NÚMERO 84

David Mayer, Efraín Bringas and Raúl García

OBEDIENCE UNDER NORMATIVE CONFLICT: A POSTCONVENTIONAL AGENCY MODEL OF MILGRAM'S EXPERIMENT

Summary

We model normative behavior when there is normative conflict between an agent and her context. We extend the postconventional agency model, in which norm-guided behavior may depend on context and incentives, by endowing the agent with self-esteem and the capacity to attribute prestige. When there is normative conflict these variables ponder the agent's dispositions against the situation's imperatives. To fix ideas we model Milgram's experiments on obedience, which have been a focus of debate on the strength of normative behavior in the face of the power of the situation. The model's results reproduce the various behaviors observed experimentally, and support an interactionist perspective. The model can be used for micro-economic institutional analysis.

Introduction

To synthesize the concepts of *Homo sociologicus*, dictated by social norms, and of *Homo economicus*, who chooses rationally (Elster, 1989), it is necessary to conceive of an economic agent who not only rationalizes her actions guided by self-interest, reacting to structures of incentives, but also acts in terms of a set of social norms and moral principles.

Several economists have incorporated the study of norms in economics. Akerlof (1982a, 1982b, 1984) develops models of "wage contracts" with social norms determined endogenously by some actions of the firm. Kahneman, Knetsch and Thaler (1986) suggest that due to "standards of fairness" people may prefer a loss to a distribution perceived as unfair. In Rabin (1993) emotions and perceptions of fairness, originated in employee-manager or consumer-monopolist relationships, have economic and welfare implications. From the point of view of health, Fuchs (1996) suggests that social norms affect preferences and these behavior, making it relevant for economists to analyze the social and economic consequences of these links. However, although norms are introduced in these works, the models describing them tend to be unrelated to the psychological mechanisms underlying normative behavior. Moving beyond social norms, Elster (1996) writes that if emotional experience is an important source of human satisfaction then "*economists have totally neglected the most important aspect of their subject matter*" (p. 1386). We shall find that describing the functioning of normative behavior will involve psychological considerations which move in the realm of emotional experience.

The main difficulty for including normativity as a fundamental aspect of the behavior of economic agents is that normative behavior may be congruent or incongruent, even in one individual, so that it does not appear to be determined by a fixed frame of reference. The problem is even more complex when the actions of individuals involve reference to two or more norms in irreducible conflict (MacIntyre, 1985). Tapp, Gunnar and Keating (1983) incorporate the possibility of flexible, reasoned normative behavior, characterizing normative reasoning in terms of a process of individual growth which can be viewed as occurring in three stages: the preconventional stage of early childhood, the conventional stage of late childhood and early adolescence, and the postconventional stage of the adult person. Normative judgment in a postconventional individual has universal values as premises from which behavioral norms depending on the context are derived. According to the theories of psychological congruence (Heider, 1958, Festinger, 1957, 1964), the force with which the individual feels she should commit to these behavioral norms derives from her need for emotional and cognitive congruence. This force confronts the forces in the situation in which she finds herself, which may include the structure of incen-

tives and in this interaction between moral dispositions and situations there arise psychologically congruent or conflictive actions, that is, moral congruence or incongruence. García-Barrios and Mayer (1995) propose a *postconventional agency* model based on these theories of the development of moral reasoning and of psychological congruence, in terms of which the normative conflict and the degree of congruence in normative behavior can be represented. The authors characterize moral strength and deficiency in economic contexts, and study their effects on efficiency in contracts with gift-exchange and asymmetric information.

The problem of normative congruence becomes particularly evident in situations in which individuals obeying the instructions of authority may break their normativity, a situation characterized by normative conflict. This is a wide-ranging phenomenon occurring in situations ranging from everyday life to war. In a behavioral study of obedience Milgram (1963) provides a striking example in an experimental situation. A high proportion of the subjects of his experiments were induced to apply high-intensity (fictitious) electric shocks to people who were supposed to be the subjects of a learning experiment. Milgram's results question in a fundamental manner the idea that behavior is guided by norms. The purpose of this article is to show that the postconventional agent introduced by García-Barrios and Mayer (1995) can be used to model behavior in the presence of normative conflict and thus to understand Milgram's experiments. We add the dimensions of self-esteem and attributed prestige to the postconventional agent and present a mathematical model which reproduces the diverse behavior patterns associated with these experiments.

The plan of the paper is as follows. We first summarize Milgram's experiments and the related debate on situations and dispositions. We then discuss Akerlof's model of indoctrination and obedience, based on cognitive salience, which is an antecedent of our own. To introduce the psychological aspects of our model, we discuss the relations between conformism and self-esteem and between cognitive dissonance and changes in self-esteem and attributed prestige, as well as resistance mechanisms to the loss of self-esteem. Introducing our model in detail, we summarize postconventional normative behavior and its relation with cognitive dissonance, and propose a dynamics of self-esteem and prestige in Milgram's experiment. Then we discuss the model's results, its wider implications for the debate on situations and dispositions and for policies based on obedience, and its uses in microeconomics.

The debate on Milgram's experiments

Milgram's experiment

In "Behavioral Study of Obedience" (1963), Milgram reports an experiment in which the subjects were induced by means of instructions provided by an experimenter to apply punishments to other people. The specific obedience consisted in

applying (fictitious) electric shocks of increasing voltage to a person who played the role of "bad learner" but who in fact was a confederate of the researcher. As the intensity of the shocks increased, the "bad learner" acted as if he were in increasing pain, until he showed signs of extreme suffering and cried for the termination of the experiment. If the subjects began to doubt their continued obedience of the experimenter's instructions, the experimenter pressed the subject by persuasive means which insisted on the necessity of their obedience. If the subject finally refused to continue the experiment was ended. Even though showing strong signs of emotional conflict, 65% of the subjects obeyed to the point of administering 450 volt (fictitious) shocks. The remaining 35% of the subjects who abandoned the experiment also showed signs of strong emotional conflict.[1] Milgram's conclusions stress two points: (a) the tendency to obey authority in the majority of individuals, even against the moral (or normative) precept of not hurting a person without her consent and (b) the fact that the experimental procedure generated extraordinary levels of tension and a serious level of difficulty (i.e., emotional effort) for the individuals to make effective their decision to abandon the experiment[2] (the italics are ours).

In another experiment by Milgram (1965a), it was found that some factors such as diminishing the prestige of the sponsoring institution or of the experimenter reduced the degree of obedience. It was shown that when the experimenter was situated at a distance and gave instructions by telephone the proportion of obedience was reduced to 25%. Another discovery was that some of the subjects who continued the experiment "cheated" by administering shocks of lower intensity without informing the experimenter. In another variant of the original experiment (Milgram 1965b) the hypothesis of directionality of social pressure was investigated, that is, its capacity to orient conduct towards obedience or disobedience. The results showed a significant difference when the pressure of the group was directed towards disobedience, since then only 10% of the subjects obeyed the instructions until the end. When the pressure of the group was oriented in the sense of reinforcing obedience, only slight increases were observed in comparison with the 1963 experiment. Milgram, interpreting these results, affirms that the social pressure exerted by the authority figure "has preempted subjects who would have submitted to group pressures" (p. 134); that is, her prestige and the elements of persuasion used had already concentrated the capacity for exerting pressure on the individual.[3]

[1] Milgram (1963) reports the misgivings of a subject abandoning the experiment: "I don't think this is very humane...Oh, I can't go on with this; no, this isn't right. It's a hell of an experiment. The guy is suffering in there. No, I don't want to go on. This is crazy"(p. 376)

[2] Ross (1988) affirms that "In fact, many subjects essentially said 'I quit', only to be confronted with perhaps the most important yet subtle feature of the Milgram paradigm, the difficulty of translating an intention to discontinue participation into effective action" (p. 103)

[3] Explaining the results of the experiment from the perspective of the law of social impact Flanagan (1995) stresses the presence of two relevant factors: immediacy (proximity in space and time) of the source of influence and its strength.

The situation-disposition debate

Milgram's experiment has been extensively discussed in the context of the situation-disposition debate. "Most researchers classify all the potential causes that we might use to explain someone's action into two kinds: situational (or external) and dispositional (or internal). Situational attributions identify factors in the social and psychological environment that are causing the person to behave in a particular way. [...] In contrast, dispositional attributions identify the causes of behavior as residing within the individual and thus reflect some unique property of that person" (Zimbardo, Ebbesen and Maslach, 1977, p. 74). Psychologists who maintain that the situation has a determinant force on the actions of the individual affirm "that many, perhaps the majority of people, can be made to do almost anything by the strength of the situation they are put in, regardless of their morals, personal convictions, and values" (Erich Fromm about Zimbardo, 1974, p. 53). Arguing against this point of view, Fromm thinks that Milgram's experiment is interesting not only as an analysis of obedience and authority, but also of *cruelty and destructivity* (*ibid.*). What is surprising to him is not the number of individuals who continued the experiment until the end, but the proportion of subjects who disobeyed in spite of the few facilities that the context made available for doing so. He also does not think that Milgram's surprise about his main observations is justified: the accumulation of tension in the subjects and the difficulty they confronted in making the decision to abandon the experiment. "The main result of Milgram's study seems to be one he does not stress: the presence of conscience in most subjects, and their pain when obedience made them act against their conscience" (*ibid.*, p. 52). For Fromm, the experimenter is not only an authority to whom obedience is owed but a representative of science and scientific institutions, something which introduces strong conflicts in individuals: performing actions which are ever more distant from their norms in order to comply with the requirements of a scientific experiment, and maintaining the idea that the experiment is scientific and pursues a worthwhile objective, when it appears to break the applicable norms.

Modeling Milgram's experiment

Akerlof's model of indoctrination and obedience

Akerlof (1991) writes a model based on the phenomenon of cognitive salience applied to the actions of the present, which he applies to explain Milgram's experiment. The concept of salience originates in psychological studies and consists in overvaluing certain aspects of the perceptual field. The model demonstrates that salience can originate inconsistency between the actions of the agent and her preferences. In spite of having a perfect knowledge of her preferences and of the future, by distort-

ing the magnitude of present actions the subject may distance herself from her objectives. Applying this mechanism, if disobeying authority represents a cost which is cognitively salient, a person or group with authority can manipulate individuals to act contrarily to their preferences if she succeeds in obtaining the deviation gradually.

For this model to explain Milgram's experiment the act of disobeying authority would have to be not only salient, but considerably more salient than applying electric shocks. The explanation of why one conduct would be more salient than the other would remit us to the original question, unless the greater salience of disobedience were fully attributed to the sharpness of the decision involved in leaving the experiment compared to the gradual increase in electric shocks. In any case, cognitive salience cannot adequately serve as a basis to describe the pain and tension expressed by the subjects.

For Akerlof, norms are included as preferences in the utility function and are therefore inherently fixed. However, salience allows the subject to deviate from them. This deviation is a function of the situation and the subject has no self-defense mechanisms. Akerlof gives additional explanations of how the subject would reduce the cognitive dissonance arising because of the inconsistency between her actions and preferences by justifying her actions ex-post: "Once people have undertaken an action, especially for reasons they do not fully understand, they find reasons why that action was in fact justified".

In Akerlof's model there is no normative conflict. His model does not explain the level of tension suffered by the subjects during and after Milgram's experiment. Other relevant facts are omitted, such as the destructive effects on the individual's self-esteem and/or on the experimenter's prestige. The agents described by Akerlof are completely manipulable within the bounds of the salience effect, lacking any mechanism of resistance. But in reality dissonance and the efforts to reduce it not only adapt the individual to an external situation but can affect her perception of it, therefore setting up a limit for the capacity of control that may be exerted on them.

Normative conflict and Milgram's experiment

For us, Milgram's experiment is characterized by normative conflict. The subject finds herself unexpectedly in the situation of having to decide between actions that break her norms of not hurting others or disobeying prestigious authorities who have contracted her. Although at first the subject accepts both norms, since the experimenter systematically supports the norm of obedience, as the experiment advances the norm of not hurting others is upheld by the subject while the norm of obedience is also upheld by the commitment with the experimenter. The result is that the experiment implicitly puts into play the self-esteem of the subject against the prestige she attributes to the experimenter. Leaving the experiment is equivalent to reducing this

attributed prestige and overcoming the difficulty of perceiving the experiment as inconsistent and participation in it as undesirable. The experimenter initially obtains obedience from her high initial prestige. As the experiment proceeds and the subject perceives her incongruous actions, the resulting tension and discomfort involve her in a dynamic which may lower her self-esteem and make her more vulnerable to the influence of the experimenter. The experimenter can aim at increasing her influence over the subject in the following ways: reducing the subject's self-esteem by obtaining further compliance; preserving the level of prestige attributed by the subject to the experimenter by not using excessive persuasion, and by raising the required voltage levels gradually, without overshooting her current level of influence. Some factors which make the change of perception on the experiment and the experimenter easier for the subject are: a firmer self-esteem, the flexibility with which she may perceive the incoherence of the experiment; the weight she gives the norm of not hurting others compared to obeying authority; the resistance she may have to persuasion; and how incoherent its excess appears to her. Eventually, the individuals may change their perception and thus establish a limit to the manipulative capacity of the experimenter.

We analyze first the relation between conformity and self-esteem and then the changes in self-esteem and prestige in relation to cognitive dissonance. Finally, we incorporate these mechanisms with postconventional normative behavior in a model of Milgram's experiment.

Conformism and self-esteem

From the perspective of influence,[4] when in the process of social interaction the norms and beliefs of different agents are confronted, some form of conformism may be generated.[5] Alternatively, a situation of indifference may arise between the parties.[6] This may be temporal or permanent, and one of the parties may submit to the other, identify or interiorize her norms or beliefs (Hollander, 1971). These different forms of conformism have been associated with characteristics of the agent exercising influence. For Aronson (1980) submission is associated with the power of the influencing agent, identification with her attraction, and interiorization with her credibility; "if the person who provides the influence is perceived to be trustworthy

[4] Influence intervenes in situations of social interaction marked by the asymmetry of the participants: differences in aptitude, status, level of anxiety, need for social approval, etc. (Moscovici and Ricateau, 1975).

[5] Aronson (1975) defines conformism as "a change in a person's behavior or opinions as a result of real or imagined pressure from a person or group of people" (p. 17).

[6] Sherif and Sherif (1969) have found that "a prestige source that contradicts our assessment of the conditions will be relatively ineffective in influencing our behavior" (p. 70)

and of good judgment, we accept the belief he or she advocates and we integrate it into our own system of values" (Aronson, *ibid.*, pp. 30).

In Milgram's experiment the elements of power, attraction and credibility are all present. We summarize them as the prestige or reputation which the subject attributes to the experimenter. As Fromm (1991) has pointed out "it is very difficult for the average person to believe that what science commands could be wrong or immoral" (p. 51). Investigating conformity, Hollander says "that the more ambiguous the stimulus presented to the subject, the greater the tendency to conform to social pressure by matching the standard response. [...] Among properties found to increase the probability of conformity are the status, power, or competence of the others representing an influence source, and their apparent unanimity" (p. 558). Milgram's experiment presented an unexpected (disconcerting rather than ambiguous) situation, which was well-designed for inducing conformity.

There are also factors which make influence more difficult. This happens when individuals have a high degree of confidence in their own perception, when they feel more competent, powerful or attractive than the other agents. Such feelings are strengthened when supported by other individuals in similar circumstances, as in Milgram's study on the *directionality* of social pressure (1965b). Thus in the process of influence the position of the person who is the object of persuasion (*i.e.*, her status, power, competence, etc.), the size of the discrepancy between her position and the influencing agent's, and the perception that the person has of herself are determining factors. Thus, besides attributed prestige, we introduce the dispositional variable *self-esteem.*[7]

In Coopersmith's definition (1967, p. 17), self-esteem is a judgment of personal self-worth which is reflected in the attitudes the individual has about herself. It indicates how capable, significant, successful or valuable the individual considers herself. An individual with low self-esteem is more easily influenced by persuasive communication than one with a higher opinion of herself. Faucheux and Moscovici (1968) have also related lower self-esteem with a higher dependency on the situation: "low self-esteem subjects (LSE) tend to be more dependent upon their environment than high self-esteem (HSE) subjects" (p. 83).

Let us also note that an individual with high self-esteem will experience a higher level of conflict when breaking her own norms. Aronson (1969, 1975) considers that the theory of dissonance makes its clearest predictions when individual's behavior violate their self concepts: "individuals with the highest self-esteem experience the most dissonance when they behave in a stupid or cruel manner" (1980, p. 143). This is because they will feel responsible for the negative consequences of

[7] Marsh (1996) reports that global self-esteem "is one of the most widely inferred constructs in personality and social psychological research. It is used as an outcome measure, as an intervening variable, and as a basis for testing theoretical models about how individual process, select (with possible biases), and integrate information about themselves" (p. 810).

their conduct, even if they were previously unaware of them, since it is hard for a person to excuse herself on the basis of insufficient previous knowledge (Wicklund and Brehm, 1976).

Cognitive Dissonance and changes in self-esteem and attributed prestige

When an individual perceives herself as acting incongruously, strong feelings arise in the form of tension and emotional discomfort. Such feelings of dissonance are stronger if the individuals have voluntarily accepted to participate or if the behavior involved clearly goes against the person's identity (Helmreich and Collins, 1968). "Dissonance arousal requires the perceptions of a strong causal link between oneself and the potentially dissonance-arousing event" (Wicklund and Brehm 1976, p. 70), as is the case when the subjects of Milgram's experiment hurt the "bad learner". Scher and Cooper (1989) stress that "aversive consequences" are necessary and cognitive dissonance, in the face of such feelings the need for individual congruence brings forth mechanisms to reduce dissonance; "a person attempts to perceive, cognize, or evaluate the various aspects of his environment and of himself in such a way that the behavioral implications of his perceptions shall not be contradictory" (Deutsch and Krauss, 1990, p. 68). There is a tendency to justify personal actions before oneself and others. Once the decisions have been taken, the subject may reduce her dissonance by changing her cognitions, so that the choice seems more valuable (Festinger, 1957; Wicklund and Brehm, 1976).

In the case of Milgram's experiment, the main cognitive changes to which these dissonance reduction mechanisms apply are changes in the perception of self-worth and changes in the attribution of the experimenter's prestige. Being a situation of strong tension, the concept of self is threatened. The subject may feel incapable of understanding the situation as it becomes removed from the expectations she originally held about it, and may generate a high degree of uncertainty on the pertinacy of her own conducts and thoughts. As the experiment proceeds and she perceives her incongruous actions, her self-esteem will suffer. This process will serve as a dissonance reduction mechanism, since the subject will feel less responsible for the negative consequences of her own conduct. In the case of the attribution of prestige, maintaining a high attribution and thus compliance with the experimenter's instructions may imply such incongruence that dissonance reduction may be achieved by reducing the attributed prestige. In common language, the subject may come to change her opinion of the experimental situation.

Resistance mechanisms to the loss of self-esteem

When individuals confront an experimental situation they do not loose their previous experience, or their personality, nor are they totally naive. Instead, they have "an implicit demand of coherence" (Leonard, 1975, p169, our translation). As Moscovici (1975, p. 78) points out, it is necessary to recognize that subjects have "a double life; [each individual] on the one hand executes what she is asked to do, and on the other elaborates her little inner theory about the experiment, about the experimenter" (our translation). Individuals do not automatically incorporate the beliefs or norms of the authorities. Instead, they possess the capacity to resist the pressure towards conformity and the reduction of self-esteem, and can manifest anticonformist behavior. (Merton, 1957; Elster, 1987). This capacity is relevant in situations restricting the free expression of the preference for disobedience, particularly when obedience implies hurting others. According to Aronson (1980), there are two major ways to reduce the psychological discomfort provoked by the discrepancy between the positions of the influencing party and the influenced one: "they can change their opinion, or they can derogate the communicator" (p. 85). Just as norms are not immovable, the capacity of a source of influence to generate conformism can change when her power, attraction or credibility changes. Credibility may weaken in the presence of more convincing contrary opinions, if there is evidence that an attempt is being made to manipulate the individual against her beliefs or in favor of a personal interest, or in some cases if communication is unilateral. Resistance to pressure will also increase when influence employs insistent means: persuasive communication tends to wear down the credibility of the source of influence.[8] Finally, when subjects loose self-esteem in their interaction with another person, the attraction felt for her tends to diminish (Aronson and Linder, 1965).[9]

Thus we have the following mechanisms of resistance. In a state of psychological conflict or tension an individual may, instead of reducing tension by modifying her norms or beliefs in the direction of the source of influence, change her perception of the source, particularly if maintaining the perception implies a loss of self-esteem. This mechanism may be strengthened by an increasing resistance to lowering her self-esteem, compared to the resistance to changing her attributed prestige.

[8] In Milgram's experiments, the experimenter did not explain or justify her lack of worry, and did not give reasons for continuing the experiment. Her reaction was in the form: "Please continue", "Please go on", "The experiment requires that you continue", "It is absolutely essential that you continue" or "You have no other choice, you *must* go on". (p. 374).

[9] Self-esteem may be manipulated in experimental procedures (Zimbardo, 1975). For example, in relation to the exposure to information, Canon (1964) found that people will listen less to arguments contrary to their own beliefs if they confront an induced diminution of their confidence in themselves.

Postconventional normative agency: the model

The concept of postconventional normativity is presented in detail in García Barrios and Mayer, 1995. Normative behavior is conceptualized as occurring in two stages. In the first the subject derives behavioral norms from a general or universal principle. In the second she decides her actions in view of the behavioral norms she has derived. The psychological processes implicit in each of these stages are understood in terms of the theories of psychological congruence. We shall outline these processes as we introduce them in our model of Milgram's experiment, in which two norms come into play: not hurting others and obeying authority.

The first stage, which corresponds to the formation of the behavioral norm is psychologically prior, in that deviations from its independent functioning represent substantive psychological disadjustments. Thus, we suppose that the subject arrives at her behavioral norms independently of the dissonance which may result from not acting according to them, or of other sources of dissonance or tension present in the situation. In this case, the subject has not developed a specific normativity adjusted to the particular conflict in which she finds herself unexpectedly, so she evalues her behavioral norms from two points of view corresponding to the two norms. In the process of minimizing her cognitive dissonance, she will arrive at a behavioral norm which represents a compromise between the dictates of either norm when viewed independently. Consequently she will be unable to act without feeling dissonant with regard to her choice of behavioral norm. The weight she gives to each norm will depend on several factors. For simplicity we suppose that the weight she gives to the norm of obeying authority will depend on the prestige she attributes to the authority issuing the commands, and to the strength of the persuasive means deployed in issuing the commands, while the weight she gives to the norm of not hurting others will depend on her current self-esteem. In mathematical terms, given a current level of attributed prestige R (for reputation) and persuasion I (for influence), if the subject has a principle for the level of obedience 1 and forms a behavioral norm for a level of obedience λ, if λ differs from 1, we will suppose that she suffers a level of dissonance $D^0_{Ob} = R\, I\, (1 - \lambda)^2$. Here marginal dissonance is an increasing function of how far the behavioral norm is from its underlying principle. The functions could be written in more general terms (see García Barrios and Mayer, 1995) but then the mathematics would be less simple. Dissonance is also increasing in R and I, with the multipli cative form taken for simplicity. Since more dissonance is felt the stronger the persuasion which has been applied we assume $I \geq 1$. Analogously, if the principle for not hurting others is represented by hurting to level 0, and the behavioral norm calls for a level of hurting y, the resulting cognitive dissonance will be represented by $D^0_{Suf} = J\, (y - 0)^2$. Here J represents the current level of self-worth attribution by the subject. By the experiment's design, given a command which implies making another suffer to a level x, the subject must form her behavioral norms λ and y si-

multaneously: obeying a proportion λ of the command x implies making another suffer to the level $y = \lambda x$. The cognitive dissonance involved in choosing λ and y, will be formed by minimizing

$$D^0 = D^0_{Ob} + D^0_{Suf} = RI(1 - \lambda)^2 + J\,y^2$$

subject to $y = \lambda x$, given the levels of R, I and J.

The agent of this model gives each norm a weight depending on the strength attributed to its proponent. The weight of obedience depends on the prestige of the experimenter and the intensity of her persuasive efforts, while the weight of not hurting others is the subject's self-worth. We distinguish self-worth from self-esteem, considering that high self-worth is a high attribution of self-value, while high self-esteem reflects the full dynamic mechanism tending to establish a high self-worth. This involves several components which will be clarified below.

As defined, the dissonance functions include the ideas that: (1) people with less self-esteem will conform more; (2) people attributing a higher prestige to the experimenter will conform more; (3) persuasion will increase conformity (although its use will also have counter-productive effects which will be included below); and (4) people with more self-esteem suffer higher dissonance for breaking their own norms.

We can solve for the behavioral norm λ. Writing $X = x^2$,

$$D^0 = R\,I\,(1 - \lambda)^2 + JX\lambda^2.$$

The solution to the minimization problem is

$$\lambda = R\,I/(R\,I + J\,X)$$

and the resulting level of dissonance is

$$D^0 = \frac{R\,I\,J\,X}{R\,I + J\,X}.$$

In the second stage of normative reasoning the subject decides her actions in terms of her behavioral norms. At this stage she might decide to depart from her norms in view of other dissonances or costs present in the situation. However, she can only do this at the cost of experiencing a level of dissonance which is a consequence of her commitment to her norms. Given that she has chosen a behavioral norm of obedience to the level λ, if she acts choosing to obey to a level μ, she will experience a level of dissonance $D^1_{Ob} = R\,I\,(\lambda - \mu)^2$ analogous to D^0_{Ob}. Likewise, given that she has chosen a behavioral norm of harming others to the level y, if she acts choosing to harm to a level a, she will experience a level of dissonance $D^1_{Suf} = J\,(a - y)^2$. In both functions marginal dissonance is increasing and the weights of prestige, persuasion and self-worth play the same role as before. Because obeying and making another suffer are tied in the experiment, $a = \mu x$. Additionally, in the

1963 version of Milgram's experiment, the subject can only choose between fully complying or else abandoning the experiment. That is, she chooses between the actions x or 0, or equivalently, between levels of obedience μ equal to 1 or 0. Since we have given psychological priority to the first stage of normative reasoning, the subject chooses her action by minimizing her dissonance taking as given the prior choice of behavioral norm. Although there will be other dissonances in the model, they will involve only the levels of prestige and self-worth, and not μ, so we can solve for the action μ. By substituting the value for λ obtained previously,

$$D^I = D^I_{Suf} + D^I_{Ob} = (JX + RI)(\lambda - \mu)^2 = \frac{(R\,I)^2}{R\,I + J\,X}, \frac{(J\,X)^2}{R\,I + J\,X} \text{ for } \mu = 0, 1.$$

Therefore the subject's decision depends on the levels of attributed prestige, applied persuasion, self-esteem, and level of suffering instructed, i.e., R, I, J, X in the following manner:

$$R\,I < J\,X \Rightarrow \mu = 0 \text{ (disobey)}, \quad R\,I \geq J\,X \Rightarrow \mu = 1 \text{ (obey)}.$$

where we have privileged permanence in the experiment in the case of equality. The total level of normative dissonance involved in each choice is:

$D^N = D^0 + D^I = R\,I,\ J\,X$ according to wether $\mu = 0,\ 1$.

It is clear that if the subject had choices in between the extremes of obeying completely or not at all, as in the version of the experiment when instructions were given over a telephone, her choice of action would be different.

Dynamics of self-esteem and prestige in Milgram's experiment

If Milgram had asked the subjects of his experiment to administer 450 volt shocks as they came in, or immediately after having applied 90 volt shocks, the compliance rate would have been practically null. Conformity was obtained gradually. What dynamic mechanism underlies this process? We have seen that conformity depends on the relative levels of prestige and self-worth. Thus during the experiment self-worth must fall more than prestige, until the subject decides to leave the experiment, a decision which must be simultaneous with the fall of the experimenter's prestige. Self-worth will fall due to the subject's perception of her own incongruity, because the situation may feel out of control, and as a mechanism of dissonance reduction. Prestige may fall due to an excessive use of persuasion, because the experiment causes a loss of self-esteem, or as part of a change of perception of the experimental situation. Finally, mechanisms preventing the continued loss of self-esteem may set in, either making it harder for further loss to occur, or making it easier for the prestige of the experimenter to fall. Thus to obtain the highest levels of obedience Milgram's experimenter aims at increasing her influence over the subject by reducing

her self-esteem (which must be done by obtaining compliance) and preserving as much as possible the level of prestige attribution.

We thus introduce the dynamics of self-worth and prestige attribution in our model. When the subject is deciding her behavioral norms and action by minimizing dissonance, she simultaneously by the same process arrives at an attribution of self-worth and prestige. The full model for the decision in period t+1 is the minimization of the dissonance:

$$\min_{J_{t+1},R_{t+1},y,a,\lambda,\mu} D^{Tot} \quad \text{subject to } y=\lambda x,\ a=\mu x,\ \mu=0 \text{ or } 1,\ \lambda=\text{Argmin}(D^0).$$

where

$$D^{Tot}(J_{t+1},R_{t+1},y,a,\lambda,\mu\ ;x,I,R_t,J_t) = D^N(J_{t+1},R_{t+1},y,a,\lambda,\mu\ ;x,I) + D^J(\rho(J_t),J_{t+1}) + D^R(\frac{\rho(R_t)}{DC(I)},R_{t+1}).$$

Here D^{Tot} represents the total dissonance which will result from the decision taking place. This is a function of the variables to be decided, which are the self-worth and prestige attributions J_{t+1}, R_{t+1} together with the behavioral norms y, λ and actions a, μ and of the experimenter's instruction x and level of persuasion I (all of which correspond to period t+1; the restrictions $y = \lambda x$, $a = \mu$ are maintained) together with the self-worth and prestige attributions J_t, R_t of the previous period. These variables represent the reality grounding of the subject. D^N represents the normative dissonance which will result from the decision. D^J and D^R represent the dissonances resulting from changes in the levels of self-worth and prestige. If these changes did not give rise to dissonance, there would be no grounding. We assume that self-worth and prestige attribution change in two steps. First, a process of adjustment occurs (independently of the process of normative decision described by the minimization of D^{Tot}) in which the attributions J_t, R_t of the previous period recuperate to levels $\rho(J_t)$, $\rho(R_t)$ given by

$$\rho(J_t) = J_t^{(1-\alpha)}J_0^{\alpha},\quad \rho(R_t) = R_t^{(1-\alpha)}R_0^{\alpha}$$

where J_0, R_0 are the initial levels of attributed self-worth and prestige and $\alpha \in [0, 1)$ represents the intensity of recuperation, which we choose equal for self-worth and prestige for simplicity. For $\alpha > 0$ this feature of the model, which represents some kind of healing or anchoring in the original levels of self-worth and attributed prestige, means that people differing in their initial attributions are different throughout the experiment. The case $\alpha = 0$ represents a model without this feature. Additionally, after this recuperation but before the subject decides upon her action, the negative effect of the level of persuasion I currently being used by the experimenter factors in,

decreasing the level of prestige to $\rho(R_t)/DC(I)$. DC is a function of discredit which we shall describe below. The second process of adjustment of the attribution of self-worth and prestige occurs jointly with the process of deciding whether to obey or not. New levels J_{t+1}, R_{t+1} for these attributes are established, at the cost of generating contributions to the total dissonance in the amounts $D^J(\rho(J_t), J_{t+1})$, $D^R(\rho(R_t)/DC(I), R_{t+1})$. These dissonances arise from a resistance to changing the attributions, which are perceptions intimately related to the reality grounding of the subject. The particular functions D^J, D^R which we choose for this purpose are

$$D^J(\rho(J_t),J_{t+1}) = \frac{[\rho(J_t)]^{(1+a_J)}}{a_J J_{t+1}^{a_J}} \qquad \text{for} \quad J_{t+1} \le \rho(J_t)$$

$$D^R(\frac{\rho(R_t)}{DC(I)},R_{t+1}) = \frac{\left[\frac{\rho(R_t)}{DC(I)}\right]^{(1+a_R)}}{a_R R_{t+1}^{a_R}} \qquad \text{for} \quad R_{t+1} \le \frac{\rho(R_t)}{DC(I)}$$

$$a_J = a_J(\frac{R_t}{J_t}), \quad a_J' > 0.$$

In the regions $J_{t+1} \ge \rho(J_t)$, $R_{t+1} \ge \rho(R_t)/DC(I)$, we choose any continuous extension of D^J, D^R which is increasing in $J_{t+1}/\rho(J_t)$, $R_{t+1}/DC(I)$ respectively. Total dissonance is always increasing in these regions so its minimum is never attained in their interior. a_J, a_R establish the relative importance of the changes in self-worth and prestige attributions. A higher parameter implies less malleability. These functions yield attractive formulae for the changes in self-worth and prestige once the minimization has been calculated. An increasing relation $a_J = a_J(R_t/J_t)$ can be used to represent a self-defense mechanism tending to set a limit to the loss of self-worth when the influence of the experimenter becomes high.

In the minimization of D^{Tot} only D^N is affected by the variables y, a, λ, μ determining normative. We solved for λ and μ previously, showing how they depend on the current values J_{t+1}, R_{t+1}, x, I. After this partial solution the minimization of D^{Tot} takes the form

$$\min_{J_{t+1},R_{t+1}} D^{Tot}(J_{t+1}, R_{t+1}) = \text{Min}\{R_{t+1}\, I, J_{t+1}\, X\} + D^I(\rho(J_t),J_{t+1}) + D^R(\frac{\rho(R_t)}{DC(I)},R_{t+1}).$$

The discredit function

To complete the model we postulate some properties of the discredit function DC(I), which it must satisfy to reflect the behavior observed in the experiments.

DC1. $DC(1) = 1$. If no persuasion is applied, there is no discredit.
DC2. $DC(I)$ is increasing. The higher the intensity of persuasion applied, the higher the discredit of the experimenter.
DC3. $DC'(1) < a_R/(1 + a_R)$. A small amount of initial persuasion produces an increment in X (see the explanation below).
DC4. $\lim_{I \to \infty} [I^{1-br}/DC(I)] = 0$. This condition implies that the experimenter cannot apply an infinite amount of persuasion without the subject abandoning the experiment. (The condition includes non-log-linearity.)
DC5. $b_R DC'(I) + I\ DC''(I) > 0$. This convexity makes most of the critical points in the model unique.

The model's results

The model is analyzed in detail in the Appendix. Here we only summarize the main results.

There is a threshold level of influence $R_0/J_0 \geq Inf_{min}$ which the experimenter must have to get any compliance at all from the subject. Given that there is some set of instructions which the subject will follow, the experimenter can push the subject only to the point at which she will change her attribution of prestige and therefore abandon the experiment, and not to the limit allowed by her behavioral norm (given by $RI > JX$). There is a maximum level of compliance X which the experimenter can obtain at any given time. To increment the level of compliance X, the experimenter must first give a sequence of commands with the purpose of maximizing his level of influence (defined as $Inf_t = R_t/J_t$). This is only possible if the experimenter's initial influence lies above another, higher threshold level Inf^{Inc}_{Min}. In each of these cases there are two kinds of subjects. For those for whom $DC'(1)(1 + a_J) \geq 1$, it will not be optimal for the experimenter to apply any level of persuasion. On the other hand, if $DC'(1)(1 + a_J) < 1$ the experimenter will obtain optimal increments in influence by applying persuasion, but only if his initial influence is above Inf^{Inc}_{Min}. Observe, however, that if the experimenter can increment her prestige, she does so following the same strategy for both kinds of persons, that is, by obtaining compliance in order to reduce their self-worth. In this case, after dedicating some periods of time to approximately reach her maximum possible influence, the experimenter can give one last command to extract the maximum compliance, knowing that the afterwards her influence will decrease.

In the simpler case in which a_J is independent of Inf_t, the dynamical system describing the influence of the experimenter is either stable or unstable, depending on whether $\alpha(1 + a_J)$ is greater than or less than 1 (the case of equality is also either stable or unstable, depending on the exact form of the function DC). In the stable case, if the experimenter has enough initial influence, she can increment her influ-

ence up to a limit, and otherwise she cannot increment it. The unstable case is similar, except that if the initial influence is high enough, the experimenter can increment her influence without limit. In the cases in which influence increases beyond $b_J DC'(1)/(b_J - DC'(1))$, the model reproduces the monotonic escalation in the applied persuasion and in the level of compliance observed in the experiment.

The condition $\alpha(1 + a_J) > 1$ combines the strength of the self-defense mechanism consisting of dissonance to a change in self-worth with the strength of the autonomous tendency to recuperate initial self-worth. One can also consider the case in which a_J depends on Inf (as would be the case if under some threshold of self-worth the dissonance to a change in self-worth increases). Then the dynamical system representing influence will be stable from those levels of influence at which $\alpha(1 + a_J(Inf)) > 1$.

Observe that in the case $\alpha = 0$ with no tendency to recuperate initial self-worth (and prestige) the dynamical system of influence is unstable (recall that in this case the initial levels of self-worth and prestige make no difference once the experiment proceeds).

The model reproduces excellently the behavior reported in Milgram's experiment. It models people who will never comply, people who will be susceptible to an increasing level of influence until some level is reached, and people over whom an unbounded influence can be exerted. What level is reached depends on a_J, DC, on the parameters a_R, α, and on the initial values of R_0, J_0. Together, a_J, α and J_0 may be taken to represent self-esteem (as opposed to self-worth), while DC and a_R regulate interaction between the self and others with a high prestige attribution. If the subject interacts with a person with enough attributed prestige, she will modify her behavior, bringing it closer to the authority's, both in the case in which she is open to persuasive means and in the case in which she is not. Assuming the individual interacts stably with people whom she attributes a high prestige to (otherwise it would probably be correct to speak of pathology), this is in fact quite reasonable behavior, in that normativity without openness to the influence of people with a high prestige attribution is a behavior that is probably too rigid and therefore counter-productive. This leads us to the following reinterpretation of the relevance of Milgram's results.

Milgram's subjects came to participate in a learning experiment, and tried to comply in good faith with the instructions they received. As is normal in everyday interaction in institutions of the prestige in which these experiments were conducted, they did not even remotely expect to be mislead, and it was only after a great degree of discomfort that some of the subjects abandoned the experiment, while others carried through to the end. It is worth commenting here that the level of damage that was supposed to be associated with the electric shocks was not credible, since the experimenter was co-responsible for their effects (this could easily be argued from a game-theoretic perspective). Thus the experiment is an experiment on deception: how far people can be made to deviate from their norms when they are deceived? It

would be interesting if Milgram had reported his own feelings on the experiment, explaining how far he thought he was deviating from what would be the applicable norms in a learning experiment, and how much he felt he was deceiving his subjects. What Milgram shows is that there are large windows of possibility in which people can be led far astray from their norm-guided behavior, by deception or by other means. But perhaps this makes the level of everyday compliance with norm-guided behavior even more remarkable, since it takes place in the presence of such large windows of opportunity for deviations from the norm. In many contexts, *people do not expect to be deceived.* This means that, in equilibrium, compliance with norms is so high that it can create the conditions which may lead some people to breaking them. Only a deeper analysis of the dynamics of normative behavior, perhaps from the point of view of evolutionary dynamics, can address these questions.

Final remarks

First, let it be said that the model presented is somewhat more complex than is usually the taste with economists. However, anything less, which would be incapable of reproducing a sufficiently wide variety of behaviors, would be unconvincing to psychologists.

Interaccionism

The model makes it clear that situation and disposition interact in highly complex ways. To begin with, disposition is adapted to a very varied environment, so is quite complex in itself. It is therefore inevitable that the interactions it may have with situation must necessarily be highly complex. Discussing the debate on external and internal determinants of behavior, Berkowitz (1986) suggests that "Our thoughts, feelings, and actions are governed by a variety of processes, and no one conceptual approach can do justice to this rich complexity" (p. 16).

Milgram's experiment hardly makes sense in a context in which people are not disposed to be norm-guided. What is important is to dilucidate by what mechanisms the situation might lead individuals to break their norms. To do this we introduce the dynamics of self-esteem and prestige attribution. These, combined with the postconventional normative behavior, can provide an explanation for the experiment's dynamics and for the several kinds of behavior observed in each variant of Milgram's experiment. The high rate of compliance in the experiment can be attributed to the degree of deception it involved, in which a fictitious researcher justified making another suffer, and which the subjects found hard to surmount. But the subjects broke their norms under the supposition that the experimenter was keeping hers. This points out the complex nature of normative behavior and its equilibria,

and raises questions pertaining to the stability of normative systems in the face of the incentives to break them. Perhaps Milgram's experiment, in showing how vulnerable its subjects were to deception and confusion, serves better to provide evidence of the high level of individual and social acceptance which norms enjoy in many contexts, in relation to which the experiment purposefully constructs itself as an exception, rather than to question how meaningful the normative references are in the first place.

The psychological factors we include allow the individual to resist authority in certain circumstances. Thus we can construct the multiple equilibria present in Hollander (1971), when he analyzes the question of whether man is intrinsically good or bad: "Given human susceptibility to the forces in the social environment, the best response seems to be that Man has the capacity for extremes of high morality and conscience as well as for the basest forms of degradation in his treatment of fellow Man" (p. 40).

Social programs involving obedience

Almost all modern theories of administration and bureaucracy coincide with the fundamental Weberian model in that they attribute technocrats and administrators an enormous power over the population (MacIntyre, 1985). In this work, however, we have seen how a program of obedience can induce individuals to destroy not only their self-worth but also the authorities' prestige, and by extension the prestige of the institutions they represent. In his article, Akerlof (1991) concludes that, since cognitive salience induces such irrational behavior as procrastination, some social programs such as forced saving could not only be feasible, if imposed sufficiently gradually, but beneficial. However, our conception implies that programs of obedience involve important costs and risks, and are not as feasible as they may appear. First, obedience can involve a decreased level of self-worth, which as we shall discuss below has many important consequences in itself. Second, if it goes against their wishes, the subjects of an obedience program will become involved in a cognitive process tending to reduce the prestige of authority. The exact cognitive outcome will tend to be unpredictable, but nevertheless will tend to increase the costs of imposing obedience, perhaps making it impossible. Thus, there are limits to people's manipulability.

Self-esteem and microeconomics

Our postconventional agent, endowed with self-esteem and the capacity to attribute prestige, allows the possibility of modeling normative conflict and introduces new dimensions to microeconomic analysis (once the total utility function is defined as

utility over goods minus dissonance). The individual's self-worth, in itself important as one of the basic psychological needs (Maslow, 1970), becomes a variable which can be correlated with decisions about work and education, personal productivity, innovation and creativity, the feasibility of gift exchange, and other phenomena. Prestige attribution can model the credibility of legal, economic or political institutions, and have consequences on the costs and efficiency of their functioning. Including these concepts makes it possible to conceive and understand psychological vicious circles which may be involved in economic backwardness and poverty. For example, (Granato, Inglehart and Leblang, 1996) find in an empirical study that cultural values including obedience affect economic growth. Individuals may find themselves in conflict with institutions which, by design or by circumstance, involve degrees of compliance or subjection which originate low levels of self-worth that generate both the inability to overcome them and inherent institutional instability. Normative conflict may appear between traditional and emerging sectors, introducing vicious cycles in already difficult situations, as individuals attempt to preserve their psychological identity. The economic reality of a social system may conflict with the ideals of justice and welfare of its members, diminishing their self-worth and undermining the credibility of its institutions. Low prestige in law enforcement institutions may fail to sustain the necessary system of loyalties and generate corruption. A lowered institutional prestige may also undermine collective action sustained on normative behavior, as in the case of the maintenance of common property when traditional cultural systems weaken or in the case of increases in tax evasion due to government loss of credibility. These are ways in which normative behavior affect economic performance. A deeper study of normativity would also account for how economic processes affect normative systems.

Modeling the behavior of the subjects of Milgram's experiment in terms of a normative agent which experiences normative conflict and is capable of different degrees of normative congruence is a first step in the direction of synthesizing *Homo sociologicus* with *Homo economicus*. Although future developments of these ideas may incorporate the logic of semantics more explicitly, this step already provides the tools to analyze problems involving psychological, sociological and economic aspects.

Appendix

We solve the model of Milgram's experiment as set out in the body of the article.

The plane (J_{t+1}, R_{t+1})

The minimization of $D^{Tot}(J_{t+1}, R_{t+1})$ proceeds as follows. Because of the form of the function $D^{Tot}(J_{t+1}, R_{t+1})$, the (J_{t+1}, R_{t+1}) plane is divided into two regions by the line $R_{t+1}I = J_{t+1}X$. If the minimum occurs in the superior region (region 1) $R_{t+1}I \geq J_{t+1}X$ then $\mu = 1$ and the subject obeys. If it occurs in the interior of the inferior region (region 0) then $\mu = 0$ and the subject decides to abandon the experiment. In the interior of region 1 self-worth diminishes as a consequence of the decision, while in the interior of region 0 it is prestige that diminishes. We have privileged permanence in the experiment in the case of equality, so in the boundary of the two regions $\mu = 1$. In this case both self-worth and reputation can diminish. The following formulae can be verified:

$$(J_{t+1}, R_{t+1}) = \left\{ \begin{array}{l} \left\{ \begin{array}{lll} \left(\dfrac{\rho(J_t)}{X^{b_J}}, \dfrac{\rho(R_t)}{DC(I)}\right) & \text{if} & \dfrac{\rho(R_t)\, I}{\rho(J_t)\, DC(I)} \geq X^{1-b_J},\ X \quad 1 \\ \left(\dfrac{\rho(R_t)\, I}{DC(I)\, X}, \dfrac{\rho(R_t)}{DC(I)}\right) & \text{if} & \dfrac{\rho(R_t)\, I}{\rho(J_t)\, DC(I)} \leq \text{Min}\{X^{1-b_J}, X\} \\ \left(\rho(J_t), \dfrac{\rho(R_t)}{DC(I)}\right) & & \text{in the remaining cases.} \end{array} \right\} \text{(region 1)} \\ \\ \left\{ \begin{array}{lll} \left(\rho(J_t), \dfrac{\rho(R_t)}{DC(I)\, I^{b_R}}\right) & \text{if} & \dfrac{\rho(R_t)\, I^{1-b_R}}{\rho(J_t)\, DC(I)} \leq X \\ \left(\rho(J_t), \dfrac{\rho(J_t)\, X}{I}\right) & \text{if} & \dfrac{\rho(R_t)\, I^{1-b_R}}{\rho(J_t)\, DC(I)} \geq X \end{array} \right\} \text{(region 0)} \end{array} \right.$$

Table 1. Table 1. The dynamics of self-worth and prestige.

where $b_J = 1/(1 + a_J)$, $b_R = 1/(1 + a_R)$. Observe that $\rho\left(\dfrac{R_t}{J_t}\right) = \dfrac{\rho(R_t)}{\rho(J_t)} = \left[\dfrac{R_t}{J_t}\right]^{1-\alpha}\left[\dfrac{R_0}{J_0}\right]^{\alpha}$.

The subjects reaction function on the plane (I, X)

According to the results above we define on the (I, X) plane the curves:

$$X_0^{Crit} = \frac{\rho(R_t)\, I^{1-b_R}}{\rho(J_t)\, DC(I)}, \quad X_1^{Crit} = \left[\frac{\rho(R_t)\, I}{\rho(J_t)\, DC(I)}\right]^{\frac{1}{1-b_J}}, \quad X_1^{*} = \frac{\rho(R_t)\, I}{\rho(J_t)\, DC(I)}.$$

There is a region of values on the (I, X) plane for which the minimum total dissonance in region 0 of the (J_{t+1}, R_{t+1}) plane occurs in its interior. These lie above the curve X_0^{Crit}. For points lying below X_0^{Crit} the minimum occurs on the boundary of region 0. Similarily the minimum total dissonance in region 1 of the (J_{t+1}, R_{t+1}) plane occurs in its interior for points lying below X_1^{Crit}. In addition, in this case, if $X \geq 1$ self-worth diminishes while if $X < 1$ it remains at its recuperated level. In the boundary case X is compared with X_1^{*} instead of 1.

If one of the regions 0 or 1 has its minimum dissonance on the boundary while the other has it in the interior, then the interior minimum is the global minimum. When in both regions the minimum is interior the values must be compared. The curve bounding these regions is:

$$X^{Int} = \left[b_J + (1 - b_J)\frac{\rho(R_t)}{\rho(J_t)}\phi(I)\right]^{\frac{1}{1-b_J}} \quad \text{where} \quad \phi(I) = \frac{I^{(1-b_R)} - b_R}{(1 - b_R)\, DC(I)},$$

and $X \leq X^{Int}$ implies permanence. It is unecessary to compare values if both minimae occur on the boundary, since permanence is priviledged in this case.

Summarizing, we have:

Regions of permanence (obeying) and exit (disobeying)	$X > X_0^{Crit}$ in region 0 the maximum dissonance lies in its interior	$X \leq X_0^{Crit}$ in region 0 the maximum dissonance lies on its boundary
$X < X_1^{Crit}$, $X \geq 1$ in region 1 the maximum dissonance lies in its interior	$X \leq X^{Int} \Rightarrow \mu = 1$ (Region Π_1) $X > X^{Int} \Rightarrow \mu = 0$	$\mu = 1$ (Region Π_2)
$X \geq X_1^{Crit}$, $X \geq X_1^*$ in region 1 the maximum dissonance lies on its boundary	$\mu = 0$	$\mu = 1$ (Region Π_3)

Table 2. Regions where the subject obeys or disobeys.

The region of permanence is $\Pi = \Pi_1 \cup \Pi_2 \cup \Pi_3$, where

$$\Pi_1 = \{(I,X) : I \geq 1, X \in [1, X^{Int}] \cap (X_0^{Crit}, X_1^{Crit})\},$$
$$\Pi_2 = \{(I,X) : I \geq 1, X \in [1, X_0^{Crit}] \cap (-\infty, X_1^{Crit})\},$$
$$\Pi_3 = \{(I,X) : I \geq 1, X \in [\max\{X_1^*, X_1^{Crit}\}, X_0^{Crit}]\},$$

(the intervals are understood to be empty if their end points are not in ascending order).

Theorem 1

(1) If two of the functions X_0^{Crit}, X_1^{Crit}, X^{Int} equal 1, so does the third and $I = 1$.

(2) For $I > 1$, $X_1^{Crit} > 1$ implies $X_0^{Crit} < X_1^{Crit}$.

(3) For $I > 1$, $X_0^{Crit} > 1$ implies $X_0^{Crit} < X^{Int}$.

(4) For $I > 1$, $X_1^{Crit} > 1$ implies $X_1^{Crit} > X^{Int}$.

(5) X_1^{Crit} and X_1^* are together greater than, less than or equal to 1.

(6) $\Pi_1 \neq \varnothing \Rightarrow \Pi_1 = \{(I,X) : I \geq 1, X \in [\max\{1, X_0^{Crit}\}, X^{Int}]\}$.

(7) $\Pi_2 \neq \varnothing \Rightarrow \Pi_2 = \{(I,X) : I \geq 1, X \in [1, X_0^{Crit}]\}$ and $X_0^{Crit} < X^{Int}$ so $\Pi_1 \neq \varnothing$.

(8) $\Pi_3 \neq \varnothing$.

(9) $\Pi = \{(I,X) : I \geq 1, X \in [1, X^{Int}]\}$.

(10) Each of the curves X_0^{Crit}, X_1^{Crit}, X^{Int}, X_1^* has a positive derivative at $I = 1$.

(11) X^{Int} and X_0^{Crit} tend to cero as I tends to infinity.

(12) X^{Int} and X_0^{Crit} are quasiconcave.

Proof. Observe the relations

$$X_0^{Crit} I^{b_R} = \left[X_1^{Crit}\right]^{1-b_J} = X_1^*$$

$$\left[X^{Int}\right]^{(1-b_J)} = b_J + (1 - b_J)\left[\frac{1 - b_R I^{-(1-b_R)}}{1 - b_R}\right] X_0^{Crit}$$

$$\left[X^{Int}\right]^{(1-b_J)} = b_J + (1 - b_J)\left[\frac{I^{(1-b_R)} - b_R}{(1 - b_R) I}\right]\left[X_1^{Crit}\right]^{(1-b_J)}$$

(1) This follows trivially.
(2) This follows from the first relation.
(3) Consider the equation

$$X^{1-b_J} = b_J + (1 - b_J)\left[\frac{1 - b_R I^{-(1-b_R)}}{1 - b_R}\right] X.$$

For $I \geq 1$ and $X = 1$ the RHS is ≥ 1. Since it also has the larger power of X, there is no solution with $X > 1$. Thus the RHS is larger than the LHS for $X > 1$. Hence $X_0^{Crit} > 1$ implies $X_0^{Crit} < X^{Int}$ because

$$\left[X_0^{Crit}\right]^{1-b_J} < b_J + (1 - b_J)\left[\frac{1 - b_R I^{-(1-b_R)}}{1 - b_R}\right] X_0^{Crit} = \left[X^{Int}\right]^{1-b}$$

(4) Consider the equation

$$X^{1-b_J} = b_J + (1 - b_J)\left[\frac{I^{(1-b_R)} - b_R}{(1 - b_R)\, I}\right] X^{1-b}.$$

For $I \geq 1$ and $X = 1$ the RHS is less than the LHS and has a smaller coefficient for X, so there is no solution for $X > 1$. Thus the RHS is smaller than the LHS for $X > 1$. Hence $X_1^{Crit} > 1$ implies $X_1^{Crit} > X^{Int}$ because

$$\left[X^{Int}\right]^{(1-b_J)} = b_J + (1 - b_J)\left[\frac{I^{(1-b_R)} - b_R}{(1 - b_R)\, I}\right]\left[X_1^{Crit}\right]^{(1-b_J)} < \left[X_1^{Crit}\right]^{(1-b_J)}.$$

(5) This follows from $X_1^{Crit} = \left[X_1^{*}\right]^{1/(1-b_J)}$.
(6) By (4).
(7) By (2) and (3).
(8) Using (5), if X_1^{Crit} and X_1^{*} are both greater than 1 then (2) implies $\Pi_3 \neq \varnothing$. But if X_1^{Crit} and X_1^{*} are ≤ 1, by the first relation $\max\{X_0^{Crit}, X_1^{Crit}\} \geq X_1^{*}$, which also implies $\Pi_3 \neq \varnothing$.
(9) By (6), (7), (8).
(10) Hypothesis DC3 can be written $DC'(1) < 1 - b_R$. The remainder of the proof is straight-forward calculation.
(11) By hypothesis DC4.
(12) Each of thes functions first increases and then decreases, because if their first derivative is cero their second derivative is negative. The calculation is the following. In the case of X_0^{Crit}, let $h(I) = I^{1-b_R}/DC(I)$.

$$h'(I) = \frac{(1-b_R)I^{-b_R}DC(I) - I^{1-b_R}DC'(I)}{DC(I)^2},$$

$$\left[DC(I)^2\, h''(I)\right]_{h'(I)=0} = -\, b_R\,(1-b_R)I^{-1-b_R}DC(I) - I^{1-b_R}DC''(I)$$

$$< -\, b_R\,(1-b_R)I^{-1-b_R}DC(I) + b_R\, I^{-b_R}\, DC'(I) = 0.$$

In the case of X^{Int},

$$(1 - b_R)\phi'(I) = \frac{(1 - b_R)I^{-b_R}DC(I) - (I^{(1-b_R)} - b_R)DC'(I)}{DC(I)^2},$$

$$\left[DC(I)^2(1 - b_R)\phi''(I)\right]_{\phi'(I)=0} = -\left(b_R(1 - b_R)I^{-1-b_R}DC(I) + (I^{(1-b_R)} - b_R)DC''(I)\right)$$

$$< -\, b_R\left((1 - b_R)I^{-1-b_R}DC(I) - (I^{(1-b_R)} - b_R)\, I^{-1}\, DC'(I)\right) = 0. \blacklozenge$$

The fact that the upper boundary of Π is the curve X^{Int} means that the experimenter can only push the subject to the point at which she would change her attribution of prestige and therefore abandon the experiment, and not to the point at which she would break her normativity. The results on the shape of X^{Int} mean that in every case the experimenter obtains an increment in X when persuasive pressure is comenced, and also that only a bounded level of persuasion is admissible in any period.

Maximization of X

Having established the reaction function of the subject our objective will be to show under what conditions the level of compliance X that the experimenter can obtain after applying a sequence of commands is bounded.

There is a maximum level of compliance X which the experimenter can obtain in the region of permanence in a given period. By Theorem 1, part (9), the maximum level of X is obtained at the maximum of the function X^{Int}. Let $K = \{I \mid \exists X : (I, X) \in \Pi\} \cap \{I \mid X^{Int} \geq 1\}$. Then

$$X^{Int}_{Max} = \sup_{I \in K} X^{Int}(I) = \left[b_J + (1 - b_J)\, \rho\left(\frac{R_t}{J_t}\right) \phi_{max} \right]^{\frac{1}{1-b_J}},$$

where $\phi_{max} = \sup_{I \geq 1} \phi(I)$ We write $I^{Int}_{Max}X$ for the value of I at which this supremum is achieved. The condition for Π to be non-empty is that $X^{Int}_{Max} \geq 1$, which occurs when the initial level of influence satisfies

$$\frac{R_0}{J_0} \geq \phi^{-1}_{max} = Inf_{Min}$$

where the last equality defines the minimum level of initial influence for which there will be a non-empty region of compliance.

To increment the level of compliance, the experimenter can first give a sequence of commands with the purpose of maximizing R_t/J_t, which we call the level of influence, and thus increase X^{Int}_{Max}.

Maximization of influence

We now study when the influence that the experimenter can obtain by a sequence of commands is bounded. The region of permanence is defined by functions which depend on the history of commands only through the level of influence $Inf_t = R_t/J_t$

attained in the last period. The change of influence is described by the following formulae:

$$Inf_{t+1} = \begin{cases} \frac{X^{b_J}}{DC(I)}\rho(Inf_t) \text{ in region } \Pi_1 \\ \frac{1}{DC(I)}\rho(Inf_t) \text{ in region } \Pi_2 \end{cases}$$

Theorem 2. In the case when $\Pi \neq \varnothing$ let $Inf_{t+1}^{Max} = \sup_{I \in K}\left[X^{Int}\right]^{b_J}/DC$ and write I_{MaxInf}^{Sup} for the point at which this supremum is achieved. Thus

$$Inf_{t+1}^{Max} = \frac{1}{DC}\rho_t\left[b_J + (1 - b_J)\,\phi\,\rho_t\right]^{b_J/(1-b_J)},$$

where ϕ and DC are evaluated at I_{MaxInf}^{Sup} and $\rho_t = \rho(Inf_t)$.

(1) Inf_{t+1} is a quasiconvex function of I on $[1, \infty)$, having a unique maximum. If $DC'(1)(1+a_J) \geq 1$ (*i.e.* $DC'(1) \geq b_J$) then $I_{MaxInf}^{Sup} = 1$. If $DC'(1)(1+a_J) < 1$ (*i.e.*), and $DC'(1) < b_J$) we suppose, first, that $Inf_t > Inf_0$ and, second, that the experimenter can increase her influence, the maximum cannot occur on a boundary point of K that is not I =1. It occurs at $I_{MaxInf}^{Sup} = 1$ if $\rho_t \in (Inf_{Min}, b_J DC'(1)/b_J - DC'(1))]$ and at an interior point of K if $\rho_t \in b_J - DC'(1)), \infty)$. Let I_{MaxInf}^{MaxInt} be the solution of the equation

$$b_J\,(1 - b_R)DC(I) - (I - b_R I^{b_R})DC'(I) = 0\,.$$

In the case when Inf_{t+1}^{Max} is achieved in the interior, $1 < I_{MaxInf}^{Sup} < I_{MaxInf}^{MaxInt} < I_{MaxX}^{Int}$. A necessary condition for the experimenter to be able to increase her influence is $\rho_t > Inf_{Min}^{Inc}$ where $Inf_{Min}^{Inc} = \left[\phi(I_{MaxInf}^{SupInt})\right]^{-1} > Inf_{Min}$.

(2) Let

$$f(\rho_t) = \frac{1}{DC}\,Inf_0^{\alpha/(1-\alpha)}\rho_t\left[b_J + (1 - b_J)\,\phi\,\rho_t\right]^{b_J/(1-b_J)},$$

and define the dynamical system

$$\rho_{t+1} = f(\rho_t)^{1-\alpha}.$$

This system generates the fastest trajectory that the experimenter can use to increase her influence.

Suppose that b_J is constant (*i.e.* independent of Inf_t). We have the following cases:

(i) $b_J < \alpha$. The dynamical system is stable. The experimenter can increase her influence up to a certain point (which depends on the initial level of influence), if Inf_0 is large enough to have $\frac{1}{DC}\left[b_J + (1 - b_J)\phi Inf_0\right]^{b_J/(1-b_J)} \Big|_{I_{MaxInf}^{Sup}} > 1$ (a necessary condition is $Inf_0 > Inf_{Min}^{Inc}$). Otherwise the experimenter cannot increase her influence.

(ii) $b_J > \alpha$. The dynamical system is unstable. Given a sufficient inicial influence, greater than the supremum of those influences satisfying equality in case *(i)* the experimenter can gain unbounded influence.

(iii) $b_J = \alpha$. This is a limit case in which one of the possibilities *(i)* or *(ii)* (see the proof for details).

Instead suppose that $a_J = a_J(Inf_t)$ is an increasing function with the following properties, written in terms of b_J: $b_J(Inf_{Min}^{Inc}) > \alpha$, but for sufficiently large Inf b_J decreases beyond α (equivalently $\alpha(1+a_J)$ increases beyond 1). Then given enough initial influence (a necessary condition is $Inf_0 > Inf_{Min}^{Inc}$), the experimenter can increase her influence up to the level at which b_J decreases beyond α. If $Inf_0 \leq Inf_{Min}^{Inc}$ then the experimenter cannot increase her influence.

Proof. By theorem 1, the maximum new influence Inf_{t+1}^{Max} is obtained on the upper bound of the region Π_1, that is, on the upper bound of Π.

(1) Let us first suppose $DC'(I) < b_J$. Then the LHS of the equation for I_{MaxInf}^{MaxInt} is positive at $I = 1$ and is a decreasing function of I, having derivative $\leq -(1 - b_J)(1 - b_R)DC'(I)$ by of Hypothesis DC5. Thus I_{MaxInf}^{MaxInt} is a well-defined number greater than 1. Since the first order condition defining I_{MaxX}^{Int} is $(1 - b_R)DC(I) - (I - b_R I^{b_R})DC'(I) = 0$, it is clear by hypothesis DC3 that $I_{MaxX}^{Int} \geq 1$ and by comparing equations that $I_{MaxInf}^{MaxInt} < I_{MaxX}^{Int}$. Consider now an interior maximum of Inf_{t+1}^{Max}, satisfying the first order condition

$$\rho_t = \frac{b_J I^{b_R}(1 - b_R)DC(I)DC'(I)}{b_J(1 - b_R)DC(I) - (1 - b_R I^{b_R})DC'(I)}.$$

Hypothesis DC5, which is equivalent to $\frac{d}{dI}\left(I^{b_R}DC'(I)\right) > 0$, implies that the RHS numerator is increasing, while the denominator is the function defining I_{MaxInf}^{MaxInt}, which is a positive function decreasing strictly monotonically to 0 on $[1, I_{MaxInf}^{Sup}]$. Thus the RHS is a function increasing from $b_J\, DC'(1)/(b_J - DC'(1))$ to infinity on

$[1, I^{MaxInt}_{MaxInf})$, and the solution to the equation is an increasing function of ρ_t. Moreover, the equation has a unique solution if it exists. This implies that $[X^{Int}]^{b_J}/D$ is a quasiconvex function on $[1, \infty)$, because its derivative cannot change sign more than once and its is eventually negative because the function tends to zero at infinity. Hence if the maximum defining Inf^{Max}_{t+1} is on the boundary of K, it occurs at the lowest level of I for which there is permanence. This occurs either at $X^{Int} = 1$ or at $I - 1$. However, for the experimenter to be able to increase her influence, it is necessary to have $Inf^{Max}_{t+1} = ([X^{Int}]^{b_J}/DC)\, Inf_t^{1-\alpha} Inf_0^{\alpha} \geq Inf_t$, that is, $[X^{Int}]^{b_J}/DC \geq (Inf_t/Inf_0)^{\alpha}$. If $X^{Int} = 1$ the LHS is less than 1, contradicting $Inf_t > Inf_0$. Thus the only boundary solutions are at $I = 1$, and these occur if $\rho_t \leq b_J DC'(I)/(b_J - DC'(1))$ or if $b_J \leq DC'(1)$. Observe also that if the experimenter can increase her influence then

$$X^{Int} = \left[b_J + (1 - b_J)\rho_t \phi(I)\right]^{1/(1-b_J)} > 1 \qquad \text{so}$$

$$\rho_t > \left[\phi(I^{Sup}_{MaxInf})\right]^{-1} > Inf^{Inc}_{Min} > \left[\phi(I^{Int}_{MaxX})\right]^{-1} = Inf_{Min}.$$

(2) In each period the recuperated influence that can be reached for the next period is an increasing function of ρ_t, so the dynamical system generates the fastest trajectory that the experimenter can use to increase her influence. The equation $f(\rho_t)^{1-\alpha} = \rho_t$ which defines the equilibria of the dynamical system is the following:

$$\left[\frac{1}{DC} Inf_0^{\alpha/(1-\alpha)} \rho_t \left[b_J + (1 - b_J)\phi\rho_t\right]^{b_J/(1-b_J)}\right]^{1-\alpha} = \rho_t.$$

At $\rho_t = 0$ the RHS behaves as $\rho_t^{1-\alpha}$ and so grows faster than the LHS. As $\rho_t \to \infty$ the LHS grows comparably to $\rho_t^{(1-\alpha)/(1-b_J)}$. Suppose that b_J is constant. The cases *(i)*, *(ii)* and *(iii)* arise from comparing $(1-\alpha)/(1-b_J)$ with 1.

(i) $b_J < \alpha$. The LHS is eventually less than the RHS, so the dynamical system is stable. The intersection will usually occur at a point but since I^{MaxInt}_{MaxInf} varies, there could conceivably be multiple intersections. If Inf_0 is large enough, the experimenter can increase her influence. The condition is obtained by substituting $\rho_0 = Inf_0$.

(ii) $b_J > \alpha$. The LHS eventually grows more than the RHS, so the dynamical system is unstable. Given a sufficient inicial influence (which must be greater than Inf^{Inc}_{Min}) the experimenter can gain unbounded influence. The condition is that the initial influence be greater than the supremum of those influences satisfying equality in condition *(i)*.

(iii) $b_J = \alpha$. This is a limit case in which one of the possibilities *(i)* or *(ii)* will prevail. If $\lim_{I \to (I^{SupInt}_{MaxInf})^-} [DC(I)]^{-(1-b_J)/b} \left[b_J + Inf_0(1 - b_J)\phi(I)\right] < 1$ then behavior is as in case *(i)*, if it is > 1 as in case *(ii)*, and otherwise it depends on the exact form of DC (and therefore ϕ).

The non-linear case is analized similarily. ◆

Observe that in the cases in which the attainable influence is bounded, the levels of persuasion $I^{Int}_{Maxinf}(\rho_t) < I^{Sup}_{Maxinf}(\rho_t) < I^{Int}_{MaxX}$ which are optimal for increasing it are bounded away from the level which is optimal to extract the maximum compliance. Thus after dedicating a number of periods of time to approximate the maximum influence, the experimenter can give one last command to extract the maximum compliance, knowing that the following her influence will be less.

It is worth noting that this analysis includes the case $\alpha = 0$. In this case $b_J > 0$ can only tend to 0 so the experimenter can increase her influence without bound.

In the cases in which the experimenter can increment her influence the optimal level of persuasion will increase with her influence while the change in b_J remains small. This reproduces the increasing level of persusion and compliance observed in the experiment.

Finally, since the prestige of the experimenter is bounded by her original prestige, we obtain the same qualitative behavior for $a_J = a_J(Inf_t)$ as we could obtain for $a_J = a_J(J_t)$. We consider the former, though, because in this model (in which α is it the same for prestige and self-worth) we can treat the dynamics in terms of the single variable Inf_t rather than in two variables R_t, J_t. The mathematical complexity would also increase if we considered the dependence in the form $a_J = a_J(Inf_{t+1})$ or $a_J = a_J(J_{t+1})$.

References

Akerlof, G. (1982b), "Labor contracts as partial gift exchange", *Quarterly Journal of Economics*, vol. 97, pp. 543-569.

_____ (1984), "Gift exchange and efficiency-wage theory: Four views", *American Economic Review*, vol. 74, pp. 79-83.

_____ (1984), "An economic theorist's book of tales", *Essays that entertain the consequences of new assumptions in economic theory*, Cambridge, Cambridge University Press.

_____ (1991), "Procrastination and obedience", *American Economic Review*, vol. 81, pp. 1-19.

Akerlof, G. and W. Dickens (1982 a), "The economic consequences of cognitive dissonance", *American Economic Review*, vol. 72, pp. 307-319.

Anderson, G. M. and W. Block (1995), "Procrastination, obedience, and public policy: The irrelevance of salience", *American Journal of Economics and Sociology*, vol. 54, no. 2, .pp. 201-205.

Andreoni, J. (1995), "Cooperation in public-goods experiments: Kindness or confusion?", *American Economic Review*, vol. 85, pp. 891-904.

Aronson, E. (1969), "The theory of cognitive dissonance: A current perspective", in L. Berkowitz (ed.), *Advances in experimental social psychology*, vol. 4, New York, Academic Press.

_____ (1980), *The social animal*, San Francisco, W.H. Freeman.

Aronson, E. and D. Linder (1965), "Gain and loss of esteem as determinants of interpersonal attactiveness", *Journal of Experimental Social Psychology*, vol. 1, pp. 156-171.

Aronson, E. and D. Mettee (1968), "Dishonest behavior as a function of different levels of self-esteem", *Journal of Personality and Social Psychology*, vol. 9, pp. 121-127.

Berkowitz, L. (1986), "A survey of social psychology", in Holt, Rinehart and Winston (eds.), USA.

Canon, L. (1964), "Self-confidence and selective exposure to information", in L. Festinger (ed.), *Conflict decision and dissonance*, Stanford, Stanford University Press.

Coopersmith, S. (1967), *The antecedents of self-esteem*, San Francisco, W.H. Freeman.

Deutsch, M. and R. Krauss, (1990), *Theories in social psychology*, New York, Basic Books.

Elster, J. (1987), *Sour grapes: Studies in the subversion of rationality*, Gran Bretaña, Cambridge University Press.

_____ (1989), "Social norms and economic theory", *Journal of Economic Perspectives*, vol. 3, pp. 99-117.

_____ (1996), "Rationality and the emotions", *The Economic Journal*, vol. 106, pp. 1386-1397.

Faucheux, C. and S. Moscovici, (1968), "Self-esteem and exploitative behavior in a game against chance and nature", *Journal of Personality and social Psychology*, vol. 8, pp. 83-88.

Festinger, L. (1957), *A theory of cognitive dissonance*, Stanford, Stanford University Press.

Flanagan, O. (1995), "Situations and dispositions", in A. I. Goldman (ed.), *Readings in philosophy and cognitive science*, Cambridge, Mass., MIT Press.

Fromm, E. (1974), *The anatomy of human destructiveness*, New York, Holt, Rinehart and Winston.

Fuchs, V. (1996), "Economics, values, and health care reform", *American Economic Review*, vol. 86, pp. 1-24.

García-Barrios, R. and D. Mayer-Foulkes (1995), *Justice and efficiency in economic relations: Explaining collaboration and conflict in the firm and choice in ultimatum games*, México, CIDE.

Granato, J., R. Inglehart and D. Leblang (1996), "The effect of cultural values on economic development: Theory, hypotheses, and some empirical tests", *American Journal of Political Science* , vol. 3, pp. 607-631.

Heider, F. (1958), *The psychology of interpersonal relations*, USA, Wiley and Sons.

Helmreich, R. and B.E. Collins (1968), "Studies in forced compliance: Commitment and magnitude inducement to comply as determinants of opinion change", *Journal of Personality and Social Psychology*, vol. 10, pp. 75-81.

Hollander, E. P. (1971), *Principles and methods of social psychology*, Oxford, Oxford University Press.

Kahneman, D., J. L. Knetsch and R. Thaler (1986), "Fairness as a constraint on profit seeking: Entitlements in the market", *American Economic Review*, vol. 76, pp. 728-741.

Leonard, F. (1975), "Un modelo del sujeto: el equilibrio de Heider", in S. Moscovici (ed.), *Introducción a la psicología social*, Barcelona, Editorial Planeta.

MacIntyre, A. (1985), *After virtue: A study in moral theory*, London, Duckworth.

Marsh, H. W. (1996), "Positive and negative global self-esteem: A substantively meaningful distinction or artifactors?", *Journal of Personality and Social Psychology*, vol. 70, pp. 810-819.

Merton, R.K. (1957), *Social theory and social structure*, Illinois, Free Press.

Milgram, S. (1963), "Behavioral study of obedience", *Journal of Abnormal and Social Psychology*.

_____ (1965 a), "Some conditions of obedience and disobedience to authority", *Human's Relations*, vol. 18, pp. 57-76.

_____ (1965 b), "Liberating effects of group pressure", *Journal of Personality and Social Psychology*, vol. 1, pp. 127-134.

Moscovici, S. (1975), "El hombre en interacción: máquina de responder o máquina de discurrir", in S. Moscovici (ed.), *Introducción a la psicología social*, Barcelona, Editorial Planeta.

Moscovici, S. and P. Ricateau (1975), "Conformidad, minoría e influencia social", in S. Moscovici (ed.), *Introduccióon a la psicología social*, Barcelona, Editorial Planeta.

Rabin, M. (1993), "Incorporating fairness into game theory and economics", *American Economic Review*, vol. 83, pp. 1281-1302.

Ross, L. (1988), "Situationist perspectives on the obedience experiments", *Contemporary Psychology*, vol. 33, pp. 101-104.

Scher, S. J. and J. Cooper (1989), "Motivational basis of dissonance: The singular role of behavioral consequences", *Journal of Personality and Social Psychology*, vol. 56, pp. 899-906.

Sears, D.O., L.A. Peplau and S.E. Taylor (1991), *Social psychology*, New Jersey, Prentice Hall.

Sethi, R. and E. Somanathan (1996), "The evolution of social norms in common property resource use", *American Economic Review*, vol. 86, pp. 766-787.

Sherif, M. and C. Sherif (1969), *Social psychology*, New York, Harper and Row.

Tapp, J. L., M. Gunnar and D. Keating (1983), "Socialization: Three ages, three systems of rules", in D. Perlman and C. Cozby (eds.), *Social Psychology*, USA, Holt, Reinman and Winston.

Wicklund, R. and J. Brehm (1976), *Perspectives on cognitive dissonance*, New Jersey, Lawrence Erlbaum.

Zimbardo, P.G. (1975), "La psicología social: una situación, una trama y una escenificación en busca de la realidad", in S. Moscovici (ed.), *Introducción a la psicología social*, Barcelona, Editorial Planeta.

Zimbardo, P.G., E.B. Ebbesen and Ch. Maslach (1977), *Influencing attitudes and changing behavior*, Massachusets, Addison-Wesley.