students quickly infer that heavy drinking is a campus norm, and try to adapt to what they perceive as a social trait essential to being identified as 'one of them'.

Another feature common to all contexts in which pluralistic ignorance occurs is the lack of transparent communication (or of any communication

at all) among individuals. This happens when people take a public stance that does not correspond to their private attitudes or beliefs. Bystanders to an emergency may feign lack of concern, and prison guards may behave less sympathetically to inmates than they would if they were to follow their inclinations. Because there is a gap between public behavior and private attitudes or beliefs, and no way to assess other people's attitudes other than observing their overt behavior, the public expressions of others are erroneously perceived to be genuine representations of their private thoughts. A good example of how damaging lack of communication can be is Schanck's 1932 study of the social dynamics of members of the Baptist and Methodist churches in the fictitiously named community of Elm Hollow. He found out that church members supported religious values more strongly in public than in private, but that the gap between the practiced and the preached was believed to be much smaller in the case of fellow church members than it was in their own case. The fact that those individuals could not have a frank discussion about proscribed activities such as card gaming was helpful in creating the illusion of uniform compliance with church teachings. On the other hand, precisely because of the perceived conformity of church members, initiating such a discussion was feared to put one at risk of being ostracized by the community. Lack of truthful communication became self-perpetuating.

Finally, pluralistic ignorance depends upon misinterpreting the similar behavior of similar others. In Tocqueville's example, the covert catholic was not comparing his position with that of the vociferous and very visible anti-religious minority. He was making comparisons with his likes, people who used to practice Catholicism and could not be suspected of revolutionary sympathies. It is their conforming to the opinions of what is to all effects a minority that is interpreted at face value. In this way, Tocqueville argued, the sentiments of a small part of the nation were mistaken as the public opinion, which became irresistible even in the eyes of those who thought and felt otherwise.

#### 4. WHY MISREPRESENTATION?

Though for our purpose it is not important to dwell on the causes of pluralistic ignorance, we shall briefly examine some possible explanations of why we so often fall prey to what might be aptly called an "illusion of transparency", the systematic misrepresentation of other people's attitudes



and beliefs. If the student is motivated to keep silent for fear of embarrassing himself, and the gang member conforms out of a desire to fit in the group, why do they have such difficulty in recognizing the same motives in other people? The motives that drive students and gang members alike are social in nature: fear of embarrassment and the desire to identify with a valued group could not exist in the absence of a reference group whose judgment one cares about. It would thus seem that people perceive social motives as more potent causes of their own than of others' behavior. It is not an inability to attribute motives or reasons for action to others which drives pluralistic ignorance: rather, it is a self/other difference in the perceived power of social motives.

It has been suggested that one possible reason for this attribution bias is that motives such as fear of embarrassment, or the desire to belong and fit in, are mainly defined by internal and unobservable cues (Miller and McFarland, 1987). Whereas emotions such as love, hatred, pride or contempt are easily observable, there are others that are almost by definition much more concealed. An individual can thus easily come to believe that she experiences such emotions more strongly than others do. This view is supported by experimental evidence that individuals tend to rate themselves as more extreme than the average person on traits pertaining to social inhibition and other states mainly defined by internal cues (McFarland and Miller, 1990). It would therefore be the very unobservability of others' motives that would lead to the self/other discrepancy.

Another possible reason for the disparity is that we are subject to a cultural propensity to underestimate the power of social motives to influence behavior (Miller and Prentice, 1994). We usually ground our assessment of our own motives on past experiences which usually provides evidence that we act to maintain our social standing, that we want to avoid embarrassment, ostracism, reproach and so on. However, we base inferences about others' motives on shared cultural representations of the relative power of different motives. Such representations induce us to overestimate the extent to which others are acting on 'private' motives, and to minimize the extent to which people are acting to maintain relations with their peers, to establish and retain a social identity, or to avoid feelings of shame.

There is some merit to this interpretation, in that it fits with other studies that report the importance of cultural biases in assessing even *our own* reasons for action. In a comprehensive study of altruistic behavior in

America (Wuthnow, 1991) – mainly expressed through voluntary caring activities such as taking meals to the elderly, visiting the sick, donating time to nursing homes and hospitals, and staffing hotlines and crisis intervention centers – it turns out that the majority of volunteers rationalize their compassionate behavior as self-interested. Altruism and compassion have to be redefined within the framework of a culture that casts a tremendous weight on self-interest in explaining behavior. Thus a volunteer in a rescue squad felt compelled to justify his seemingly self-sacrificial activities by reference to ‘perks’ such as “being able to drive as fast as I want on the highways” and “being able to park wherever I want to”, hardly sensible reasons to undergo the costs and risks of such activity.

Though we do not want to deny the importance of cultural biases in assessing other people’s motives, it seems that the very nature of motives such as fear of embarrassment, rejection and ostracism presupposes a self/other disparity. To exist at all, one’s fear of rejection presupposes the existence of someone that has the power and willingness to sanction one’s deviant behavior. I want to conform because – among other things – I dislike the social consequences of transgressing; would I be as uncomfortable at the prospect of breaking the rules if I knew that everyone else complies for the same reasons I do? Those who are expected to punish cannot be presumed to share the same motives for compliance, otherwise the very threat of a sanction would become void. To recognize other people’s fears would render the norm as insubstantial and volatile as the Emperor’s new clothes. Thus fears of social sanctions, to be justified, need to be supported by the belief that others are committed to the norm for very different reasons.

All the above explanations of motivational biases seem to presuppose that individuals are able to recognize the true causes of their own behavior. Experimental evidence, however, only supports the claim that individuals consistently report acting in accordance with social motives, whereas they tell different stories when asked to explain other people’s behavior. All we can infer is that most of us act like poor scientists, in that we discount an important piece of evidence (our own alleged motives) without apparent good reasons. This is particularly true since the victims of pluralistic ignorance are not just observers: they are participants in the group dynamics who know that their own behavior belies their internal state and cannot be taken at face value. We cannot (and should not) infer from the available experimental data that individuals have a privileged access to the true causes of their behavior. What matters to pluralistic

ignorance is that the purported cause of one's behavior is not deemed to be sufficient to produce similar behavior in others. Even if we just grant that individuals know how they feel, and infer from others' observable behavior that they feel differently, this simple inference is sufficient to generate pluralistic ignorance.

## 5. WHY DO PEOPLE CONFORM?

The examples of pluralistic ignorance one finds in the literature refer to two major classes of situations in which pluralistic ignorance (PI) occurs. One is when PI yields the illusory belief that others hold the group's values more strongly than does oneself (schools, churches, prisons, etc.). We know from studies of students' drinking habits and sexual stereotyping among young children that it does not matter whether the group is transient or well-established: PI will arise in the presence of both. Another class of situations where PI occurs is when individuals correctly identify the positive value a group places upon a characteristic, but fail to realize that other group members are just pretending to have these characteristics, like 'keeping cool' in emergencies or not looking stupid in the classroom. What is underestimated here is others' strength of motivation to avoid acting inconsistently with this value.

Given that people seem to make such huge mistakes in assessing others' motives, it remains to be explained why they choose not to raise hands, intervene in emergencies, or otherwise behave in accordance with their true preferences. That is, even if we can tell a plausible story about how PI comes about, we have not yet explained how and why beliefs about others' motives can have prescriptive power. Why would a perceived self/other difference push one to conform to what is taken to be a collective value or norm? It seems that two conditions are necessary to produce conformity to the perceived value or norm: one is that the people observed must serve as a reference group, the second is that they are seen as uniform in their opinion.

In all those cases in which others' behavior is just taken as an indicator of a given state, people will simply choose an action that best fits their motives and information. In the case of the students, or of bystanders to an emergency, other students or bystanders serve as a reference group in that one gathers information from their behavior about some piece of reality (the lecture's difficulty, the severity of the emergency), and one

wants to avoid acting stupidly in public. Behavior may however just reflect the belief, as in emergencies, that there is nothing to worry about.

Juvenile gangs and other cohesive social groups (like churches and schools) offer a more interesting perspective. Here individuals do not simply try to gather information or not to embarrass themselves. Rather they want to behave in accordance with what they perceive to be the group norms. The gang or the church are valuable social groups, and individuals strive to be accepted as good-standing group members. It thus seems that group-identification lies at the root of many cases of pluralistic ignorance. Individuals express different views than they actually hold and act in ways they privately disapprove of because they believe those views and behavior to be consensual within a valued group.

We realize that 'social identity' is an elusive concept. Here we use it in a very limited and circumscribed sense to refer, in Tajfel's own words, to "that part of an individual's self-concept which derives from his knowledge of his membership of a social group (or groups) together with the value and emotional significance attached to that membership" (Tajfel, 1981, p. 255). Our preoccupation is with the effects on group behavior of the significance granted by individuals to group membership. There are many groups one simultaneously belongs to, and some of these memberships are more salient than others, whereas some may vary in importance with time due to changes in social and individual circumstances.

A crucial feature of the concept of social identity as we use it here is that identification with a group is in some sense a conscious choice: one may accidentally belong to a group, but it is only when being a group member becomes at least partly constitutive of who one is that we can meaningfully talk of social identifications. Being born in Northern Italy, up to a few years ago, was a mere geographical accident. With the advent of the Lombard League, it has become a reason of pride and distinction for many. That social identity considerations may motivate behavior is less contentious than the reasons why this happens. Identifying with a particular ethnic or geographical group, for example, might hold the promise of future tangible rewards; the small industrial and commercial businesses that are the political bedrock of the Lombard League stand to gain from a program of at least partial fiscal and financial separation from the rest of the nation. Membership in the League can thus be seen as a rational choice, strictly motivated by self-interested considerations.

At other times, however, group memberships' benefits are more elusive: Ash's experiments on conformity and Tajfel's study of "minimal groups"

suggest that social identity effects may occur even in the absence of the tangible or intangible rewards that membership in an established group affords. There is a crucial difference between motives derived from personal self-interest and those derived from concern for the interests and outcomes of others. Identification with a valued group can stem from individual or collective welfare considerations: one may want to belong to a group because of the prospect of future personal rewards, or just because one values the group and takes the group's goals and interests as one's own, even at the cost of overlooking or restricting individual gains. Be it as it may, we shall keep using social identity as a motivational factor, even in those cases in which it can apparently be further decomposed into self-interested motives.

## 6. THE CONSEQUENCES OF PLURALISTIC IGNORANCE

Our goal has been that of examining some of the social consequences of PI. In particular, we want to claim that PI plays a role in the perpetuation of unpopular and inefficient social norms. PI explains how individuals might wrongly believe that a certain norm, attitude or belief they personally condemn is widely held among a population or group to which they belong. For example, several studies done in the 60s and 70s uncovered a marked tendency for white Americans to overestimate private white support for forced racial segregation. In fact, only 18% of those polled favored segregation, but 47% believed that most did so (O'Gorman, 1975). If the overestimators acted according to the perceived majority opinion, a racist norm might have survived in spite of being privately endorsed only by a small minority.

When people act on their wrong beliefs, and conform to what they take to be the majority's position, their public behavior will provide further evidence for the validity of their beliefs. Illusory private deviance will be experienced as real as people perceive the norm, attitude or belief to have universal support. In this case, if individuals come to embrace the norm or attitude they originally erroneously attributed to others, an initial wrong perception will become accurate at both private and public levels.

Embracing a norm that is perceived as widely supported is only one of the possible ways of reducing the self/other discrepancy. To use the wording of Hirschman (1970), *exit*, *voice* and *loyalty* are all potential means to solve the conflict raised by pluralistic ignorance. Loyalty

involves changing one's attitude towards the perceived norm: a person who adopts this strategy will eventually come to internalize the norm. The Princeton freshman and the third grader can ultimately internalize drinking norms or sexual stereotyping because their private attitudes toward the norm are presumably not well established. If the perceived norm does not clash with preexisting values and beliefs, it is reasonable to assume that – after a more or less lengthy period of time – what was wrongly perceived as a norm endorsed by the majority will in fact become the majority's norm. In this case, pluralistic ignorance will disappear. To say that a norm is internalized does not necessarily mean that it becomes part of one's deepest system of values. We use internalization in a much simpler sense: an internalized norm is an accepted norm, a norm that one is prepared to defend and rationalize as having positive value.

Exit involves rejection of the group that upholds the norm, and voice entails an attempt to bring the norm closer to one's attitude. Yet exit is not an easy option and, in many situations, it is not an option at all. Voice, too, is a difficult and costly choice, since it involves an attempt to change the shared norm. Exit and voice presuppose that private preferences and attitudes are well established. In this case, individuals dislike the perceived norm, and the question is whether it is feasible or it pays to leave the group or express different preferences.

Take the case of corruption. In a system where corruption is the norm, there may be many reasons why denouncing corrupt transactions is a costly option. Typically in such systems the prices of public contracts are much higher than they would be in a non corrupt system, therefore many firms have at least a short-term incentive to keep the system in place. Those who recognize its inefficiency may prefer to be honest, but expect everyone else to be dishonest. The firm who decides not to bribe or even to denounce corrupt practices can thus expect to be an isolated case that will be excluded from future, lucrative interactions (Bicchieri and Rovelli, 1995). For most firms, moving to another country or denouncing the system are just not feasible or reasonable options.

Turnbull, in his book about the Iks of Uganda, reports that they went to great lengths to avoid being caught in a situation that dictated reciprocity, or even accepting help from another person with the intent of generating a future debt. Yet norms of reciprocity were upheld, even if the situation of extreme hardship in which the Iks were living had made them ineffective, if not harmful. The very fact that everyone tried to avoid situations where such norms would have normally applied testified to their resil-

ience, as well as to their being privately disliked by all. The norms' resilience was apparently due to the widespread belief that they enjoyed universal support (Turnbull, 1972).

In both examples, exit, voice or loyalty are not options. Individuals disapprove of the norm, but refrain from open dissension because they interpret others' behavior as signaling support. The unpopular norm is an *equilibrium*, in the sense that no individual has an incentive to try to influence other people or just reveal her true preferences by not conforming to behavior that she perceives as universally endorsed (Bicchieri, 1990). The equilibrium is self-perpetuating, since the belief that the norm is universally endorsed generates widespread conformity, and observation of conformist behavior further confirms the expectation of universal endorsement. The equilibrium however is also fragile, because small shocks are sufficient to generate large shifts in behavior. If the huge, silent majority of 'dissenters' become aware that their private preferences are shared by many, they would presumably shift their behavior in the direction of their true preferences. The prescriptive force of a norm is derived by its perceived universality: if people come to recognize that support for a given norm is limited or wavering, its power to induce conformity will be greatly reduced or even nullified.

When assimilation or rejection mechanisms fail, the veil of pluralistic ignorance can be lifted through an external intervention or by an endogenous mechanism. Misperception about a norm's endorsement could be eliminated by the government or media agencies releasing public information about people's true preferences or opinions, or about some new facts that would make it easier for those who dislike the norm to openly dissent. Smoking in public places, gender-biased language and strict sexual mores are all examples of norms that changed very quickly in response to the diffusion of public information about, say, the availability of contraception or the fact that many women consider offensive the exclusive use of male pronouns in all kind of prose. What matters is not the quantity of information released (it may take surprisingly little) but its quality, in the sense that the source must be credible. This condition is important in all cases of pluralistic ignorance, not just those that effect the upholding of unpopular norms. In Andersen's tale, it takes a child – who is innocent and truthful – to make the emperor's nakedness common knowledge. Similarly in a silent classroom the first question will usually generate a cascade of further questions, since it indicates to everyone that indeed the lecture was a difficult one.

Alternatively, it may just take a few vocal deviants to generate a major shift in public opinion. All revolutions were initiated by small minorities: their visibility, as well as their ability to provide or at least present an alternative, gave voice to popular dissatisfaction; but a major political overturn would hardly have occurred had the majority of people not been already privately disappointed with the status quo. In this case, the endogenous mechanism exclusively relies on the existence of a group of individuals who for some reason refuse to conform to the established system. Though it may be true that different people have different degrees of preference for conformity, an endogenous explanation of how an established norm may suddenly break down should not be confined to the existence of a few unconventional characters. Deviant behavior is frequently the result of a momentary slip or even a mistake in processing relevant information. In the following section we show with a formal model how such actions may have a disproportionate effect, as they can lead to sudden and large shifts in collective behavior. Since our explanation of such shifts does not rely on the existence of a few non-conformists, the combination of pluralistic ignorance with the possibility of 'contravening a norm by mistake' make the collapse of unpopular norms much more likely than it would otherwise be.

## 7. INFORMATIONAL CASCADES

To model the effects pluralistic ignorance may have on the dynamics of unpopular norms, we need to model how people may rapidly converge on a common behavioral pattern on the basis of very little information, and how even a little new information – suggesting that a different course of action is optimal – may shift collective behavior in a direction opposite to the status quo. To do so, we want to show that pluralistic ignorance is likely to generate a so-called *informational cascade* (Bichchandani et al., 1992). Informational cascades occur when it is optimal for an individual – having observed the actions of other individuals – to follow their behavior regardless of his own preferences or information. Once an individual acts only on the information obtained from others' actions, his decision conveys no truthful information about his private information or preferences. Because the conformity of individuals in a cascade has no informational value, cascades are fragile and could be upset by the arrival of new (truthful) public information.



In a state of pluralistic ignorance, individuals have private information about their preferences and beliefs, but can only infer other people's preferences or beliefs from observing their choices. If they assume that other people's choices truthfully reveal their preferences, beliefs, or attitudes, they may find it rational to conform to patterns of behavior they privately dislike. To model the kind of situation in which the 'wrong' norms are likely to emerge, we have to modify the assumption of sequential choices made by informational cascades models. In our model choices are simultaneous; moreover, we make an explicit assumption of pluralistic ignorance. The interesting conclusion we draw is that it takes very little to reverse a cascade: even in a population almost entirely made of conformists, a few transgressions (wrongly interpreted as revealing true preferences) may induce a sudden change of behavior in the direction of the true majority's preferences .

### *The Assumptions*

In order to avoid unnecessary complications, we assume that people choose their own actions by observing others' actions. We model binary choices, so agents choose one action in the set  $\{x_1 = 0, x_2 = 1\}$ . For example, they can choose to drink either beer or soda, or choose to bribe a public officer or behave honestly. Individuals have a common prior belief that a certain percentage of the population is 'deviant', but do not know the direction of deviance. That is, individuals have common priors on the distribution of the majority and minority. For example, if 10% of the population is 'wet' and 90% is 'dry', individuals believe that 90% of the population is in the majority and 10% is in the minority. However, they have no idea which trait distinguishes the majority (minority). Therefore, in the absence of any background information, they put the same prior probability (50%) on two possibilities: the majority (90%) is wet (and the minority (10%) is dry), or the majority is dry (and the minority is wet).<sup>3</sup>

Individuals have varying degrees of conformist preferences, but believe (the majority of) others have "normal" preferences, i.e., they really prefer what they choose. We are including here a crucial feature of pluralistic ignorance: individuals assess others' reasons for action as different from their own. Thus the fact that everyone is conforming is not common knowledge. The utility (loss) function of an individual is  $U = \{-(x_i - \hat{x})^2$

$-(\beta/2)*(x_i - y)^2\} - \partial$ , where  $y$  is the individual's privately preferred action,  $\hat{x}$  is the perceived majority preference (which takes the value 0 if  $x_1$  is believed to be the majority's preference, 1 if  $x_2$ , and 1/2 if there is uncertainty as to what the majority prefers),  $\partial$  is a discount factor (explained later) and  $\beta$  stands for a person's degree of non-conformism. To make the argument as simple as possible, we assume  $\beta$  can take only two values, 0 and 1. When  $\beta$  is 0, an individual conforms no matter what. The population contains a small (relative to the whole population) number of trendsetters, whose  $\beta$  is 1. They care about expressing their private preferences but do not want to be deviants. Their number ( $z$ ) is exogenously given. The number of individuals, excluding trendsetters, is  $N$ . Individuals are rational, in that they maximize expected utility, and follow a Bayesian decision pattern whenever the use of Bayes' rule conflicts with private information.

The only difference between conformists and trendsetters is as follows: if a conformist cannot infer which is the majority on the basis of observed behaviors and his preferences, he will choose his action by flipping a coin (since in this case the probability that the majority is, say, wet, is equal to the probability that it is dry), and the conformist wants to maximize the utility of conforming to the majority, whatever it happens to be. For example, if a privately dry conformist observes only one person drink beer, he has two conflicting pieces of information: one wet person and one dry person (himself). In this case, the dry conformist will drink either beer or soda equiprobably. In the presence of conflicting observations, the trendsetter will choose according to his true preferences. That is, if  $(x_1 - \hat{x}) = (x_2 - \hat{x}) = 1/2$ , a positive  $\beta$  will induce him to choose on the basis of the second term  $[-(\beta/2)*(x_i - y)^2]$  of his utility function. However, in a situation in which there is a well established norm and the trendsetter thus believes the majority to prefer, say,  $x_1$ , the trendsetter will choose  $x_1$  irrespective of his private preference, because  $\beta/2 = 1/2 < 1$ .

We are interested in modeling how an unpopular norm might emerge. In the absence of an established norm, how will an individual choose? We assume people can choose at either time 1 or time 2, and only these two time-points exist. The discount factor,  $\partial$ , takes a small value ( $\ll 1$ ) if the trendsetter chooses at time 2 and zero otherwise, in order to reflect the eagerness of the trendsetters to express themselves earlier, *ceteris paribus*.

At time 1, a norm has not yet been established, thus there is no information as to the majority's preference; at time 2 instead some information is present, but it might not be enough to decide that a norm is

now in place. At time 1 an individual has to decide whether to choose now or to defer his choice to time 2. He will thus compare the utility of choosing now ( $U_1$ ), to his present expectation of what the utility of choosing later would be ( $E_1[U_2]$ ). The conformist will always wait and see because, since his  $\beta = 0$ ,  $U_1 = -1/4 < E_1[U_2] = -\gamma/4$  (where  $\gamma < 1$  is the probability that information will be indecisive at time 2). Since his  $\beta = 1$ , what will the trendsetter do? If he acts at time 1,  $U_1 = -1/4$ , because he chooses according to his private preference. If he waits and acts at time 2,  $E_1[U_2] = \{-1/2*(1-\gamma)/2\} + (-1/4*\gamma) - \partial = -1/4 - \partial$ . Therefore,  $U_1 > E_1[U_2]$ , and he always acts at time 1. Thus in a new situation the kind of norm that gets established will depend on the trendsetters, who choose first and simultaneously. Their choices may coincide, in which case those who observe them will infer that the observed choices represent the majority's preference. In such case, all will conform. Suppose instead that the initial trendsetters' choices are different, and thus not decisive. In this case, a person whose  $\beta = 0$  will toss a coin, and act according to the result of the toss.

We define  $p$  as the size of the majority (expressed as a percentage). Depending on the situation studied, the magnitude of the majority will make an important difference. In corruption and elections, even a narrow majority (say, 51%) will matter, since everyone wants to bet on the winning horse. In the case of drinking norms, it might not be reasonable to assume that people fear being in the minority if  $p$  is close to 0.5.

*The Model*

We define  $\pi$  as the posterior odds of  $x_1$  being the majority preference to  $x_2$  being the majority preference. Let  $\theta_i$  denote that  $x_i$  is the majority preference;  $z_i$  denotes the action (either  $x_1$  or  $x_2$ ) of trendsetter  $i$ , and  $y$  is a private preference. For example,  $p(z_1 = x_1|\theta_1)$  is the probability that trendsetter 1 has chosen action  $x_1$  conditional on  $x_1$  being the majority preference. By assumption, the prior odds are 1 to 1, that is,  $p(\theta_1)/p(\theta_2) = 1$ . Then,

$$\begin{aligned} \pi &= p(\theta_1|y, z_1, z_2, \dots)/p(\theta_2|y, z_1, z_2, \dots) \\ &= [\prod p(z_i|\theta_1)/\prod p(z_i|\theta_2)] \cdot [p(y|\theta_1)/p(y|\theta_2)]. \end{aligned}$$

In logarithm,

$$\begin{aligned} \ln \pi &= \ln[p(\theta_1|y, z_1, z_2, \dots)/p(\theta_2|y, z_1, z_2, \dots)] \\ &= \sum \ln[p(z_i|\theta_1)/p(z_i|\theta_2)] + \ln[p(y|\theta_1)/p(y|\theta_2)]. \end{aligned}$$

All but trendsetters want to wait and see at time 1, and to conform to the expected majority at time 2. Therefore, the conformist's decision rule is:

$$x = \begin{cases} \text{no action} & \text{time 1} \\ x_1 & \text{time 2 and } \pi > 1 \\ x_1 \text{ or } x_2 \text{ equiprobably} & \text{time 2 and } \pi = 1 \\ x_2 & \text{time 2 and } \pi < 1. \end{cases}$$

Trendsetters are assumed to have a relatively strong preference for expressing themselves. Therefore, they act at time 1 as shown above, while conformists defer their actions to time 2. Note that at time 1  $p = 1$ , because we assume the common prior is 1. Then,

$$x = \begin{cases} x_1 & \text{if the true preference is } x_1 \\ x_2 & \text{if the true preference is } x_2 \end{cases}$$

### Results

Let us assume that the true majority ( $pN$ ) prefers  $x_1$ . Suppose  $z = 1$  and his (the only trendsetter's) action is  $x_1$ . After observing the trendsetter, all conformists choose simultaneously. In this case,  $(1-p)N$  people toss a fair coin, whereas  $pN$  people choose  $x_1$ . Then, the distribution of  $x_1$  and  $x_2$  will arbitrarily approach  $[(1+p)/2]$  and  $[(1-p)/2]$ , respectively, for very large  $N$ . If the trendsetter's action is  $x_2$ , the distribution of  $x_1$  and  $x_2$  will approach  $(p/2)$  and  $[(2-p)/2]$ , respectively, for very large  $N$ . In either case, the actual distribution of actions is different from the distribution of true preferences, because those whose true preferences are different from that of the trendsetter flip a coin to choose an action. We may call this phenomenon a *partial cascade*.

Unless  $N$  is infinite, both positive and negative (complete) cascades can occur. However, even if  $N$  is small, the probability that either a positive

or a negative cascade occurs is very small. For example, if  $N = 9$ , the probability of a negative cascade conditional on the (single) observation of  $x_2$  is  $(1/2)^9 \approx 0.002$ .

If we have two trendsetters, there are three possibilities:  $\{x_1, x_1\}$ ,  $\{x_1, x_2\}$  and  $\{x_2, x_2\}$ . The collective behavior that follows will be, respectively:

$N$  choose  $x_1$  if  $\{x_1, x_1\}$ ;  $pN$  choose  $x_1$  and  $(1-p)N$  choose  $x_2$  if  $\{x_1, x_2\}$ ;  $N$  choose  $x_2$  if  $\{x_2, x_2\}$ .

Only in the second case,  $\{x_1, x_2\}$ , individuals will reveal their true preferences, hence  $p$  will express the true proportion of people who prefer (and choose)  $x_1$ .

Although the situation looks similar to that of a sequential setting, a significant difference exists: because the simultaneous setting needs neither a sequentially specified order, which the sequential setting needs, nor does it make possible coin flipping among trendsetters, the probability that no cascade occurs is higher in the simultaneous setting than in the sequential one. For example, if  $z = 4$ , the probability of no cascade in a simultaneous choice setting is:  ${}_4C_2 \cdot p^2 \cdot (1-p)^2 = 6p^2 \cdot (1-p)^2$  instead of  $p^2 \cdot (1-p)^2$  (which is the probability of no cascade if choices are sequential). However, if  $z$  is large (with respect to  $N$ ), the probability of no cascade in the simultaneous case cannot be distinguished from that of the sequential case.

### *Distribution of the trendsetters' tastes and simulation results*

So far, we have not specified what determines the distribution of tastes (and choices) among trendsetters. Since the trendsetters' tastes distribution is crucial in determining the likelihood of cascades, we now specify how it is determined and show some simulation results based on this specification. Let  $N$  be very large, and represent the whole population (including trendsetters). Suppose the distribution of the entire population's tastes is given:  $pN$  are dry and  $(1-p)N$  are wet. Without loss of generality, assume that  $p > 0.5 > 1-p$ . In our example, dry people belong to the majority and wet people to the minority. Nature randomly picks out of  $N$  a small number ( $z \ll N$ ) of trendsetters. The distribution of tastes among trendsetters thus becomes:

$$\begin{aligned}
 p(\text{number of wet} = 0) &= {}_z C_0 \cdot p^0 \cdot (1-p)^z \\
 p(\text{\# of wet} = 1) &= {}_z C_1 \cdot p^1 \cdot (1-p)^{z-1} \\
 \dots\dots\dots \\
 p(\text{\# of wet} = q) &= {}_z C_q \cdot p^q \cdot (1-p)^{z-q} \\
 \dots\dots\dots \\
 p(\text{\# of wet} = z) &= {}_z C_z \cdot p^z \cdot (1-p)^0
 \end{aligned}$$

If (# of wet - # of dry)  $\geq 2$ , negative cascades occur. If (# of dry - # of wet)  $\geq 2$ , positive cascades occur. In the case of an even number of trendsetters, it is possible that the number of wet is equal to the number of dry trendsetters, in which case cascades do not occur, though this possibility does not exist in the case of an odd number of trendsetters. On the other hand, it is possible that  $|\# \text{ of wet} - \# \text{ of dry}| = 1$ , in which case partial cascades would occur in the presence of an odd number of trendsetters, but not if the number of trendsetters is even. As can be easily inferred, the actual distribution of trendsetters' tastes depends upon the number of trendsetters, as well as upon the distribution of the entire population's tastes. Table 1 shows the simulation results with the size of the (dry) majority ranging between 55% and 90% of the population, and the number of trendsetters going from 1 to 20. As we already mentioned, the even/odd difference matters. However, we shall focus upon the even-number case because the general tendencies are the same, though we should substitute partial cascades for no cascades in the odd-number case.

Figures 1, 2, and 3 graphically depict the results for the even-number case: the probability of positive cascades increases monotonically with increases in the size of the majority (as a percentage of the population) and the number of trendsetters. The probability of no cascades increases monotonically with decreases in the size of the majority and the number of trendsetters. The probability of negative cascades increases monotonically with a decrease in the size of the majority, though the relation between the probability of negative cascades and the number of trendsetters is not monotonic.

The last finding deserves a closer scrutiny. As Figure 3 shows, in the range of large majorities, i.e., when the majority comprises 70, 80 or 90% of the population, seemingly intuitive monotonicity is preserved, that is, the smaller the number of trendsetters, the higher the probability of negative cascades. However, when the majority is just 55 or 60% of the population, monotonicity breaks down. Rather, monotonicity is reversed between 2 and 10 trendsetters, though the relation is reversed again

between 10 and 20 trendsetters. In other words, if the majority – as well as the number of trendsetters – is small, say 55% and 10, respectively, decreasing the number of trendsetters leads to a *decrease* in the probability of negative cascades. Actually, in the case of a 55% majority, the probability of negative cascades is globally maximized with 10 trendsetters (26.14%). As the case of negative cascades shows, monotonicity does not hold for some combinations of parameters.

One conclusion one can draw is that – whenever a large majority of the population prefers, say, to be dry, a very small number of trendsetters can have a disproportionate effect on the probability that an unpopular drinking norm will emerge. On the contrary, if the majority of dry people is quite small, unpopular drinking norms are more likely to be established in the presence of a sizable number of trendsetters. Many social contagion phenomena like college students' alcohol consumption, binge eating, teenage smoking and even widespread illegal behaviors such as bribing practices seem to originate from the actions of a relatively *small* group of individuals. It is always surprising to realize that many of those who adopt these behavioral patterns have a negative attitude toward them, because we have a tendency to expect consistency between attitudes and behavior. Our model shows that such inconsistencies are not necessarily a sign of irrationality: people in the grip of pluralistic ignorance may rationally choose to behave in ways they privately dislike. The conformists in our model choose to conform to whatever they perceive to be the majority preference, often at the expense of neglecting their own tastes or values. Furthermore, it is important to recognize that – whenever most people share a given private preference – the presence of pluralistic ignorance makes it easy for even an extremely small “contrarian” minority to steer public behavior in the direction of *their* preferences. This is how unpopular norms come into existence.

*What happens if some happen to tremble?*

Once a cascade occurs, there is no incentive for anyone to deviate even if the majority of people hate the status quo. A norm has been established, and since everyone believes other people's compliance to reveal a genuine preference, nobody wants to bear the cost of deviating from the norm. However, some may eventually deviate, because they either slip into revealing their private preferences or just make a mistake. The term

'mistake' encompasses several possible reasons why one would deviate from the norm. A 'dry' person may order beer in a bout of depression, and a manager intent on bribing a public officer might mistakenly interpret some piece of information as suggesting that in that particular moment it might be unwise to offer a bribe.

Given our assumptions, it follows that people are unlikely to consciously choose to deviate, but it is not unreasonable to assume that others *believe* the deviant's off-equilibrium choice to reveal his true preference. In a situation of pluralistic ignorance, this is precisely what would be presumed. We thus assume the common belief about deviations to be as follows: the probability that a deviation from the norm reveals a true preference is taken to be  $1 - \epsilon$ , and that of a simple mistake is taken to be  $\epsilon$  ( $\ll 1$ ). Note that  $\epsilon$  is a function of individuals' belief about how many deviants are conformists (and thus can only make a mistake).<sup>4</sup> We also have to assume that people believe that some among them (the falsely perceived minority) are conformist.

The interesting question to ask is under which conditions a negative cascade will be reversed, i.e., under which conditions an established norm that most people privately dislike will collapse. As an example, let us consider the case of a negative cascade where the current norm is  $x_2$  but privately the majority ( $p \cdot N$ ) prefers  $x_1$ . Take  $\#x_2 - \#x_1$  to be the difference between the number of type 2 and type 1 observed actions taken by trendsetters before the current cascade occurs. For example, if there are two trendsetters and both choose  $x_2$ , the difference  $\#x_2 - \#x_1$  will be 2 ( $2 - 0$ ). Then it will take  $\#x_2 - \#x_1$  observations of 'trembled'  $x_1$  actions to induce people who privately prefer  $x_1$  to reveal their true preferences. The general principle is that if  $n$  ( $\#x_i - \#x_j$ ) actions of type  $i$  taken by trendsetters were sufficient to generate a cascade, it will take  $n$  actions of type  $j$  to reverse the cascade. The reason is simple. Once a cascade occurs, individuals' actions no longer depend on their private information (their preferences), hence their behavior is uninformative to others. Thus a cascade aggregates the information of only a few early individuals' actions. In our example, the relevant information is that provided by the actions of trendsetters. To shatter a cascade, individuals will only need to observe a number of 'deviant' actions sufficient to offset the information conveyed by the trendsetters' actions. The fact that the majority follows a norm does not therefore entail that it will take a major release of alternative public information to abandon it. Very little public information, in the form of very few observable 'deviant' actions, may be sufficient.



For example, if  $\#x_2 - \#x_1 = 2$ , only one observation of  $x_1$  is not sufficient to break the equilibrium. Before observing the ‘tremble’  $x_1$ , the odds of  $x_1$ -preferring people are:

$$\frac{(1-p) \cdot (1-p)}{p \cdot p} \cdot \frac{p}{1-p} = \frac{1-p}{p} < 1$$

The odds after the  $x_1$  tremble are:

$$\frac{p}{1-p} \cdot \frac{p \cdot (1-\varepsilon) \cdot (1-p) \cdot \varepsilon}{p \cdot \varepsilon + (1-p) \cdot (1-\varepsilon)} = \frac{p}{1-p} \cdot \frac{p - (2p-1) \cdot \varepsilon}{1-p + (2p-1) \cdot \varepsilon} < 1 \quad (\because 2p-1 > 0).$$

But if two people tremble, the odds for  $x_1$  and  $x_2$ -preferring people are, respectively:

$$\frac{p}{1-p} \cdot \left\{ \frac{p \cdot (1-\varepsilon) \cdot (1-p) \cdot \varepsilon}{p \cdot \varepsilon + (1-p) \cdot (1-\varepsilon)} \right\}^2 = \frac{p}{1-p} \cdot \left\{ \frac{p - (2p-1) \cdot \varepsilon}{1-p + (2p-1) \cdot \varepsilon} \right\}^2 < 1.$$

$$\frac{(1-p)^3}{p^3} \cdot \left\{ \frac{p \cdot (1-\varepsilon) \cdot (1-p) \cdot \varepsilon}{p \cdot \varepsilon + (1-p) \cdot (1-\varepsilon)} \right\}^2 = \frac{(1-p)^3}{p^3} \cdot \left\{ \frac{p - (2p-1) \cdot \varepsilon}{1-p + (2p-1) \cdot \varepsilon} \right\}^2 < 1.$$

Then, everyone will be induced to truthfully reveal his preferences because  $x_1$ -preferring people switch to  $x_1$  and  $x_2$ -preferring people stick to  $x_2$ .

It should be noted that some of the assumptions we made are important in generating a cascade. For example, we have assumed that a conformist, in the absence of relevant information about what the majority is, will choose by flipping a coin. If we were to assume that people use their true preference as a tie breaker, information cascades would be mitigated. In the case of  $z = 1$ , whatever behavior this lone trendsetter takes, people would act on the basis of their true preferences. For example, if the trendsetter orders beer, wet people will order beer; dry people will choose soda because two conflicting pieces of evidence (one person is wet and one person (himself) is dry) give no clue beyond the common prior belief, and accordingly people act on their true preferences. Actually partial cascades, which can only occur in the case of an odd number of observa-

tions, would never occur even in our simultaneous setting with true revelation of preferences as a tie-breaking rule. Another interesting case is that of people who are not perfect Bayesians. For example, people may take others' behavior into account more or less than predicted by Bayes' theorem, and in this case information cascades might be amplified or mitigated. Conservative belief revision, i.e., less than optimal revising from a Bayesian viewpoint, could be a useful psychological mechanism in terms of blocking the emergence of information cascades. For example, in our strict Bayesian model, only two initial observations of  $x_2$  are sufficient to generate a cascade. But "conservative" people would need more definite pieces of evidence to conform to the perceived majority. Thus if people were *less* rational in processing their information, it would be more difficult for unpopular norms to get established.

The important conclusion we can draw, however, is another. Unpopular or dysfunctional norms may survive even in the presence of a huge, silent majority of dissenters. They refrain from open defiance because of social pressures they themselves help to sustain through actions that stem from pluralistic ignorance. Our model shows that it may take surprisingly little new public information to reverse the original cascade. Interestingly enough, we need not assume much about the sources of such information. It would be a mistake to suppose that only the actions of a 'subversive' minority or the availability of public information about what most people really think (or like) can be expected to generate sudden and unexpected changes in well-established norms. Deviant behavior may occur for many other reasons, and it may well be unintended: what matters is that it may take very few observations to convince people to change their behavior in the direction of what they truly prefer.

#### NOTES

<sup>1</sup> For an analysis of some possible dynamics of corruption norms, see Bicchieri and Rovelli (1995), and Bicchieri and Duffy (1997).

<sup>2</sup> In some experiments, however, bystanders did not communicate or observe each other, and the 'victim' was heard but not seen. In other experiments, bystanders were face to face, but again the emergency was only heard and not directly observed. Their results have been replicated in other experiments in which the emergency was not directly observed (Schwartz and Clausen, 1970). When the victim, as well as other bystanders, are observed, results are mixed. Piliavin and Rodin's 1969 subway field study suggests that diffusion of responsibility increases as the cost of helping increases and the cost of not helping decreases

<sup>3</sup> Our assumptions can be relaxed, however. What is crucial to our model is that individuals assign the same prior probability (50%) to the trait characterizing the majority (minority). It does not matter that people may have different beliefs about the size of the majority (minority): that is, even if some believe the majority to be 51%, and others believe it is 99% of the population, we get the same results. Hence the assumption of a common prior distribution is unnecessary.

<sup>4</sup> We are assuming here that  $\epsilon$  is fixed. In this case, the mistake people make is only about who belongs to the majority, not how large the majority is. If  $\epsilon$  is not fixed, then it must vary with  $p$ : as an agent observes, say, more and more  $x_1$  actions,  $p$  will increase and  $\epsilon$  will decrease. In this case, false beliefs about who is in the majority will be reinforced with time.

## REFERENCES

- F.H. Allport (1924) *Social Psychology*. Boston: Houghton Mifflin.
- H.C. Andersen (1994) *Fairy Tales*. London: Penguin Books. First edition, 1835.
- S. Asch (1952) *Social Psychology*. Englewood Cliff, NJ: Prentice Hall.
- C. Bicchieri (1990) "Norms of cooperation", *Ethics* 100.
- C. Bicchieri and C. Rovelli (1995) "Evolution and revolution: The dynamics of corruption", *Rationality and Society* 7.
- C. Bicchieri and J. Duffy (1997) "Corruption cycles", *Political Studies* 1.
- S. Bikhchandani, D. Hirshleifer, and I. Welch (1992) "A theory of fads, fashion, custom, and cultural change as informational cascades", *Journal of Political Economy*, 100.
- A. de Tocqueville (1955) *The Old Regime and the French Revolution*. New York: Doubleday. (First edition, 1856).
- A. Hirschman (1970) *Exit, Voice and Loyalty*. Princeton: Princeton University Press.
- K. Kauffman (1981) "Prison officer attitudes and perceptions of attitudes", *Journal of Research in Crime Delinquency*, 18.
- B. Latane and J.M. Darley (1968) "Group inhibition of bystander intervention", *Journal of Personality and Social Psychology*, 10.
- B. Latane and J.M. Darley (1970) *The unresponsive bystander: Why doesn't he help?*. New York: Appleton Century Crofts.
- C. McFarland and D.T. Miller (1990) "Judgments of self-other similarity: Just like other people, only more so", *Personality and Social Psychology Bulletin* 16.
- D. Matza (1964) *Delinquency and drift*. New York: Wiley.
- D.T. Miller and C. McFarland (1987) "Pluralistic ignorance: When similarity is interpreted as dissimilarity", *Journal of Personality and Social Psychology* 53.
- D.T. Miller and C. McFarland (1991) "When social comparison goes awry: The case of pluralistic ignorance". In J. Suls and T. Wills (eds.), *Social Comparison: contemporary theory and research*. Hillsdale, NJ: Erlbaum.
- D.T. Miller and D.A. Prentice (1994) "Collective errors and errors about the collective", *Personality and Social Psychology Bulletin*, 20.
- H.J. O'Gorman (1975) "Pluralistic ignorance and White estimates of White support for racial segregation", *Public Opinion Quarterly* 39.
- J.S. Packard and D.J. Willower (1972) "Pluralistic ignorance and pupil control ideology", *Journal of Education Administration*, 10.

- I.M. Piliavin and J. Rodin (1969) "Good samaritanism: an underground phenomenon?", *Journal of Personality and Social Psychology* 13 (4).
- D.A. Prentice and D.T. Miller (1996) "Pluralistic ignorance and the perpetuation of social norms by unwitting actors". In M. Zanna (ed.) *Advances in Experimental Social Psychology*. San Diego: Academic Press.
- R.L. Schanck (1932) "A study of community and its group institutions conceived of as behavior of individuals", *Psychological Monographs* 43 (2).
- S.H. Schwartz and G.T. Clausen (1970) "Responsibility, norms, and helping in an emergency", *Journal of Personality and Social Psychology* 16 (2).
- H. Tajfel (1981) *Human Groups and Social Categories*. Cambridge: Cambridge University Press.
- C. Turnbull (1972) *The Mountain People*. New York: Simon and Schuster.
- S. Wheeler (1961) "Role conflict in correctional communities". In D.R. Cressey (ed.) *The Prison: Studies in Institutional Organization and Change*. New York: Holt, Rinehart and Winston.
- R. Wuthnow (1991) *Acts of Compassion: Caring for Others and Helping Ourselves*. Princeton: Princeton University Press.

*Cristina Bicchieri*  
*Carnegie Mellon University*

*Yoshitaka Fukui*  
*Carnegie Mellon University*

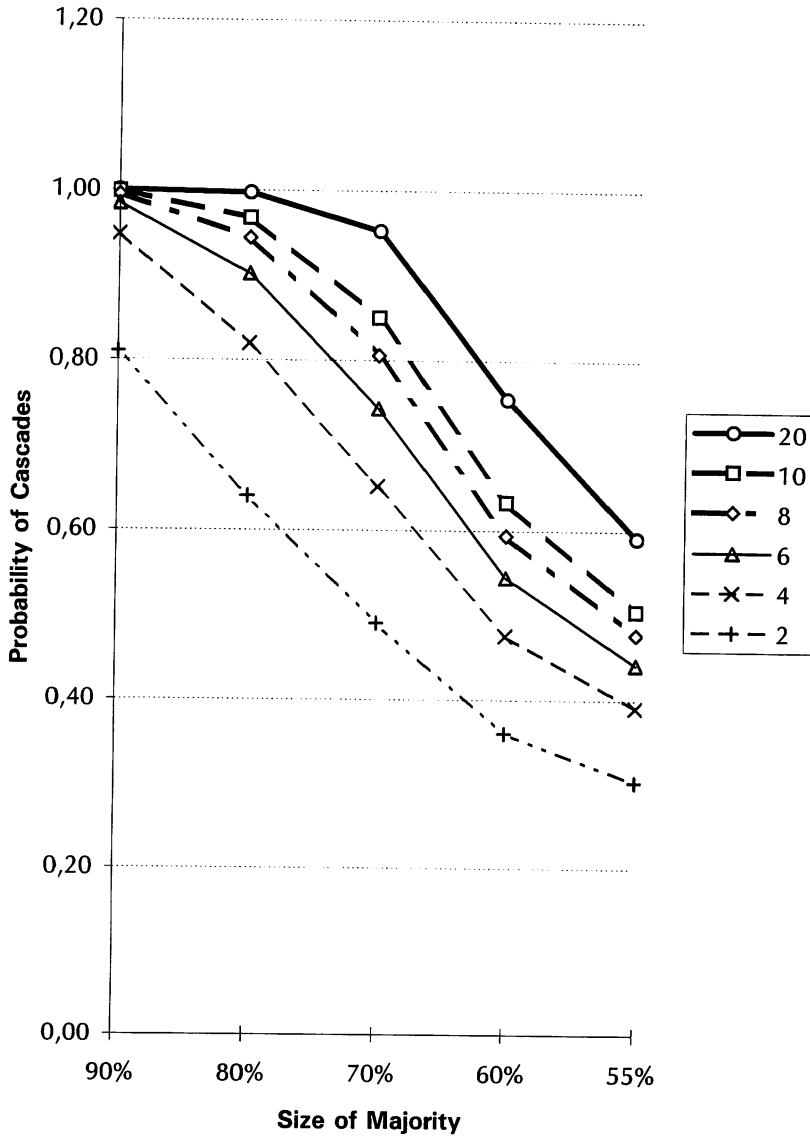
**Table 1: Simulation Results (Even)**

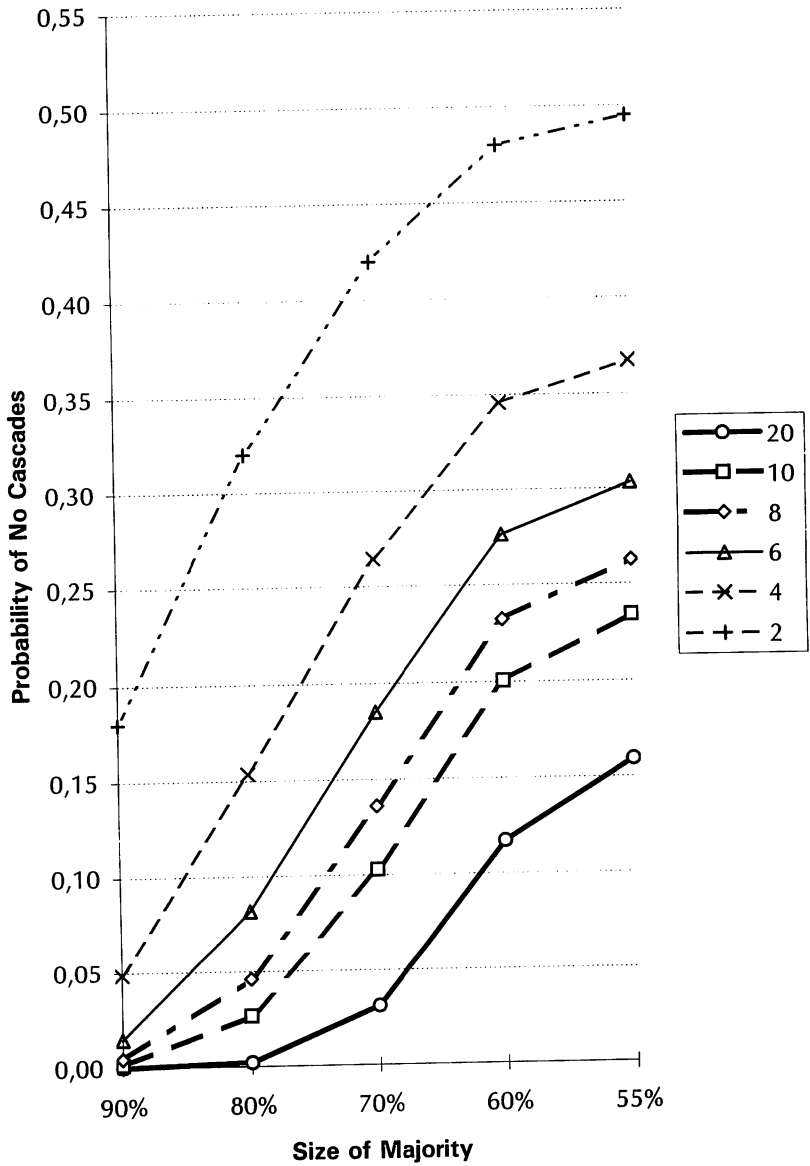
Number of Trendsetters	20	20	20	20	20
Size of Majority	0,90	0,80	0,70	0,60	0,55
Size of Minority	0,10	0,20	0,30	0,40	0,45
Prob. of Negative Cascades	0,00	0,00	0,02	0,13	0,25
Prob. of Positive Cascades	1,00	1,00	0,95	0,76	0,59
Prob. of Partial Cascades	--	--	--	--	--
Prob. of No Cascades	0,00	0,00	0,03	0,12	0,16
Number of Trendsetters	10	10	10	10	10
Size of Majority	0,90	0,80	0,70	0,60	0,55
Size of Minority	0,10	0,20	0,30	0,40	0,45
Prob. of Negative Cascades	0,00	0,01	0,05	0,17	0,26
Prob. of Positive Cascades	1,00	0,97	0,85	0,63	0,50
Prob. of Partial Cascades	--	--	--	--	--
Prob. of No Cascades	0,00	0,03	0,10	0,20	0,23
Number of Trendsetters	8	8	8	8	8
Size of Majority	0,90	0,80	0,70	0,60	0,55
Size of Minority	0,10	0,20	0,30	0,40	0,45
Prob. of Negative Cascades	0,00	0,01	0,06	0,17	0,26
Prob. of Positive Cascades	0,99	0,94	0,81	0,59	0,48
Prob. of Partial Cascades	--	--	--	--	--
Prob. of No Cascades	0,00	0,05	0,14	0,23	0,26
Number of Trendsetters	6	6	6	6	6
Size of Majority	0,90	0,80	0,70	0,60	0,55
Size of Minority	0,10	0,20	0,30	0,40	0,45
Prob. of Negative Cascades	0,00	0,02	0,07	0,18	0,26
Prob. of Positive Cascades	0,98	0,90	0,74	0,54	0,44
Prob. of Partial Cascades	--	--	--	--	--
Prob. of No Cascades	0,01	0,08	0,19	0,28	0,30
Number of Trendsetters	4	4	4	4	4
Size of Majority	0,90	0,80	0,70	0,60	0,55
Size of Minority	0,10	0,20	0,30	0,40	0,45
Prob. of Negative Cascades	0,00	0,03	0,08	0,18	0,24
Prob. of Positive Cascades	0,95	0,82	0,65	0,48	0,39
Prob. of Partial Cascades	--	--	--	--	--
Prob. of No Cascades	0,05	0,15	0,26	0,35	0,37
Number of Trendsetters	2	2	2	2	2
Size of Majority	0,90	0,80	0,70	0,60	0,55
Size of Minority	0,10	0,20	0,30	0,40	0,45
Prob. of Negative Cascades	0,01	0,04	0,09	0,16	0,20
Prob. of Positive Cascades	0,81	0,64	0,49	0,36	0,30
Prob. of Partial Cascades	--	--	--	--	--
Prob. of No Cascades	0,18	0,32	0,42	0,48	0,50

**Table 1: Simulation Results (Odd)**

Number of Trendsetters	19	19	19	19	19
Size of Majority	0,90	0,80	0,70	0,60	0,55
Size of Minority	0,10	0,20	0,30	0,40	0,45
Prob. of Negative Cascades	0,00	0,00	0,01	0,09	0,18
Prob. of Positive Cascades	1,00	0,99	0,92	0,67	0,49
Prob. of Partial Cascades	0,00	0,01	0,07	0,24	0,32
Prob. of No Cascades	--	--	--	--	--
Number of Trendsetters	9	9	9	9	9
Size of Majority	0,90	0,80	0,70	0,60	0,55
Size of Minority	0,10	0,20	0,30	0,40	0,45
Prob. of Negative Cascades	0,00	0,00	0,03	0,10	0,17
Prob. of Positive Cascades	0,99	0,91	0,73	0,48	0,36
Prob. of Partial Cascades	0,01	0,08	0,25	0,42	0,47
Prob. of No Cascades	--	--	--	--	--
Number of Trendsetters	7	7	7	7	7
Size of Majority	0,90	0,80	0,70	0,60	0,55
Size of Minority	0,10	0,20	0,30	0,40	0,45
Prob. of Negative Cascades	0,00	0,00	0,03	0,10	0,15
Prob. of Positive Cascades	0,97	0,85	0,65	0,42	0,32
Prob. of Partial Cascades	0,03	0,14	0,32	0,48	0,53
Prob. of No Cascades	--	--	--	--	--
Number of Trendsetters	5	5	5	5	5
Size of Majority	0,90	0,80	0,70	0,60	0,55
Size of Minority	0,10	0,20	0,30	0,40	0,45
Prob. of Negative Cascades	0,00	0,01	0,03	0,09	0,13
Prob. of Positive Cascades	0,92	0,74	0,53	0,34	0,26
Prob. of Partial Cascades	0,08	0,26	0,44	0,58	0,61
Prob. of No Cascades	--	--	--	--	--
Number of Trendsetters	3	3	3	3	3
Size of Majority	0,90	0,80	0,70	0,60	0,55
Size of Minority	0,10	0,20	0,30	0,40	0,45
Prob. of Negative Cascades	0,00	0,01	0,03	0,06	0,09
Prob. of Positive Cascades	0,73	0,51	0,34	0,22	0,17
Prob. of Partial Cascades	0,27	0,48	0,63	0,72	0,74
Prob. of No Cascades	--	--	--	--	--
Number of Trendsetters	1	1	1	1	1
Size of Majority	0,90	0,80	0,70	0,60	0,55
Size of Minority	0,10	0,20	0,30	0,40	0,45
Prob. of Negative Cascades	0,00	0,00	0,00	0,00	0,00
Prob. of Positive Cascades	0,00	0,00	0,00	0,00	0,00
Prob. of Partial Cascades	1,00	1,00	1,00	1,00	1,00
Prob. of No Cascades	--	--	--	--	--

**Figure 1: Probabilities of Positive Cascades**



**Figure2: Probabilities of No Cascades**



**Figure3: Probabilities of Negative Cascades**

