Taboos: Considering the Unthinkable

November 2008

Chaim Fershtman[*]
Uri Gneezy[**]
Moshe Hoffman[***]

Abstract: A taboo is an "unthinkable" action, that is, even the thought of violating it triggers social punishment. Taboos are the social "thought police," discouraging individuals from considering certain type of actions. We consider a simple model in which taboos are part of the definition of one's identity. Deliberating over breaking the taboo adds the action to the individual's choice set and provides information on possible private benefits but is costly because it contradicts one's identity. The strength of the taboo is endogenously determined by the number of individuals that obey it without any consideration of its violation. We model stable taboos and examine how they can change and disappear over time as a result of changes in the distribution of private benefits gained from its violation. We assume that individuals are heterogeneous with respect to their attitudes towards social punishment. We then analyze the relationship between social heterogeneity and the strength as well as effectiveness of taboos, i.e., are taboos stronger in homogenous or heterogeneous societies? We extend our analysis and examine societies in which individuals may choose among several identities, characterized by different taboos or varying strengths of taboos. Having such a choice defines an evolutionary process with respect to identity: Some identities disappear while others flourish. We examine the characterization and the conditions giving rise to a multi-identity society.

[#]We thank Sagit Bar-Gill for excellent research assistance.
[*] The Eitan Berglas School of Economics, Tel Aviv University.
[**] Rady School of Management, University of California, San Diego.
[***] Graduate School of Business, University of Chicago.

1. Introduction

An important part of the characterization of societies is the different social norms and taboos that govern their members' behavior. In some societies, eating beef is a taboo; in others, eating pork is a taboo; while in many other societies eating dog meat is a taboo. Numerous taboos relate to sexual activity, with many taboos common to most societies. Trade in human organs is another example of a common taboo (see Roth (2007)). Yet, the list of taboos is not necessarily stable. Some taboos may weaken over time or even disappear, while others may become stronger and more dominant.[1] Taboos can be repugnant and appalling actions or behavior.[2] Other types of taboos involve "taboo tradeoffs" that can entail putting a monetary value on "sacred" values like life, friendship, religion, etc.

What is the difference between a taboo and a social norm? One interpretation states that a taboo is a strong social norm. Every time an individual's behavior diverges from a norm, this act impacts on the other members of society, who then punish the deviant individual (see for example Akerlof, 1976, 1980). Hence, a taboo is a strong social norm that is also supported by severe socially imposed punishment. On the other hand, taboos are sometimes referred to as doing the "unthinkable". Under this interpretation, even thinking about deviating from a taboo is problematic. The sanctions associated pertain not just to the behavior that contradicts the taboo but also merely thinking or considering such behavior (see Fiske and Tetlock (1997) and Benabou and Tirole (2006) for a similar interpretation of taboos)[3]. A taboo is therefore a form of "thought police" that governs not just our behavior but also our thoughts. The economic consequences of not being able to consider the act are

[1] For an example of the disappearance of a taboo see Zelizer (1978, 1981), who describes how child insurance and life insurance were first taboo and then became acceptable.

[2] Some of the repugnance and appalling actions as well as markets/tradeoffs are culturally specific and may change over time. For many examples and discussion see Haidt, Rozin, McCauley and Imada (1997) and Roth (2007). For a discussion of taboo tradeoffs, see Fiske and Tetlock (1997) and Tetlock, Kristel, Elson, Green and Lerner (2000)).

[3] If an act is unthinkable it is not necessarily the case that the act of thinking about the act is itself unthinkable. For example, suppose the reader found out his friend was considering committing incest. The reader might recognize this to be disgusting. However, the reader is not disgusting for having come to such recognition, even though the reader has thought about the act of thinking about incest. In fact, it seems the permissibility of meta-thinking is a necessary component for the functioning of taboos. How else would we be able to sanction others for thinking about the unthinkable? How else could we teach our children what taboos our society upholds?

twofold: one needs to think about the act for it to be in one's choice set and one needs to think about the act in order to determine the exact costs and benefits of the act.

Setting a penalty for thinking about something is puzzling[4]. What advantage may society obtain if individuals do not think about or consider certain actions or behaviors? Why should there be a penalty for considering an action if one does not actually engage in that action? One explanation is that sometimes thoughts per se create negative externalities. For example, consider the taboo against eating dog meat: Not only do we not do it, we also do not contemplate doing it and most of us do not want to associate with people who do so or consider doing so. Some of our readers may object or be offended from the very discussion of such an example, which clearly reflects the act's status as a taboo. Also, what benefits emerge from having such a taboo? Perhaps it is the other side of the coin of having dogs as man's best friends. It seems that one cannot survive without the other. That is, it would be difficult to define a dog as a friend when your other friends consider eating it. In this case, the taboo provides some public benefit or public good – dogs as friends – that all members of society can share.[5]

An alternative interpretation is that taboos that penalize consideration of an action are effective instruments for deterring certain types of actions. Consider an action that yields uncertain private benefits. Thinking about such an action is a necessary part of the process by which individuals learn about the possible private benefits they can obtain from deviating from the taboo. In particular, consider a situation having a low probability of obtaining high private benefits. Social norms penalizing the actions only are ineffective as deterrence in such cases. Only high social penalties would be sufficient to deter deviation by an individual who knows beforehand the private benefits to be gained. On the other hand, taboos that penalize for merely considering a deviation may impose a much lower penalty that is nonetheless sufficient to deter individuals from thinking about the option to deviate. To illustrate this structure, consider the Hindu taboo of not eating beef. When

[4] Henceforth, we restrict our discussion of taboos to actions for which there is a non-legal penalty for thinking about the action. For instance, kosher dietary restrictions are not included in our discussion of taboos; while it is prohibited to eat pork, it is nevertheless acceptable to think about eating pork. Additionally, legal prohibitions against incest would not be included in our discussion of taboos; while such laws may have initially resulted from socially regulated taboos, the law now takes on a life of its own, which is beyond the scope of this paper.

[5] Another example is a taboo prohibiting a direct payment for human organs for transplanting which may benefit a society by eliminating incentives for a violent activity and other types of abuse that may be a result of such a trade.

someone is starving, the private benefit of deviating from this taboo is large. Thus the only way to deter such a behavior is to impose a taboo such that people will not even think about this possibility.[6] However, this example also demonstrates that taboos may have higher costs for a society as individuals may forgo large private benefits – in this case, staying alive – by not violating a taboo.

One type of social punishment involves the attitudes and reactions of other of society's members. For such social punishment to be effective, behavior must be (partly) observable. How can someone be punished for having "dirty" thoughts? Thoughts are not observable. But there is another type of social punishment, one that is self-inflicted. Taboos are an important part of any social identity (see Akerlof and Kranton (2000) for a discussion on social identity). Adopting an identity implies accepting the taboos associated with this identity. As in Benabou and Tirole (2006), the relationship between identity and taboos introduces a cost of thinking or considering deviation from a taboo. The desire to acquire an identity and to view oneself as a moral person as defined by one's identity is important considerations that define the self-inflicted cost of thinking about violating a taboo. As Haidt et al. (1997) and Fessler and Navarrete (2003) argue, these costs may involve negative emotions such as fear or disgust.

Taboos have an important aspect of social interaction. The benefits from a taboo – like the severity of the social punishment enforcing it – depend on the behavior of members of the society. A taboo's strength affects an individual's incentives to deviate from it or to consider such a deviation. On the other hand, the percentage of individuals who deviate or think about deviation affects the strength of the taboo (see Fessler and Navarrete (2003); for a similar argument with respect to social customs see Akerlof (1980) and Romer (1984)).

As opposed to similar models with respect to social norms, this model assumes that the strength of a taboo is affected also by the percentage of people that consider deviating from it despite the fact that such considerations or any such thoughts are not directly observable.[7] We do not specify the underlying social interaction process that establishes this relationship. It is possible that with some

[6] For most of the readers the taboo against eating human flesh is probably more convincing.
[7] As oppose to a model in which the strength of a taboo is only a function of the percentage of people that actually deviate from it.

probability individuals' thoughts are transparent.[8] Or that the strength of the taboo is determined by social interaction among individuals that may express their opinions and those that consider deviating from a taboo contribute to its weakening.

In this paper we present a simple model that formalizes the concept of a taboo and endogenously determines its strength. We consider a society in which there is one taboo. Behaving in a manner that violates the taboo or considers deviation from the taboo is costly; the severity of the punishment is positively related to the strength of the taboo. We further assume that individuals may differ with respect to their social concerns and the severity that they associate with social punishment (for evidence on individual variations in susceptibility to emotional of fear and disgust in reaction to taboo stimuli see Fessler and Navarrete (2003)).

Taboos that do not conflict with self interest do not need to be supported by punishment. People deviate from a taboo only because they may receive a potential private benefit from doing so; this benefit creates the temptation for the deviation. Without foregoing some private benefit, the presence of a taboo is redundant as no one would think of or wish to deviate from it. We assume that private benefits are stochastically distributed such that considering deviation from a taboo is a learning process by which an individual becomes aware of his/her own potential private benefit. Deviation from a taboo is possible only after an individual has completed the learning process.

We assign to each taboo a measure of its strength, determined by the percentage of individuals that follow the taboo without considering it together with the percentage of individuals that consider but do not violate it. We assume that the public benefit from a taboo and the severity of the social punishment depends on the taboo's strength. Thus, an individual's behavior is affected by the strength of the taboo, which is an outcome of individual's decision to consider deviation or actually deviate from the taboo. We define a stable taboo as one outcome of such a process. We then investigate the effect of changes in the private benefits culled from deviation from the stable taboo while demonstrating the possibility of a non-continuous adjustment process by which a relatively small increase in private benefits destroys the taboo.

[8] In some cases considering an act may require engaging in an actual search, which is not simply a mental task, but involve fact finding which is partly observable.

Societies may differ in their level of social heterogeneity. One aspect of such heterogeneity is the distribution of attitudes towards social norms and social punishment. Another outcome of such heterogeneity is the range of different identities, different taboos or multiple identities. Since taboos are affected by social interaction, social heterogeneity may affect the stable taboos.

We first consider heterogeneity with respect to social preferences. We introduce a measure of heterogeneity that reflects the distribution of individual concerns regarding social punishment. In our model, the effect of social heterogeneity on stable taboos depends on the type of private benefits to be obtained. When the potential private benefits from deviation are relatively small, then increasing the level of homogeneity implies a weaker stable taboo with more people considering deviation. Alternatively, when the potential private benefits from deviation are relatively large, a higher degree of homogeneity leads to a stronger stable taboo, with a smaller percentage of the population considering deviation from the taboo.

We then consider a society in which there are (potentially) two identities. Each identity is characterized by only one taboo. Individuals need to first choose which identity to adopt and then whether they are going to follow the taboo associated with this identity or to consider a deviation from it. Allowing individuals to choose defines an evolutionary process with respect to identities. Some identities may disappear while others may flourish when more individuals choose to adopt them. We specify the conditions that give rise to a two-identity society and investigate its structure.

The two-identity society has an interesting structure in our model. There is one relatively homogenous group that adopts the identity with the weaker taboo. These individuals moderately care about their identity; they all consider deviating from the taboo. The second group is heterogeneous with respect to their social concerns and adopts the identity with the strong taboo. This group consists of two subgroups, one of which consists of individuals with weak social concerns who consider deviating from the taboo while the other consists of individuals with strong social concerns who are not going to consider any deviation from the taboo. This structure provides a framework that enables integration of population mobility with formation of identity.[9]

[9] For example, the effect of immigration on the formation and shaping of identities in a society.

2. Taboo

We consider a society that has one taboo. We normalize the size of the population in this society to 1 and assume that individuals are heterogeneous with respect to their concern for social punishment. We denote the individual's type as ϕ, letting the distribution of ϕ be $G(\phi)$, defined over the support $[\bar{\phi},\underline{\phi}]$. An individual of type ϕ=0 is concerned with neither his identity nor the social implications of his actions. A higher ϕ implies higher social punishment/identity concerns.

Not all the taboos are of the same strength, with the strength of the taboo endogenously determined by the percentage of people that accept the taboo. We thus assume that the strength of a taboo is a decreasing function of the percentage of people that consider breaking the taboo and of the number of people that actually break it. Letting

- N_c = the percentage of people that consider deviating from the taboo.
- N_d = the percentage of people that actually deviate from the taboo.

We denote the strength of the taboo as $T(N_c, N_d)$ where $T(\cdot,\cdot)$ is declining in both arguments.

When the taboo is stronger, the social cost of deviating or of considering deviation from it is higher. These costs are also a function of the individual's type, ϕ, as there are individuals who are more sensitive to social pressures and punishments while others are less so. Specifically, let

- $C(\phi,T)$ = the cost associated with considering a deviation from the taboo
- $D(\phi,T)$ = the cost associated with deviating from the taboo.

Individuals in this society benefit from having the taboo. These benefits may be a function of the number of people that "maintain" the taboo as well as a function of the number of people that do not even consider deviating from it. Since the strength of the taboo is a function of these two variables, we assume that the social benefit from a taboo is an increasing function of its strength. These benefits may be interpreted differently. They derive from simple affiliation with a group having a certain identity or, alternatively, the standard externalities affected by the taboo itself, the number of people that maintain it and the number that consider deviating from it. The taboo's overall public benefit to the members of society is denoted by E(T); it is

assumed that these benefits are shared by all members of the society and can thus be viewed as a public good.

A taboo has meaning only if there are private benefits attached to deviating from it. We therefore assume that people may obtain private benefits from deviating from the taboo. However, these benefits are not apparent without individuals first considering the deviation. This consideration process is also a learning process in which the private benefits of deviation are revealed.

An individual who considers deviation from a taboo will observe the realization of his own private benefits from such a deviation. The private benefit, denoted by b, is privately observed whenever an individual is considering a deviation from the taboo. The distribution of b is given by *F(b)*, and is commonly known. We assume for convenience that all individuals have the same distribution of private benefits.

After observing the realization of private benefits an individual may decide whether he/she wishes to deviate from the taboo. But this deviation is costly; therefore, deviation from the taboo will occur only when the realized benefits are greater than the costs of deviation, i.e., whenever $b \geq D(\phi,T)$.

Considering deviation from a taboo is similar to buying an option, that is, the option to deviate from a taboo whenever such a deviation is beneficial. In order to obtain such an option, the individual pays the consideration costs $C(\phi,T)$. The individual then observes a realization of the random benefit *b* and, given his/her type, he/she decides whether to deviate or not. The value of such an option for an individual of type ϕ is denoted by $V(F(b),\phi,T)$. An individual will consider deviating from the taboo only when the value of this option is greater than the cost of acquiring it, which is the cost of considering violation of the taboo, i.e., whenever $V(F(b),\phi,T) \geq C(\phi,T)$ [10].

Clearly, some individuals who deviate from a taboo regret doing so. When an individual decides on deviation, the cost of considering a deviation, *C(ϕ,T)*, is a sunk cost. Thus, whenever $C(\phi,T) + D(\phi,T) > b > D(\phi,T)$, the individual will deviate from the taboo but regret the fact that deviation was even considered.

[10] The fact that individuals are able to costlessly calculate whether or not it is worth buying this option rests on the assumption that it is permissible to think about thinking about a taboo. Likewise, this assumption is necessary for us to permit punishment, of oneself or an other, without violating the taboo. See footnote 3 for a discussion of this assumption.

Definition: A **Taboo** is defined by $\{T^*,\ N_C^*, N_D^*\}$ such that

- $T^* = T(N_C^*, N_D^*)$.
- $N_C^* = \#\ \{\phi \mid \mathrm{V(F(b)}, \phi, T^*) \geq C(\phi, T^*)\}$ is the expected percentage of individuals for whom the value of the option of deviating from the taboo is greater than the cost of considering such a deviation.
- $N_D^* = \int_{\{\phi \mid \mathrm{V(F(b)}, \phi, T^*) \geq C(\phi, T^*)\}} (1\text{-F}(D(\phi, T^*))) \, g(\phi) \, d\phi$ is the expected percentage of individuals that actually deviate from the taboo. These individuals have considered deviation and realized a private benefit that is above their cost of deviation, i.e., $b \geq D(\phi, T^*)$.

Dynamic adjustment process: In order to consider the stability of a taboo, we define a dynamic adjustment process as follows: Starting from any (N_C^1, N_D^1), the corresponding strength of the taboo is $T^1 \equiv T(N_C^1, N_D^1)$. Given T^1, we can find the optimal number of individuals that consider deviation from the taboo, $N_c(T^1)$, and those who actually deviate, $N_D(T^1)$. Define $(N_C^2, N_D^2) \equiv (N_C(T^1), N_D(T^1))$. We can now define $T^2 = T(N_C^2, N_D^2)$ as the taboo's strength at the second iteration. We can proceed in the same manner to define the dynamic adjustment sequence $\{T^k(N_C^{k-1}, N_D^{k-1}); (N_C^k(T^{k-1}), N_D^k(T^{k-1}))\}$.

A **stable taboo** is a taboo that has the following property: For every (N_C, N_D) in the neighborhood of (N_C^*, N_D^*), the dynamic adjustment process converges to (N_C^*, N_D^*).

The cost of a taboo: Maintaining a taboo implies that people will not take certain actions that may benefit them. We define the social cost of a taboo, denoted as SC(T), as the expected unexploited benefits for the types of individuals who do not consider deviation plus the sum of positive benefits foregone by those who consider deviation but do not deviate whenever the benefits from deviation are smaller than the cost. Hence,

$$SC(T) = (1-N_c)\int_0^{\infty} bf(b)db + \int_{\{\phi|V(F(b),\phi,T^*)\geq C(\phi,T^*)\}} g(\phi)\, F(D(\phi,T)) \int_0^{D(\phi,T)} bf(b)\, db d\phi \quad .$$

Our setup distinguishes between social norms and taboos. Social norms are actions that are subject to social sanctions or punishments. Any deviation from a social norm entails a social punishment. Taboos, on the other hand, "prohibit" even thinking about deviations. Such thoughts results in penalties that may be self-inflicted. This difference between social norms and taboos determines their effectiveness. Our setup is such that an individual may learn about the realization of private benefits before he has taken the required actions. Thus, when private benefits are such that with a small probability there are large private rewards, the effectiveness of social norms is limited. For social norms to be effective and block deviation, the social punishments for deviation should exceed the large private benefit.[11] For a taboo to be effective, it is sufficient for the cost of considering deviation to exceed the expected private benefit.

3. Stable Taboos and Private Benefits

3.1 A Simple Setup

Consider a simpler version of the above model. Specifically, let

- $C(\phi,T) = \lambda T\phi$ = the cost of considering deviation from a taboo.
- $D(\phi,T) = \delta T\phi$ = the cost of deviating from a taboo; $\delta > \lambda > 0$.
- $T(N_c, N_d) = \alpha_c(1-N_c) + \alpha_d(1-N_d)$ = the strength of the taboo, where $\alpha_d > \alpha_c$.
- $G(\phi)$ = the distribution of types in the population, assumed to be uniformly distributed on [0,1],
- The benefits from deviating from the taboo = b with probability *q* and 0 with probability (1-*q*).
- The public benefits of deviation are *E(T)=ET*

where λ, δ, α, $\mu > 0$.

[11] Large punishments are sometimes problematic and not effective.

3.2 Stable Taboos

Consider the behavior of an individual of type ϕ who belongs to a society that maintains a taboo of strength T. If the individual has already considered violating the taboo, he will violate it whenever the benefit from doing so is greater than the associated cost, i.e., whenever $b \geq \delta T\phi$. The cost of considering deviation is $\lambda T\phi$; if an individual does not plan to deviate from the taboo upon obtaining a positive realization of his/her private benefits, there is no reason for him/her to consider such a option. Thus, the value of the option to deviate from the taboo is

$$(1) \qquad V(F(b),\phi,T) = q[b - \delta T\phi].$$

An individual will consider deviating from the taboo only when the value of the option $V(F(b), \phi, T)$ is greater than the cost of deviation, i.e., whenever

$$(2) \qquad q[b - \delta T\phi] \geq \lambda T\phi$$

Since the cost of considering a deviation as well as the cost of deviation itself are increasing in ϕ, there is a threshold $\phi_c(T,b,q)$ such that only individuals of type $\phi \leq \phi_c$(T,b,q) will consider deviation. Using (2), ϕ_c is defined as follows:

$$(3) \qquad \phi_c(T,b,q) \equiv Min\left\{\frac{qb}{(\lambda + q\delta)T}, 1\right\}.$$

The strength of the taboo T is endogenously determined by the proportion of the population that considers breaking it together with those that actually break it. Letting ϕ_c be the proportion of individuals that consider breaking the taboo, then $q\phi_c$ will be the proportion of individuals that actually break it.[12] $T(\phi_c)$, the strength of the taboo, is therefore given by:

$$(4) \qquad T(\phi_c) = \alpha_c(1-\phi_c) + \alpha_d(1-q\phi_c) = (\alpha_c + \alpha_d) - (\alpha_c + q\alpha_d)\,\phi_c \quad .$$

[12] Since proportion q of the individuals that consider deviating actually deviate (after observing a positive private benefit realization), we get that $N_d = q\phi_c$.

An individual of type ϕ=0 will always consider deviating as such an individual does not suffer from any costs associated with violating or considering violation of the taboo. The presence of individuals of this type implies a setting in which there is always some proportion of the population that violates the taboo.[13]

Our definition of a taboo is therefore a couple (T, ϕ_c) that satisfies equations (3) and (4). We depict these two conditions in a $(T \times \phi_c)$ space (see Figure 1). Condition (3) describes the percentage of individuals who consider deviation as a function of T (the taboo's strength) while condition (4) describes the strength of the taboo as a function of the percentage of people that consider deviation, ϕ_c. Condition (3) is a convex declining curve while condition (4) is a linearly declining line.

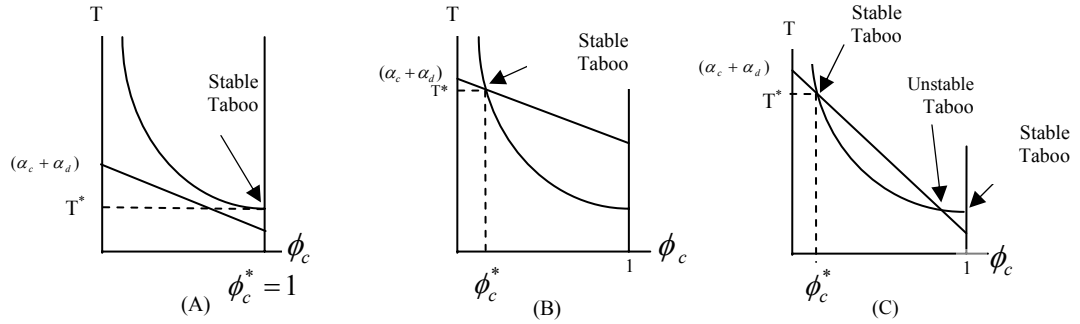

Figure 1: Stable and unstable taboos in the simple setting.

Three possible scenarios are depicted in Figure 1:

Case (A): When there is no intersection between conditions (3) and (4), the only solution is no taboo (ϕ_c^*=1), i.e., all individuals consider deviation. This is a stable situation. If there are some individuals who do not consider deviating from the taboo, the dynamic adjustment process will converge back to ϕ_c^*=1 since the strength of the taboo is sufficiently low to induce these individuals to change their behavior and consider violating the taboo after all.

Case (B): There is only one intersection between conditions (3) and (4). This point, denoted as (T^*, ϕ_c^*), is a stable taboo.[14] In terms of the dynamic adjustment process,

[13] Whenever there are fixed costs of deviation (independent of the type) or when the distribution of ϕ is such that $\phi > \phi_{\min} > 0$, we may obtain stable taboos that no one considers violating.

[14] Observe that the point ϕ_c=1 is not a stable taboo.

the curved line describes the adjustment function $\phi_c^t(T^{t-1})$ while the straight line describes $T^t(\phi_c^{t-1})$, i.e., the strength of the taboo at period t as a function of the percentage of people that consider deviating in the previous period. Suppose that at period t, $\phi_c^t > \phi_c^*$. Using the dynamic adjustment process we can define $T^{t+1}(\phi_c^t)$ and to obtain that $\phi_c^{t+2}(T^{t+1}(\phi_c^t)) < \phi_c^t$, with the adjustment process converging to (T^*, ϕ_c^*).

Case (C): There are two intersections between conditions (3) and (4) and three possible situations that we need to consider. When the intersection on the left has the same properties as the intersection in case (B) and therefore defines a stable taboo, the second intersection is not a stable taboo. The point $\phi_c=1$ has the same properties as described in case (A) and therefore defines a second stable taboo.

Note that the social cost of the taboo is defined by the private benefits that are foregone from maintaining such a taboo, i.e.,

$$(5) \qquad SC(T)=(1-\phi_c(T))qb \ .$$

3.3 The Effect of Greater Private Benefits on Taboos

Taboos change over time; some become stronger while others disappear. Part of this process is clearly a result of social as well as demographic changes. But taboos may also change as a result of variations in the distribution of private benefits. New inventions and ideas as well as new opportunities may imply higher private benefits. Having a higher *b* affects the incentives available to individuals should they deviate from the taboo or even just consider such a deviation. Greater private benefits therefore lead to more deviations and a weaker taboo, which in turn encourages further deviations. Such positive feedback implies that the process of weakening the taboo is not necessarily continuous. We may therefore reach a critical level of private benefits that abolishes the taboo.

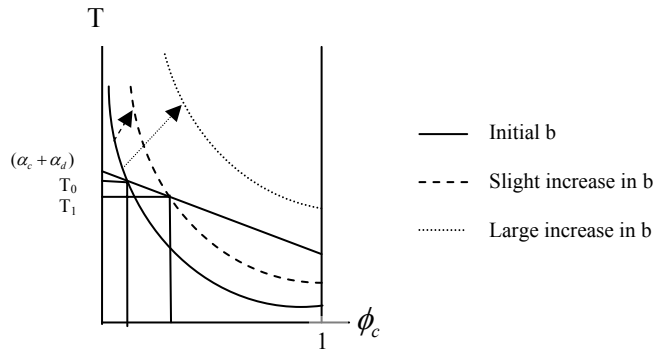

Figure 2: Effect of increasing b on stable taboos.

The effect of a higher *b* is described in Figure 2. Changing *b* does not affect condition (4) as it describes the strength of the taboo as a function of individual behavior. Condition (3) describes the incentives to deviate as a function of the taboo's strength. These incentives increase with the private benefits gained from deviation. It is evident from Figure 2 that when *b* increases, the intersection points between the two curves shift to the right, which implies a weaker taboo as more people consider deviation.

Weakening the taboo is not necessarily a continuous process. There is a critical level of b, denoted as $\hat{b}$, such that whenever $b > \hat{b}$, condition (3) will be above condition (4) (i.e., moving from case B to case A in Figure 1). The critical $\hat{b}$ is given by

$$(6) \quad \hat{b} \equiv \frac{(\alpha_c + \alpha_d)^2(\lambda + q\delta)}{4q(\alpha_c + q\alpha_d)}.$$

This point also defines the weakest taboo that may still hold:

$$(7) \quad T^* = \frac{\alpha_c + \alpha_d}{2}; \phi_c^* = \frac{\alpha_c + \alpha_d}{2(\alpha_c + q\alpha_d)}$$

Clearly, the necessary condition for such discontinuity is that $\phi_c^* < 1$, which will be satisfied whenever $q > \frac{\alpha_d - \alpha_c}{2\alpha_d}$. When q is below this level, the adjustment of the taboo's strength is continuous until the taboo slowly disappears. Only when b is sufficiently large is the process discontinuous. We can thus conclude the following:

Proposition 1: A higher *b* (potential private benefit) implies an erosion of the taboo's strength, i.e., a lower T^*, with more people considering deviating (a higher ϕ_c). The process of taboo erosion is not necessarily continuous. When $q > (\alpha_d - \alpha_c)/2\alpha_d$, a small increase of *b* above $\hat{b}$ eradicates the taboo and induces all members of society to consider deviation.

Proof: See Appendix 1.

As in the case of other models of multiple stable equilibria, this model also contains situations in which temporary changes in the private benefits may cause permanent changes in the taboo's strength. Suppose for example that the starting point is case (c) (see Figure 1), with the stable taboo at the point on the left that represents a sufficiently strong taboo so that a proportion of the population does not consider deviation. Suppose now that b rises sufficiently such that the curved line is above the straight line (case (a) in Figure 1). The stable taboo has reached the point $\phi_c = 1$ and all individuals consider deviation. Now suppose that b falls back to its original level – so we are back to case (c). According to the dynamic adjustment process specified in Section 1, in such circumstances the taboo remains at point $\phi_c = 1$ and does not converge to its original value.

To illustrate, think about the case in which kidney transplants become easy and the private value of trade in kidneys becomes sufficiently high so that many people deviate from the taboo of trading in kidneys and consider this possibility. Suppose now that due to some technological breakthrough it becomes easy to transplant artificial kidneys. The private benefit from trading in kidneys will then decline although we do not expect that the taboo against such trade will return to its original level of strength.

4. Heterogeneity of Types

Does the effectiveness of taboos depend on social homogeneity? Some societies are characterized by a homogenous population while others have a more diverse population with a more heterogeneous preference distribution. Heterogeneity can be with respect to the distribution of private benefits or in the effectiveness of social costs and punishment. The question is whether it is easier to maintain a stronger

taboo in homogeneous or in heterogeneous societies. Clearly, the answer depends on the type of preferences individuals have in the society. In one homogeneous society individuals may not care about social costs while in another everyone is very much concerned with social costs. Thus, in order to examine the relationship between the strength of taboos and the homogeneity of society, we vary the degree of homogeneity without changing the average type of individual.

Let's consider our benchmark model but now assume that the type ϕ is uniformly distributed over [μ,1-μ], with $0 \leq \mu \leq 1/2$. Changing μ does not change the average type in the society but a higher μ implies a more homogeneous society. We can thus interpret μ as the degree of population heterogeneity. When μ=0 we are back with our benchmark case; when μ=1/2 we have a homogeneous society in which all individuals are of the same type.

In order to examine the effect of μ on the taboo's strength, let us first examine the two equilibrium conditions, (3) and (4). Condition (3) is derived from the individual's cost-benefit considerations, which depend on his type but not on the distribution of types. Therefore:

$$(3') \qquad \phi_c = \min\{\max\{\frac{qb}{(\lambda + q\delta)T}, \mu\}, 1-\mu\} \quad .$$

Condition (4) defines the strength of the taboo as a function of the percentage of people that maintain it; it thus depends on N_c and N_d, which are given now by:

$$N_c = \frac{\phi_c - \mu}{1-2\mu}; N_d = q\frac{\phi_c - \mu}{1-2\mu}.$$

Inserting these terms into (4) we derive the new condition as follows:

$$(4') \qquad \begin{aligned} T(\phi_c) &= \alpha_c(1 - \frac{\phi_c - \mu}{1-2\mu}) + \alpha_d(1 - q \cdot \frac{\phi_c - \mu}{1-2\mu}) = \\ &= [\alpha_c + \alpha_d + \frac{\mu(\alpha_c + q\alpha_d)}{1-2\mu}] - \frac{1}{1-2\mu}(\alpha_c + q\alpha_d)\phi_c \end{aligned}$$

We can now illustrate these two conditions (Figure 3). Note that as μ increases, condition (4') has a steeper slope and a higher intercept.

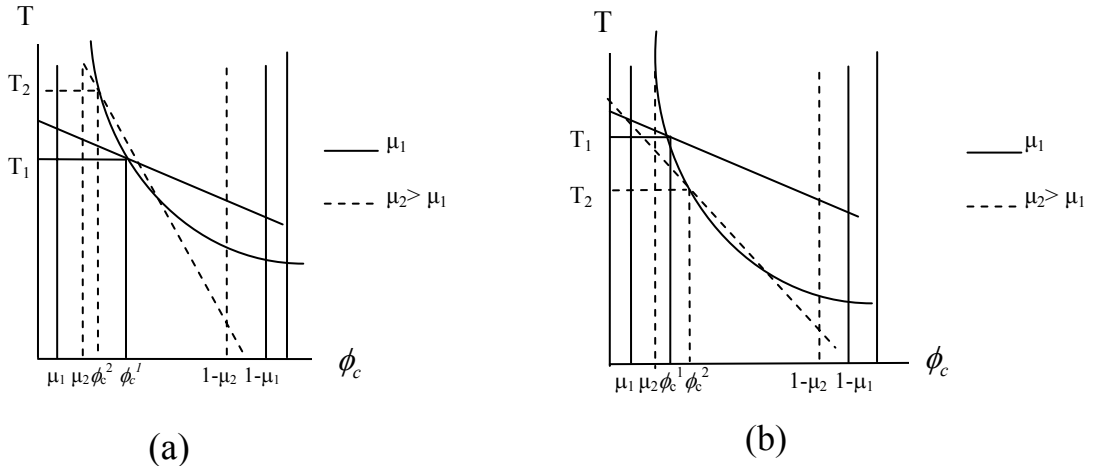

Figure 3: The effect of population homogeneity on a stable taboo.

Figure 3 illustrates two possible scenarios. In both cases we compare the stable taboo in two societies, letting $\mu_2 > \mu_1$, thus assuming that the second society, with μ_2, is more homogeneous than the first society, μ_1. In Figure 3a we describe a homogeneous society, where the stable taboo is stronger as there are fewer people who consider deviation from the taboo. But this is not necessarily the only outcome of a homogeneous society: In Figure 3b we demonstrate the opposite possibility. In this case, the second society is even more homogeneous but there are also more people who consider deviating from the taboo that, as a result, is weaker than in the less homogeneous society.

Proposition 2: (i) For low levels of b, increasing the level of homogeneity will cause the stable taboo to weaken, with more people considering deviation, while for high levels of b, more homogeneity causes the stable taboo to strengthen, with a smaller percentage of the population considering deviation from the taboo.
(ii) The weakest taboo that can be supported in a society is stronger in a more-homogeneous society (given a higher μ).

Proof: See Appendix 2. ■

We prove Proposition 2 by solving the two conditions (3') and (4') and analyzing the effect of μ on the stable taboo. In the Appendix we show that there are two critical levels of private benefits, $\bar{b}$ and $\tilde{b}$, where $\bar{b} < \tilde{b}$ such that in case A

(Figure 1), where the only stable situation is a weak (or no) taboo in which everyone considers deviating from the taboo, then:

(i) Whenever $b \geq \tilde{b}$, increasing μ and having a more homogeneous society may induce taboo formation.

(ii) Whenever $b < \tilde{b}$, decreasing μ and having a more heterogeneous society may induce taboo formation.

If, however, we have case B or C (Figure 1), such that there is an effective stable taboo (i.e., $\mu \leq \phi_c \leq 1-\mu$), then:

(i) Whenever $b \geq \bar{b}$, increasing μ and having a more homogeneous society results in a stronger taboo with less deviation (i.e., for $b \geq \bar{b}$ we have $\frac{\partial \phi_{c_1}}{\partial \mu} < 0$).

(ii) Whenever $b < \bar{b}$, increasing μ and having a more homogenous society results in a weaker taboo with more deviation (i.e., for $b < \bar{b}$ we have $\frac{\partial \phi_{c_1}}{\partial \mu} \geq 0$).

Proposition 2 implies that in more-homogeneous societies, a taboo that is associated with low private benefits provides less temptation to deviate and is weaker than the same taboo in a more heterogeneous society. Alternatively, a taboo that is associated with high private benefits is stronger. This property may have an effect on how different societies adjust to changes in potential private benefits. Returning to our example of organ transplants, if a new invention implies higher benefits for deviating from this taboo, we should expect more resistance to such changes in the more homogeneous society in which the taboo would still be stronger than would the same taboo in a heterogeneous society.

5. Multiple Identities

In some societies, different identities may be available to its members. Although some individuals may be born with a specific identity inherited to from their tribe, family, gender, etc., others may be able to decide which identity to adopt. This choice defines a dynamic process by which some identities can become more widespread, as more individuals choose to adopt them, while others disappear.

Consider, for example, religious and secular groups that coexist in a society. In such a case, individuals may be allowed to choose the group to which they wish to belong. In addition, each identity comes with a list of taboos and social norms; an individual needs to choose which identity he prefers to adopt and which set of norms and taboos to accept.

The Interaction between identities may not be constrained to competition over the "hearts and minds" of individuals. It is possible that the behavior of the individuals who adopt one identity shapes the taboos and the norms associated with another identity. For example, in our definition of a taboo, the taboo's strength is endogenously determined by the people who obey the taboo's rules without considering any deviation from them. When a taboo is common to different identities, deviation by members of one identity may affect the strength of the taboo in the other identity. In this case, interactions between the identities may occur via the joint reference group that holds the taboo in common and thus jointly determines the strength of the taboos in both identities. This creates a setup in which we can endogenize the concept of identity. Some identities will not survive the competition with other identities and will disappear. This process may be the result of individuals' choice of identities or a result of the interaction between identities that change and shape the taboos and the social norms of the different identities.

In this section we consider only one simple scenario that would clarify and demonstrate the endogenous shaping of identities. In the model, there are two competing identities, each with one taboo. The taboos may be the same or different. The strength of a taboo is endogeneously determined in each identity by the percentage of individuals that deviate or consider deviation from that taboo. Hence, there is no influence across identities regarding the strength of the taboos. The only interaction is the choice made individuals regarding the identity they choose to belong to.

Consider a society in which there are two possible identities A and B.[15] There is only one taboo and the two identities differ with respect to the strength of this taboo. The assumption of one taboo is not restrictive in our case as the taboo's strength is endogenously determined in each identity free of any influence from the behavior of individuals of the other identity. Each individual needs to choose the

[15] As we prove later, in our framework it is impossible to have a stable taboo system with more than two identities.

identity to belong to and then whether he/she would consider deviating from the taboo and suffering the associated penalty.

We further assume that the taboo's externality function is identical in the two identities; hence, the public benefits from the taboos are only a function of the strength of the taboo.[16] Thus, the two identities and taboos are identical ex-ante. One of the questions to be explored is whether population dynamics may result in ex-post asymmetric identities.

A community with two identities is defined as $\{(T_A,T_B),(\Omega_A,\Omega_A^c),(\Omega_B,\Omega_B^c)\}$ such that (T_A,T_B) are the strength of the taboos in the two identities. (Ω_A,Ω_B) is the assignment of individuals of different types to identities A and B, respectively; Ω_A^c and Ω_B^c are the groups of individuals that consider deviating from each of the taboos.[17]

The conditions that need to be satisfied are: (i) T_A and T_B – the strength of the taboo in each group – should be consistent with the percentage of people that consider deviation and those who actually deviate in each group; (ii) given T_A and T_B, each individual of type ϕ optimally choose his/her identity and whether to consider deviating from the taboo.

Since we assume that the two taboos are associated with the same externality function, i.e., $E(T_A)=E(T_B)$, one possible equilibrium occurs when individuals are equally distributed between the two identities.[18] The number of people who consider deviation and actually deviate among the two groups is therefore identical and the strength of the taboos is also identical. Individuals would therefore be indifferent to which identity to belong, and the strength of the identity would be endogenously determined as in the previous sections. The question remaining is if, despite the assumed symmetry, it is possible to have two different types of identities with one taboo stronger than the other, with different types of individuals and with different behaviors for each group.

[16] Clearly in a more general setup we may assume different externality functions but since the externality function is totally exogenous to our model we assume that it is identical in both societies, in order to have a setup in which the choice of individuals will be determined solely on their type and on the behavior of other people in the two groups.
[17] Note that under our simple framework q percent of those who consider deviation actually deviate.
[18] This is only one possibility. We do not need that individuals would be equally divided between the two identities. We need that the distribution of types in each group would be the same.

The utility of an individual of type ϕ who adopts an identity k in which the strength of the taboo is T_k is:

$$(8)\quad U_k(T_k,\phi) \equiv Max\{E(T_k); E(T_k) + qb - (\lambda + q\delta)T_k\phi\} \quad k=A,B.$$

Individuals of type ϕ will choose identity A if and only if $U_A(T_A, \phi) \geq U_B(T_B, \phi)$. Assume now, w.l.o.g., that $T_B > T_A$. Since the two taboos have the same externality function, $T_B > T_A$ implies that $E(T_B) > E(T_A)$. Note that when type $\phi = 0$, individuals are unaffected by social punishment; they will therefore always find it optimal to choose identity B and consider deviating from its taboo.

The possibility of choosing one's identity implies that all the individuals who choose identity A would consider deviating from its taboo. To see this consider an individual who chose an identity A but without considering deviating from its taboo. This individual does not suffer any social punishment but does enjoy the externality associated with the taboo. Since $E(T_B) > E(T_A)$, this individual would be better off belonging to identity B, which supports a stronger taboo. Thus, in a stable society with two stable identities, all the individuals belonging to the identity with the weaker taboo would consider deviating from its taboo.[19]

In Figure 4 we show the utilities arising from choosing identities A and B as a function of type ϕ. The horizontal part of the graph presents the choice of obeying the taboo and thus enjoying the benefits ET_j, whereas the declining part is the utility enjoyed by individuals who adopt the identity but also consider deviating from the taboo. Note also that because $T_B > T_A$, U_B is steeper than U_A since $ET_B + qb - (\lambda + q\delta)T_B\phi$ declines faster with ϕ than does $ET_A + qb - (\lambda + q\delta)T_A\phi$.

[19]This result is sensitive to our specific setup. If we change the externality function or the other primitives, the model can generate a multi-identity society in which there is a mixture of individuals who follow or deviate from the taboo in both identities.

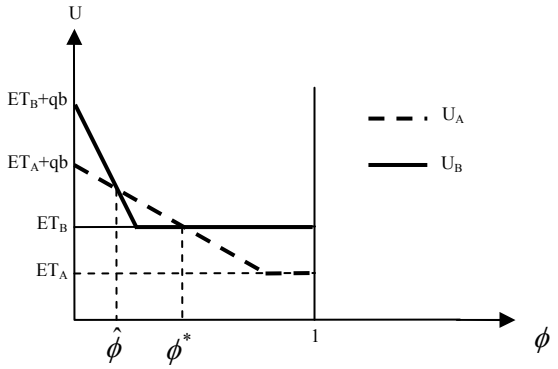

Figure 4: Utilities in a two-identity society.

Figure 4 illustrates the type equilibria that can emerge in this society. When T_A is sufficiently low such that U_A lies fully below U_B, then all individuals prefer identity B, which will become the only identity in the society. Otherwise a two-identity society will emerge, to be characterized by two parameters, $\hat{\phi}$ and ϕ^*. Identity B will consists of types $\phi < \hat{\phi}$, who choose identity B but also consider deviating from its taboo, and types $\phi > \phi^*$, who choose identity B but without considering deviation from the taboo. Identity A would consist of types $\hat{\phi} \le \phi \le \phi^*$, who all consider deviation from its taboo.

The two identity society thus has the following interesting structure: There is one relatively homogeneous group that adopts the identity with the weaker taboo. These individuals are moderately concerned about their identity and so they all consider deviating from the taboo. The second group is heterogeneous with respect to their social concerns. Although they adopt the identity with the strong taboo, they consist of two groups. One group is comprised of individuals with low social concerns who consider deviating from the taboo while the other group consists of individuals with strong social concerns who indeed adopt the strong taboo but without considering a deviation from it.

Assume now that we have three identities such that $T_1^* < T_2^* < T_3^*$. We can repeat our previous analysis to claim that in group 1, all individuals consider deviation from the taboo. If some do not consider deviation, then they are better off belonging to groups 2 or 3 because the social punishment is not applicable to them; they may therefore enjoy higher public benefits (which increase with the strength of the taboo). We can repeat the same argument with respect to the individuals that belong to group

2: If they do not consider deviation from the taboo, they are better off belonging to group 3. Hence, group 1 and 2 consist only of individuals that consider deviation. In consequence, the taboo has the same strength in both groups, i.e., $T_1^* = T_2^*$, and the identities do not differ from one another. We can thus conclude the following:

Proposition 3: When there is only one taboo and identities differ only in the strength associated with this taboo then,

(i) A two-identity society is stable. A multi-identity society with more than two identities is not stable.

(ii) In a society with two identities, some of the individuals who adopt the identity with the weaker taboo will always consider deviating from it. The identity with the stronger taboo is more heterogeneous and is adopted by two groups of individuals. Individuals of type $\phi < \hat{\phi}$ consider deviating from the taboo and individuals of type $\phi > \phi^*$ do not consider such a deviation.

Proposition 3 considers the individuals' choice of identity when a taboo's strength is exogenously given. The next step is to relate the strength of the taboos to the individuals' choices as well as to identify the conditions under which we may have one society with two competing identities and conditions that would lead to a uniform society with only one identity.

The first condition is that of $\phi^* < 1$. When $\phi^* \geq 1$, all individuals consider deviation from both identities, which implies that the taboos are of the same strength. For a two-identity society, we require that $\phi^* < 1$. The critical level ϕ^* can be derived from the indifference between belonging to identity A with considering deviation and belonging to identity B without considering deviation. Thus, the first condition is:

$$(9) \quad \phi^* = \frac{ET_A - ET_B + qb}{(\lambda + q\delta)T_A} < 1 \ ,$$

which implies the following restrictions on parameters:

$$(10) \quad qb < ET_B - ET_A + (\lambda + q\delta)T_A \, .$$

The second condition is that the line $U_A(T_A, \phi)$ will not lie entirely below $U_B(T_B, \phi)$. In other words, there exists a type ϕ for which $U_A(T_A, \phi) > U_B(T_B, \phi)$. This condition will be satisfied when $U_A(T_A, \hat{\phi}) > ET_B$.

To identify type $\hat{\phi}$, note that by definition, this type is indifferent between identities A and B, such that no matter what his/her choice is, he/she would consider deviating from the taboo. The type $\hat{\phi}$ indifference condition is:

$$(11) \qquad ET_A + qb - (\lambda + q\delta)T_A\hat{\phi} = ET_B + qb - (\lambda + q\delta)T_B\hat{\phi},$$

which yields:

$$(12) \qquad \hat{\phi} = \frac{ET_B - ET_A}{(\lambda + q\delta)[T_B - T_A]} = \frac{E}{(\lambda + q\delta)}.$$

Inserting this value into $U_A(T_A, \hat{\phi})$ implies that $U_A(T_A, \hat{\phi}) = qb$; hence, the second requirement is that $qb > ET_B$. The necessary condition for a two-identity society is, then,

$$(13) \qquad ET_B < qb < ET_B - ET_A + (\lambda + q\delta)T_A .$$

A necessary condition for (13) is that $\lambda + q\delta > E$, which holds when public benefits are not too high and the expected punishment from deviation is not too low. But note that T_A and T_B are endogenously determined, so the question is: Is there a range of parameters for which condition (13) is satisfied, with T_A and T_B consistent with the individuals' behavior and $T_B > T_A$?

Proposition 4: A necessary condition for having a two-identity society is that $\lambda + q\delta > E$. If this condition is satisfied, then for every q there is a range of private rewards b such that $T_B > T_A$, condition (13) is satisfied and T_A as well as T_B are consistent with the individuals' behavior.

Proof: See Appendix 3.

Concluding Remarks

Talking about taboos is talking about the unthinkable. It was not easy to choose and describe the examples presented in this paper, not because there are insufficient examples but because discussing deviation from specific taboos made us uncomfortable: The discussion itself is a violation of a taboo. Using rational

terminology to discuss the possibility of thinking about eating human flash is not a simple task and may be repel some of the readers. In our society one does not need to justify or explain the taboo of not eating dog meat. We simply do not think about such a possibility. We changed our examples a number of times but even then one may ask why did we chose this example and not another? Is the taboo of not eating dog meat less controversial than other taboos? Does it say something about us, about our preferences, or our identities? Are other examples more unthinkable than the ones we chose (e.g., eating human flash)? Then again, the examples may be so remote and so controversial that it is "safe" to use them. These worries simply reflect the fact that simply discussing these issues imposes some personal costs.

Our paper suggests a framework with which we can endogenize our attitudes towards different taboos. By letting the penalties of considering a deviation or actually deviating from a taboo become a function of what other people, who share the same identity, do and think, we endogenously determine the strength of the taboo. Letting people choose their identity then defines a process by which some identities disappear while others flourish. The passage between identities may affect the strength of taboos (and social norms) for each identity. As a result, we propose viewing identity not as prescribed set of values, taboos, norms, etc. but as a dynamic concept that evolves, changes and adapts in response to shifting environments and population dynamics.

References

Akerlof, G. A. (1976). "The Economics of Caste and the Rat-Race and Other Woeful Tales". *Quarterly Journal of Economics*, (November) 599-618.

Akerlof, G. A. (1980) "A Theory of Social Custom, of Which Unemployment May Be One Consequence". *Quarterly Journal of Economics*, June, 749-775.

Akerlof, G. A. and R. Kranton (2000) "Economics and Identity". *Quarterly Journal of Economics*, 115, 716-753.

Benabou, R. and J. Tirole (2006) "Identity, Dignity and Taboos: Beliefs as Assets".

Fessler, D.M.T. and C.C. Navarrete (2003) "Meat is Good to Taboo: Dietary Proscriptions as a Product of the Interaction of Psychological Mechanisms and Social Processes", *Journal of Cognition and Culture*, 3(1):1-40.

Fiske, A. and P. Tetlock (1997) "Taboo Trade-offs: Reaction to Transactions that Transgress the Sphere of Justice" *Political Psychology*, 18, 255-297.

Haidt J., P. Rozin, C. McCauley and S. Imada (1997) "Body, Psyche, and Culture: The Relationship between Disgust and Morality" *Psychology and Development Societies*, 9, 107-131.

Romer, D. (1984) "The Theory of Social Custom: A Modification and Some Extensions", *Quarterly Journal of Economics*, November, 717-727.

Roth, A.E. (2007) "Repugnance as a Constraint on Markets", *Journal of Economic Perspectives*, 21(3), 37-58.

Tetlock, P.E., O.V. Kristel, S.B. Elson, M.C. Green and J.S. Lerner (2000) "The Psychology of the Unthinkable: Taboo Trade-Offs, Forbidden Base Rates, and Heretical Counterfactuals", *Journal of Personality and Social Psychology*, 78(5), 853-870.

Zelizer, V.A. (1978) "Human Values and the Market: The Case of Life Insurance and Death in 19th Century America". *The American Journal of Sociology*, 84(3), 596-610.

Zelizer, V. A. (1981) "The Price and Value of Children: The Case of Children/Child Insurance" *The American Journal of Sociology*, 86(5), 1036-1056.

Appendix 1: Proof of Proposition 1

The conditions that give rise to the three cases:

Equations (3) and (4) yield a quadratic equation:

$$(\alpha_c + q\alpha_d)(\lambda + q\delta)\phi_c^2 - (\alpha_c + \alpha_d)(\lambda + q\delta)\phi_c + qb = 0$$

We define $\Delta \equiv (\alpha_c + \alpha_d)^2(\lambda + q\delta)^2 - 4(\alpha_c + q\alpha_d)(\lambda + q\delta)qb$

When $\Delta < 0$, we are in case (A).

When $\Delta = 0$, condition (3) is tangent to condition (4).

When $\Delta > 0$, cases (B) and (C) arise:

There are two solutions for the quadratic equation –

$$\phi_{c_{1,2}} = \frac{(\alpha_c + \alpha_d)(\lambda + q\delta) \pm \sqrt{\Delta}}{2(\alpha_c + q\alpha_d)(\lambda + q\delta)}$$

where $\phi_{c_2} > \phi_{c_1}$

When $\phi_{c_2} > 1$ we are in case (B).

When $\phi_{c_2} < 1$ we are in case (C).

Using an upper bound for Δ, it is possible to show that $\phi_{c_1} > 0,\ \phi_{c_2} < \frac{\alpha_c + \alpha_d}{\alpha_c + q\alpha_d}$.

Conditions on the parameters:

We define $\hat{b} \equiv \frac{(\alpha_c + \alpha_d)^2(\lambda + q\delta)}{4q(\alpha_c + q\alpha_d)}$

$b = \hat{b} \Rightarrow \Delta = 0$

Thus, for $b > \hat{b}$, we have case (A), and for $b < \hat{b}$ we are in case (B) or (C).

A distinction between cases (B) and (C) is achieved by looking at the value of ϕ_{c_2}, as stated above. A sufficient condition could be derived:

We are in case (B) when $\phi_{c_2} > 1$, which is ensured by:

$$\frac{\alpha_c + \alpha_d}{2(\alpha_c + q\alpha_d)} > 1 \ \Leftrightarrow\ (1 - 2q)\alpha_d > \alpha_c$$

Since $\alpha_d > \alpha_c$, we can see that for small values of q, we have one stable taboo (case B). ■

Appendix 2: Proof of Proposition 2 (Heterogeneity of Types):

Equations (3') and (4') yield the following quadratic equation:

$(\alpha_c + q\alpha_d)(\lambda + q\delta)\phi_c^2 - (\lambda + q\delta)[(\alpha_c + \alpha_d)(1-2\mu) + \mu(\alpha_c + q\alpha_d)]\phi_c + qb(1-2\mu) = 0$.

Define:

$\Delta' \equiv (\lambda + q\delta)^2[(\alpha_c + \alpha_d)(1-2\mu) + \mu(\alpha_c + q\alpha_d)]^2 - 4(\alpha_c + q\alpha_d)(\lambda + q\delta)qb(1-2\mu)$.

$\tilde{\Delta} \equiv \Delta'/(\lambda + q\delta) = (\lambda + q\delta)[(\alpha_c + \alpha_d)(1-2\mu) + \mu(\alpha_c + q\alpha_d)]^2 - 4(\alpha_c + q\alpha_d)qb(1-2\mu)$
$M \equiv (\lambda + q\delta)[(\alpha_c + \alpha_d)(1-2\mu) + \mu(\alpha_c + q\alpha_d)]$.

Solving the above quadratic equation yields that:

$$\phi_{c1,2} = \frac{(\lambda + q\delta)[(\alpha_c + \alpha_d)(1-2\mu) + \mu(\alpha_c + q\alpha_d)] \pm \sqrt{\Delta'}}{2(\alpha_c + q\alpha_d)(\lambda + q\delta)} = \frac{M \pm \sqrt{\Delta'}}{2(\alpha_c + q\alpha_d)(\lambda + q\delta)} \quad .$$

To examine how changes in μ affect the taboo, we examine the following derivatives:

$$\frac{\partial M}{\partial \mu} = -(\lambda + q\delta)[\alpha_c + (2-q)\alpha_d] < 0 \ ,$$

and

$$\frac{\partial \tilde{\Delta}}{\partial \mu} = 2(\lambda + q\delta)[(\alpha_c + \alpha_d)^2(4\mu - 2) + (\alpha_c + \alpha_d)(\alpha_c + q\alpha_d)(1-4\mu) + \mu(\alpha_c + q\alpha_d)^2] + 8qb(\alpha_c + q\alpha_d)$$

Define:

$$\tilde{b} \equiv \frac{(\lambda + q\delta)[(\alpha_c + \alpha_d)^2(2-4\mu) + (\alpha_c + \alpha_d)(\alpha_c + q\alpha_d)(4\mu - 1) - \mu(\alpha_c + q\alpha_d)^2]}{4q(\alpha_c + q\alpha_d)} .$$

$$b > \tilde{b} \Rightarrow \frac{\partial \tilde{\Delta}}{\partial \mu} > 0$$

$$b = \tilde{b} \Rightarrow \frac{\partial \tilde{\Delta}}{\partial \mu} = 0$$

$$b < \tilde{b} \Rightarrow \frac{\partial \tilde{\Delta}}{\partial \mu} < 0$$

We now examine the effects of changes in the degree of homogeneity μ on the percentage of deviation from the taboo:

$$\frac{\partial \phi_{c1}}{\partial \mu} = \frac{1}{2(\alpha_c + q\alpha_d)(\lambda + q\delta)}\left\{\frac{\partial M}{\partial \mu} - \frac{(\lambda + q\delta)\frac{\partial \tilde{\Delta}}{\partial \mu}}{2\sqrt{\Delta'}}\right\} =$$

$$= \frac{-\sqrt{\Delta'}[\alpha_c + (2-q)\alpha_d] + (\lambda + q\delta)[(\alpha_c + \alpha_d)^2(2-4\mu) + (\alpha_c + \alpha_d)(\alpha_c + q\alpha_d)(4\mu - 1) - \mu(\alpha_c + q\alpha_d)^2] - 4qb(\alpha_c + q\alpha_d)}{2(\alpha_c + q\alpha_d)\sqrt{\Delta'}}$$

Define:

$$\bar{b} = \frac{(\lambda + q\delta)[(\alpha_c + \alpha_d)^2(2-4\mu) + (\alpha_c + \alpha_d)(\alpha_c + q\alpha_d)(4\mu - 1) - \mu(\alpha_c + q\alpha_d)^2] - \sqrt{\Delta'}[\alpha_c + (2-q)\alpha_d]}{4q(\alpha_c + q\alpha_d)}$$

We see that for:

$$b < \bar{b} \Rightarrow \frac{\partial \phi_{c1}}{\partial \mu} > 0$$

$$b = \bar{b} \Rightarrow \frac{\partial \phi_{c1}}{\partial \mu} = 0$$

$$b > \bar{b} \Rightarrow \frac{\partial \phi_{c1}}{\partial \mu} < 0$$

Thus, for a small level of b, raising the degree of homogeneity μ will cause the stable taboo to be weaker, with more people considering deviation. For high levels of b, more homogeneity causes the stable taboo to be stronger, with a smaller percent of the population considering deviation from the taboo.

Note also that $\bar{b} < \tilde{b}$. We thus have several possible domains with respect to the value of *b* and the initial state of the taboo:

Start with case A (in Figure 1) such that in effect there is no taboo:

(i). $b > \tilde{b}$: Increasing μ could induce taboo formation.

(ii). $\bar{b} < b < \tilde{b}$: Increasing μ will never create a taboo. On the other hand, lowering μ could induce taboo formation. If a taboo is formed in this domain, further lowering μ would weaken this taboo.

(iii). $b < \bar{b}$: Increasing μ will never create a taboo. Lowering μ could induce taboo formation. If a taboo is formed, lowering μ further would strengthen this taboo.

Now assume that we are in case B or C (Figure 1), in which there is an effective taboo:

(i). $b > \tilde{b}$: Increasing μ strengthens the taboo.

(ii). $\bar{b} < b < \tilde{b}$: Increasing μ strengthens the taboo but can also induce a switch to case A.

(iii). $b < \bar{b}$: Increasing μ weakens the taboo and could also induce a switch to case A. Lowering μ strengthens the taboo but does not induce a switch to case A.

Next, find the weakest taboo that still holds as a function of the degree of population homogeneity:

$$\tilde{\Delta} = 0 \Leftrightarrow \hat{b}_{\mu} = \frac{(\lambda + q\delta)[(\alpha_c + \alpha_d)(1-2\mu) + \mu(\alpha_c + q\alpha_d)]^2}{4q(\alpha_c + q\alpha_d)(1-2\mu)} \quad .$$

$$\tilde{\Delta} = 0 \Leftrightarrow \Delta' = 0 \;\Rightarrow\; \phi^*_{c_\mu} = \frac{\alpha_c + \alpha_d}{2(\alpha_c + q\alpha_d)} - \mu[\frac{\alpha_c + \alpha_d}{(\alpha_c + q\alpha_d)} - ½] \; .$$

We can now derive the weakest taboo as a function of population homogeneity:

$$T^*_{\mu} = [\alpha_c + \alpha_d + \frac{\mu(\alpha_c + q\alpha_d)}{1-2\mu}] - \frac{1}{1-2\mu}(\alpha_c + q\alpha_d)[\frac{\alpha_c + \alpha_d}{2(\alpha_c + q\alpha_d)} - \mu(\frac{\alpha_c + \alpha_d}{\alpha_c + q\alpha_d} - ½)] =$$

$$= \frac{\alpha_c + \alpha_d}{2} + \frac{\mu(\alpha_c + q\alpha_d)}{2(1-2\mu)}$$

thus: $\mu > 0 \Rightarrow \; \phi^*_{c_\mu} < \phi^*_c$ and $T^*_{\mu} > T^*$.

For higher levels of homogeneity, we need a stronger threshold taboo.

We now derive conditions on μ, q that ensure $\phi^*_{c_\mu} \in [\mu,\; 1-\mu]$:

$$\mu \le \phi^*_{c_\mu} \le 1-\mu \;\Leftrightarrow\; \mu \le \frac{\alpha_c + \alpha_d}{2(\alpha_c + q\alpha_d)} - \mu[\frac{\alpha_c + \alpha_d}{(\alpha_c + q\alpha_d)} - ½] \le 1-\mu$$

$$\frac{(1-2q)\alpha_d - \alpha_c}{\alpha_d(2-3q) - \alpha_c} \le \mu \le \frac{\alpha_c + \alpha_d}{2(\alpha_c + \alpha_d) + (\alpha_c + q\alpha_d)} .$$

Similarly,

$$\frac{(1-2\mu)\alpha_d - (1-\mu)\alpha_c}{\alpha_d(2-3\mu)} \le q \le \frac{(1-3\mu)\alpha_c + (1-2\mu)\alpha_d}{\mu\alpha_d} . \quad \blacksquare$$

Appendix 3: Proof of Proposition 4.

Eq. (13) states the necessary condition for a two-identity society. This condition requires that

$$\text{(A1)} \qquad ET_B < qb < ET_B - ET_A + (\lambda + q\delta)T_A$$

A necessary condition for (A1) to be satisfied is that $\lambda + q\delta > E$.

Yet, a taboo's strength in our model is endogenously determined by the percentage of people who consider deviation from the respective taboo.

First let us find T_A. Since all individuals in group A consider deviating from the taboo, $N^A_c = 1$, which implies that $N^A_d = qN^A_c = q$ and therefore:

$$T_A = \alpha_c(1 - N^A_c) + \alpha_d(1 - N^A_d) = \alpha_d(1-q) .$$

The percentage of individuals who consider deviating from the taboo in group B is:

$$\text{(A2)}\ N_c^B = \frac{\hat{\phi}}{1-\phi^* + \hat{\phi}} = \frac{ET_A}{(\lambda + q\delta)T_A - qb + ET_B}.$$

The strength of the taboo is defined by

$$\text{(A3)}\ T_B = (\alpha_c + \alpha_d) - (\alpha_c + q\alpha_d)N_c^B.$$

Substituting for N_c^B, we derive a quadratic equation, which we can solve for T_B.

We now need to establish that $T_B > T_A$. Using the above terms, this requires that $\alpha_c + \alpha_d - (\alpha_c + q\alpha_d)N_c^B > \alpha_d - q\alpha_d$; collecting terms, this condition is equivalent to requiring that $(\alpha_c + q\alpha_d)(1 - N_c^B) > 0$. This condition is satisfied only when $N_c^B < 1$, which is guaranteed by condition (A1). But condition (A1) does not hold for all combinations of q and b. Solving for T_B and inserting in (A1) yields the following condition:

$$\frac{(\alpha_c + \alpha_d)E + qb - \alpha_d(1-q)(\lambda + q\delta) + \sqrt{\Delta}}{2} < qb < \alpha_d(1-q)(\lambda + q\delta - E) + \frac{(\alpha_c + \alpha_d)E + qb - \alpha_d(1-q)(\lambda + q\delta) + \sqrt{\Delta}}{2}$$

Note that whenever $\lambda + q\delta > E$, the right hand side of this inequality is greater than the left-hand side. Thus, for every q that satisfies $\lambda + q\delta > E$, we can find a range of values of b for which condition (A1) is satisfied and we obtain a non-redundant two identity society. ■