RATIONAL CHOICE AND FREUDIAN ACCOUNTS OF COOPERATION

Debra Friedman and Barry R. Weingast*

April 1993

Abstract

In this paper we discuss one of the richest psychological conceptions of individuals – that offered by Sigmund Freud – in relation to rational choice theory. The purpose is twofold. First, we use coordination games to translate some of Freud's ideas into a choice theoretic framework; including the three central components of Freudian psychoanalysis, the id, ego, and superego. Second, we use this framework to suggest a deeper and richer psychological characterization of the individual within the choice theoretic framework. Freud's more complex view of individuals provides a potentially deeper understanding of social interaction within the rational choice approach.

We focus on Freud’s Civilization and Its Discontents (1930) to discuss both the rational choice and the Freudian conception of the relationship between the individual and society, to note points of commonality and divergence, and to understand the consequences of these differences for models of social behavior. Freud makes two theoretical contributions in this work. The first parallels recent scholarship in rational choice in discussing the relationship between individual behavior and social outcomes. Both approaches emphasize, for example, that individuals bear a cost for participating in schemes of social cooperation and interaction.

In Civilization and Its Discontents, Freud goes beyond this idea, attempting to show the implications of his interpretation of society for the individual, an analysis absent from rational choice analysis. Freud argues that individuals bear a psychological cost in fulfilling the demands and obligations of civilization. Participation in society requires that individuals suppress their desires in order to realize the gains from cooperation. The implications from the second part of Freud’s analysis allow us to study the limits of the standard rational choice conception of the individual.

NB. Were we to rewrite this paper today, we would embed the discussion in the explosion of literature in behavioral economics drawing on aspects of psychology to enrich models of human behavior.

*Dean of the Arizona State University College of Public Programs; and Senior Fellow, Hoover Institution, and Ward C. Krebs Family Professor, Department of Political Science, Stanford University. This paper was conceived while Friedman was a National Fellow at the Hoover Institution, Stanford University;
I. Introduction

Rational choice theory studies social phenomena that result from the interaction among individuals. Focusing on social outcomes, rational choice theory rests on a simplistic conception of individuals. For many choice situations, qualitative differences among choices are of no consequence, and the characterization of the choice process is adequately captured by standard assumptions about preferences. But not all choice situations readily fall into this simple framework, and rational choice’s simplistic conception of the individual comes into question.

In this paper we discuss one of the richest psychological conceptions of individuals – that offered by Sigmund Freud – in relation to rational choice theory. The purpose is twofold. First, in the context of coordination problems, we suggest how to translate some of Freud’s ideas into a choice theoretic framework, including the three central components of Freudian psychoanalysis, the id, ego, and superego. Second, we use this framework to suggest a deeper and richer psychological characterization of the individual within the rational choice framework. Freud’s more complex view of individuals provides a potentially deeper understanding of social interaction within the rational choice approach.

We focus on Freud’s *Civilization and Its Discontents* (1930) to discuss both the rational choice and the Freudian conception of the relationship between the individual and society, to note points of commonality and divergence, and to understand the consequences of these differences for models of social behavior. Freud makes two theoretical contributions in this work. The first parallels recent scholarship in rational choice in discussing the relationship between individual behavior and social outcomes. Both approaches emphasize, for example, that individuals bear a cost for participating in schemes of social cooperation and interaction.
In *Civilization and Its Discontents*, Freud goes beyond this idea, attempting to show the implications of his interpretation of society for the individual, an analysis absent from rational choice analysis. The implications from the second part of Freud’s analysis allow us to study the limits of the standard rational choice conception of the individual.

This essay proceeds as follows. Section 2 introduces the Freudian conception of the individual in the context of coordination problems. Section 3 contrasts a rational choice and a Freudian account of sharing. Section 4 discusses the costs of cooperation from a Freudian perspective. Section 5 models Freud's account of civilization as a coordination game. Section 6 turns to the individual's internal psychic struggle generated by civilization. Section 7 discusses aspects of the individual’s internal psychic struggle. Our conclusions follow, including a discussion of the Freudian concept of renunciation.

### 2. The Coordination Problem

In both the Freudian and rational choice approaches, modern society presents individuals with an elaborate coordination problem. In both approaches, individuals find cooperation costly. In rational choice theory, the individual will do so when the net benefits from cooperating are larger than those from defecting. In Freudian theory, the individual also may cooperate when the benefits are sufficient, but the individual's decision to do so entails more than the juxtaposition of costs and benefits.

The added complexity arises from Freud's conception of the tripartite personality structure comprising the id, ego, and superego. As infants, individuals begin in an ego-undifferentiated state wherein there is no separation of the external world from the ego (Freud, 1930:15). Yet through social interaction, infants learn that at least some of their desires and
impulses will be met with external sanction. These desires and impulses, many of which are instinctual, comprise the id, the part of the individual that focuses on wish-fulfillment and that knows no boundaries. The id is unconcerned with constraints, contradictions among desires, or the passage of time.

Because the desires of the id bring external reactions, the individual seeks to avoid the sanction (because of the fear of the potential loss of love) and so develops a powerful internal authority that regulates desire. Freud called this internal authority the superego. Were the individual to fail to develop the ability to regulate desire from within, the attempt to satisfy those boundless desires would meet with external sanction: "An unrestricted satisfaction of every need presents itself as the most enticing method of conducting one's life, but it means putting enjoyment before caution, and soon brings its own punishment" (Freud, 1930:27).

It is left to the third part, the ego, to regulate the internal struggle between the id and the superego, as well as to consider the environment in which decisionmaking takes place. The ego is the mediator between the impulsive id and the restrictive superego. It is also the locus of rational action: to the extent that individuals understand time, opportunity costs, consistency, consequences, and so forth, it is the ego that embodies these understandings and logical constraints. The ego thus serves three harsh masters: (1) the omnipresent desire and appetitious aspects of the individual – the id; (2) the demanding conscience of the individual – the superego; and (3) the external constraints imposed by reality.

For Freud, civilization requires that individuals suppress their libidinous urges: civilization becomes possible only with this suppression; it is also the way in which the cooperation dilemma is ultimately resolved. "A good part of the struggles of mankind," Freud observes, "center round the single task of finding an expedient accommodation... between this
claim of the individual and the cultural claims of the group” (Freud, 1930:50). A large component of this accommodation arises through the internalization of control found in the superego; it is as if society sets up "an agency within [the individual] to watch over it, like a garrison in a conquered city” (Freud, 1930:84).

The decision to cooperate results from the interplay of id, ego, and superego.

To illustrate the differences between the two approaches, we offer an analysis of sharing.

3. Two Analyses of Sharing

Sharing is a quintessential element of cooperation among individuals. It is to be distinguished from exchange. Consider two individuals. In a standard exchange, the two individuals trade at two objects, and each individual gains more by the exchange than she gives up. Sharing, by contrast, involves one, indivisible object together with two strong and opposing desires, made stronger by the other's actual or potential possession of it. How do individuals evolve to comply with the norm of sharing? We offer two analyses, first from rational choice theory and then from Freudian psychoanalysis.

A Rational Choice Account

Calvert's (1989) model of the norm of reciprocity offers a way to model sharing from the rational choice perspective. In his model, one of two individuals holds initial possession of a desired object in each period. In the first period, one individual initially possesses the object, and she may retain complete possession or share the object with the other (see Figure 1). In the second period, the situation is reversed: the other individual possesses the object and may choose to retain or to share it. This sequence is repeated indefinitely.
The payoffs are straightforward. Complete possession is worth three; sharing, two; and no possession, zero. In this scenario, sharing can emerge as an equilibrium provided that neither individual discounts the future too heavily. The logic is familiar. Each individual is willing to share as long as the other shares. Each individual might prefer complete possession, but also prefers that the other shares rather than retains complete possession. Thus, if the first individual anticipates that, should she share, the other will follow suit in the second period, she will choose sharing: A payoff of two in both periods exceeds three in the first period and zero in the second.

Three observations are worth noting about this analysis. First, and most important, individuals anticipate the responses of others to their actions and predicate their choices on that knowledge. No accumulation of knowledge is necessary, no learning required: although rational choice theory encompasses several models of learning, nearly all game theoretic models assume that the players know the equilibrium from the start of the game. Second, no prospect for reciprocity exists, individuals will not share. And third, this analysis readily extends to a community of individuals who behave reciprocally with one another even if, for any one individual, there is no immediate occasion for reciprocity from the individual with whom she shares today. Reciprocity can be supported here, provided that an entire community responds to the behavior of an individual by refusing to share with her in the future if she refuses to share now.

---

1 This account can easily be extended to include a community of individuals in which one might share with another even though he does not expect to be in a reciprocal situation. This can be supported when an entire community responds to the behavior of an individual; that is, by punishing one who did not share by refusing to share with him in the future.
A Freudian Account

Unlike in the rational choice account, we cannot begin the analysis in the immediate situation in which the individual is called upon to share. This is because the decision to share comes from learning that has taken place in the past, most likely when the individual was a young child.

By their nature, young children live in the present had have at best dim perceptions and understandings of the future. Sharing among young children presents difficulties for the rational choice approach, for it fundamentally rests on the individual's estimation that giving up something today yields something the future. Without this sense of time and interconnectedness of action, sharing will not emerge spontaneously among young children. No endogenous peaceful coordination emerges: the child in possession of the valued object will not give it up and will fight (literally) to keep it.

Learning to share requires outside intervention. To induce sharing, parents and teachers typically begin by appealing to a fairness rule: you may play with X for five minutes, and then your playmate gets it for five minutes. You may have it back for five minutes after that, and so on. As an appeal, this never works; though direct enforcement combined with careful monitoring can work. The next appeal is to the child's future self interest: if you don't share your toy, then your friend won't share his. This appeal also falls on deaf ears, perhaps because of future discounting, or perhaps because, at that moment, there is no toy as dear as the one in contention.
Parents enforce the norm of sharing in the following way: if a child does not respond to a call for sharing, the desired object is taken away with the dictum, "If you can't share it, no one will have it."\(^2\)

We model this setting in the following way. In the young child's initial estimation, the decision problem is simple. The child must decide whether to keep or share the object when called on by the parent to do so (see Figure 2). If the child keeps it, the payoff is three; if the child shares, the payoff is two. Hence the child wishes keeps the object. But the parent wishes to teach the child to share and so intervenes by removing the object if the child fails to share. Thus, figure 2 does not accurately represent the choice situation: failing to share yields a payoff of zero. With experience, the child learns to take account of the behavior of the parent and therefore comes to understand the more complex choice situation is shown in Figure 3.

Parental reactions affect the child’s choice between keeping and sharing, and in time the child learns to associate decisions about sharing with parental response. Sharing means the object will remain; refusing to share means the object will be removed.

As the child learns the sharing rule, she will attempt to avoid the external sanction by internalizing the parental response. In Freudian terms, the child internalizes the parental response as part of the repertoire of the superego.\(^3\) In the superego are sets of norms and rules of behavior reflecting sets of rewards and punishments associated with specific classes of action. In subsequent situations similar to the ones in which a norm was learned, the superego will exert the control over choice that the parent, the external agent, once did. We represent this choice setting

---

\(^2\) We do not model this interaction as a game; rather we focus on the choice problem facing the child, taking the parent's behavior as exogenous.

\(^3\) Rational learning models may be applied to this process, but for simplicity we suppress this.
in figure 4. Once an individual has internalized a sharing norm as part of her superego, even in
the absence of a parent, the individual will choose to share in a situation where the norm calls for
sharing. This behavior follows because the superego colors the ego's perceptions of the world
prior to ego's rational calculations about choice. The analysis in figure 4 shows that an individual
without a superego will choose to keep over sharing. But the superego, internalizing the (often
long ago) parental response, alters the individual’s perceptions of the payoffs so that she will
share.

As with the rational choice account, the notion of the community is important. Parents do
not teach their children to share indiscriminately. Children come to learn the boundaries of the
community by the application of the norm enforcement.4

The model of rational choice reciprocity and the Freudian analysis allow us to compare
the two methods of choice. In the rational choice approach, an individual shares because of
anticipated reciprocity and a knowledge that reciprocal sharing yields the highest payoff over the
long-term. In contrast, in the Freudian approach an adult who has internalized the sharing norm
and is faced with the same situation of reciprocity as the rational actor above, will not make her
choice as the rational actor did. She will share, not because she is playing a contingent strategy of
reciprocal sharing, but because her superego will bring internal punishment if she fails to share in
a setting that calls for sharing. Of course, both analyses produce the same outcome: both the

4 This setting gives rise to a further coordination problem. Given a society of \( n \) individuals, every individual perceives themselves to be part of a community, some subset of the entire
society. To be a community, (1) members must perceive themselves as such; (2) they must
perceive others as non-members; and (3) non-members must perceive themselves as such. This
is a problem with multiple equilibria, and for a given set of \( n \) individuals, the bound on the
number of potential communities is \( 2^n \).
rational individual and the individual who has internalized the rule of sharing will share. But the mechanisms that produce the same end are different.

In the rational choice perspective on norms, upon which our approach is based, the anticipation of punishment for failure to observe the norm is essential to maintain it. In equilibrium, punishment should never be observed. Indeed, individuals follow the behavior to be an equilibrium. Were no punishment credible, the norm could not be sustained. In the Freudian account, in contrast, punishment in the next period for failure to share today need not be credible for sharing to occur.

We might expect, therefore, that individuals of the Freudian account will choose to share more often than the individuals of the standard rational choice account, withdrawing from sharing only if reciprocity fails repeatedly. In novel situations for which there has been no prior learning, the superego remains silent, and ego acts just as an individual in the standard rational choice account.

This analysis suggests an important comparative statics result: some norms are easier to establish than others. In the Freudian perspective, punishment need not be credible in some settings for a norm to be maintained. The more an individual has internalized rules governing a situation due to prior learning, the less credible (or large) need be the sanctions in order for the norm to be observed.

The view that some norms are internalized via the superego does not imply that rational calculus is irrelevant. Indeed, other norms can be supported via the standard rational choice method, that is, by threat of sanctions. We hypothesize that there exists two categories of norms,

---

5 An interesting hybrid might be worth exploring, i.e., where an individual of the Freudian account meets one from the rational account.
one principally enforced internally, the other principally enforced externally. Those enforced internally may produce an added emotional or affective component upon violation: not only is a defector punished, but his behavior provokes outrage. Furthermore, for a given norm, to the degree to which external constraints have a corollary within the internal personality structure, their effect is multiplied.  

4. The Cost of Cooperation

Another fundamental difference emerges between the rational choice and Freudian accounts. The ability of the superego to exercise control over the desire for exclusive possession of a valued object does not mean that the wish for exclusive possession has been eradicated. On the contrary. In general, individuals bear a psychological cost in fulfilling the demands and obligations of civilization. Participation in society requires that individuals suppress their desires in order to realize the gains from cooperation. Civilization requires above all instinctual renunciation in order that the power of the group is preeminent over the power of the individual:

Human life in common is only made possible when a majority comes together which is stronger than any separate individual and which remains united against all separate individuals. The power of this community is then set up as "right" in opposition to the power of the individual, which is condemned as "brute force." This replacement of the power of the individual by the power of a community constitutes the decisive step of civilization. The essence of it lies in the fact that the members of the community restrict themselves in their possibilities of satisfaction... (Freud, 1930:42).

In both approaches, individual renunciation (of the desire to defect rather than cooperate) is central to the analysis. The difference is that in the Freudian account, the act of renunciation required of the individual gives rise to a by-product that threatens the very basis of civilization,

---

6 Were we to follow this logic, we think that it might lead toward an explanation of why certain norms exist, as against others.
namely aggression. Aggression comes to be controlled both externally and internally. The internal dynamic is one of introjection: the individual turns the aggressive impulse in upon himself through the creation of the harsh superego which comes to oppress the ego. This, in turn, leads to unhappiness.

5. Modeling Freud's Account as a Coordination Game

Just as for rational actors, in order to live with others, Freudian individuals must give up the essential egoism of their constitution. In standard rational choice accounts, this sacrifice is borne of a calculated exchange: when the gains from cooperation exceed those of purely short-term self-interested action, and individual renounces defection without regret. In the Freudian scheme, however, the cost of renunciation includes a component beyond that captured by the notion of an opportunity foregone.

Before turning to a full consideration of the costs of renunciation, in this section we show that there exists a striking resemblance to the results described by Freud in *Civilization and Its Discontents* and those derived from standard rational choice analyses of social cooperation.

We employ the standard model of social cooperation, the familiar repeated prisoners' dilemma (Axelrod 1984, Calvert 1995, Greif 1992, Kreps 1989, Milgrom, North and Weingast 1990, and Taylor 1976). The main conclusion of this approach is that social cooperation can be maintained as an equilibrium in which individuals forego their short-run interests in order to gain over the long-run. Herein lies a direct correspondence between the repeated prisoners' dilemma and the Freudian conception of renunciation. Individuals must be willing forego tempting alternatives in order to sustain social cooperation.
Mixed Prisoners' Dilemma - Coordination games (MPDC) provide further insights. In a standard prisoners’ dilemma game, the pareto optimal choice that makes the community best off is unique. In a MPDC game, more than one pareto-optimal equilibrium exist, and individuals in society have countervailing preferences over which one is better. The countervailing preferences arise because the gains from cooperation can be distributed in many ways. This game allows distributional conflicts, and so provides a more interesting way of modeling the dilemmas of social cooperation than does the simple prisoners' dilemma.

Both games show that Pareto optimal outcomes can be assured. Yet neither game provides much insight into the problem of unhappiness. We now turn to this question.

6. The Individual's Residual Internal Struggle

To live in civilization, individuals must limit their pleasure seeking, here modeled as forgoing short-term temptations to defect. In this the Freudian and rational choice accounts concur. Yet Freud goes beyond this analysis to argue that an individual bears consequences from renunciation. Standard rational choice theory has no such counterpart. As we will show, the intrapsychic consequences of renunciation in turn have social implications.

In the Freudian scheme, individuals turn away from pleasure-seeking not only because of external sanction, but because of their internal constitution which early on in development places limits on the rewards of pleasure-seeking. It is tempting to conclude that the tripartite structure of personality itself achieves a stable equilibrium, yet Freud cautions against the notion that a

\[\text{\footnote{Thus we ignore an obvious, but unhelpful, interpretation of unhappiness in these settings as reflecting the multiple equilibria and the potential inability of individuals to attain the Pareto optimal equilibrium.}}\]
permanent balance among the id, ego, and superego can be achieved. For one, the ego cannot always win: "even in normal people, the id cannot be controlled beyond certain limits. If more is demanded of a man, a revolt will be produced in him or a neurosis, or he will be made unhappy" (Freud, 1930:108). For another, it is impossible to attain all that is desired:

Happiness, in the reduced sense in which we recognize it as possible, is a problem of the economics of the individual's libido. There is no golden rule which applies to everyone: every man must find out for himself in what particular fashion he can be saved. All different factors will operate to direct his choice. It is a question of how much real satisfaction he can expect to get from the external world, how far he is led to make himself independent of it, and finally, how much strength he feels he has for altering the world to suit his wishes. In this his psychical constitution will play a decisive part, irrespectively of the external circumstances (Freud, 1930:34).

Far from the satisfaction or utility which the individual in the standard rational choice account achieves through cooperation and the successful attainment of her ends, by any action at all, the Freudian actor has only succeeded in recasting the terms of the battle to attain happiness.

Additionally, the very development of the superego precludes the possibility that the individual might find happiness through pursuit of the pleasure principle: "A threatened external unhappiness – loss of love and punishment on the part of the external authority – has been exchanged for a permanent internal unhappiness..." (Freud, 1930:89). As a result, the individual must cooperate with others in order to attain some measure of happiness.

At the same time, however, the act of cooperation produces aggressive impulses towards others. Unfulfilled wishes lead to aggressiveness in dealings with others:

The existence of this inclination to aggression, which we can detect in ourselves and justly assume to be present in others, is the factor which disturbs our relations with our neighbor and which forces civilization into such a high expenditure of energy. In consequence of this primary mutual hostility of human beings, civilized society is perpetually threatened with disintegration. The interests of work in common would not hold it together; instinctual passions are stronger than reasonable interests (Freud, 1930:69).
In the Freudian account, aggressiveness serves as the psychological basis for the actor's search for a strategy of defection. The flaw in civilization as it is presently constituted is that it rewards those who pursue this search:

The commandment, "love thy neighbor as thyself," is the strongest defense against human aggressiveness and an excellent example of the unpsychological proceedings of the cultural super-ego. The commandment is impossible to fulfil; such an enormous inflation of love can only lower its value, not get rid of the difficulty. Civilization pays no attention to all this; it merely admonishes us that the harder it is to obey the precept the more meritorious it is to do so. But anyone who follows such a precept in present-day civilization only puts himself at a disadvantage vis-a-vis the person who disregards it. What a potent obstacle to civilization aggressiveness must be, if the defence against it can cause as much unhappiness as aggressiveness itself! (Freud, 1930:109)

Thus the unbridled pursuit of happiness cannot bring happiness, nor can submission to the demands of civilization. Instead what is left is the continual struggle of ego to chart a course between the two.

How does this intra-psychic struggle affect social outcomes? At the individual level, Freud emphasizes the dual – and conflicting – impulses reside simultaneously within the individual – toward cooperation and aggression, toward egoism and altruism. He further asserts that the satisfaction of one impulse through choice of action is the very source of the competing impulse. At least two corollaries follow at the social level. Cooperative equilibria and pervasive hostile aggression exist side by side. So too do patterns of egoistic action exist alongside patterns of altruistic action. Although it has proven quite difficult to accommodate both observations in standard rational choice accounts, these twin realities arise from the logic of Civilization and Its Discontents.
7. Conclusion: The Importance of Renunciation

To live in society requires renunciation. Both the rational choice and Freudian approaches concur on this point. Beyond this, an examination of Freud's *Civilization and Its Discontents* teaches that both the process by which renunciation occurs and the cost to the individual matter in understanding the social contract.

In the Freudian accounts, external sanctions are less necessary for cooperation among adults than in a standard rational choice account. This conclusion is particularly true in cases where individuals learn the relevant form of cooperation early in their life as a child. This conclusion follows because early learning of norms leads to the internalization of sanctioning, rendering external sanctions at least partially redundant in adult life.

Moreover, the development of the superego renders happiness unlikely, for the individual can no longer pursue the pleasure-principle without paying an internal price. Happiness can ultimately come only through concert with others, but the act of cooperation produces its own antithesis.

Most importantly, the internal control exercised by the superego has consequences that may undermine the very cooperative equilibria that it is designed to sustain. To cooperate means to suppress the rampant desires of the id. Freud suggests that this is the basis of aggression, and the source of the greatest threat to civilization.

We conclude our essay by discussing some of its limitations and by suggesting the direction of our future work. We have emphasized that both the rational choice and Freudian

---

8 Neurosis is another possible outcome. In rational choice terms, neurosis is like a path-dependent state at the individual level: it is, by definition, a departure from a utility-maximizing strategy.
accounts emphasize cooperation and renunciation as a basis for society. The central difference is that, in the Freudian account, rules of behavior governing a present choice situation are taken from the internalized repertoire of those learned from similar situations in the past.

But when are such rules optimal? Our analysis of sharing suggests that the development of the rule in the first place may well have been optimal. Yet the application of the rule in the second instance similar to the first may not be.

The central weakness of this view lies in the process of identification. What is the mechanism by which an individual identifies current circumstances with earlier ones so that internalized rules are applied? This question holds one of the keys to operationalizing the Freudian theory, something that we hope to attend to in the future. Only then will we be able to move beyond something akin to assuming a taste for certain norms.

As we suggested above, comparative statics results seem to be one of the most fruitful approaches for operationalizing the theory. The difference between the rational choice and Freudian accounts of cooperation lies in the mechanism supporting a given norm. The first emphasizes external mechanisms in the form of sanctions; the second, internal mechanisms in the form of the superego.⁹

A hybrid approach, combining the insights of both, hold the possibility for a variety of comparative statics results. For instance, a norm may be observed in the absence of strong, or even any external sanctions, provided that there are internal mechanisms supporting it. Put

---

⁹ Another interesting contrast can be drawn here. Standard rational choice conception of the individual focus on the external constraints governing choice, but outcomes are ascribed to internal states (i.e. utilities). Freud's conception of the individual focuses principally on internal states, but these are ascribed to external forces (such as parental intervention in sharing dilemmas of young children).
another way, some norms may be easier to establish than others precisely because they appeal to internal sanctioning mechanisms.

In sum, the developmental conception of the personality embodied in Freudian psychoanalysis can enhance the shadow individual of the standard rational choice model without abandoning its powerful analytic framework. Combined, the two models offer new hypotheses in the area of social cooperation, and perhaps beyond it, as well.

References


Figure 1: A Rational Choice Model of Sharing.
Figure 2: Child's Initial Perception.
Figure 3: Parental Response.
Figure 4: Internalized Parental Response via the Super-ego.