

AN ECONOMIC MODEL OF MORAL MOTIVATION¹

Kjell Arne Brekke^{*}

Snorre Kverndokk^{**}

Karine Nyborg^{***}

Abstract

In this paper, we present an economic model of moral motivation. Consumers prefer regarding themselves as socially responsible individuals. Voluntary contributions to public goods are motivated by this preference. The self-image as socially responsible is determined by a comparison of one's actual behavior against an endogenous morally ideal behavior. Public policy influences voluntary contributions through its effects on relative prices and budget or time constraints, but also indirectly through the policy's effect on the morally ideal contribution. This implies that economic incentives may have adverse effects on contributions. We present survey data on recycling behavior and voluntary community work, which is consistent with the model predictions.

Keywords: Voluntary contributions, economic incentives, warm glow, identity theory.

JEL classification: D11, H41.

^{*} Centre for Development and the Environment, University of Oslo, P.O. Box 1116 Blindern, 0317 Oslo, Norway. E-mail: k.a.brekke@sum.uio.no

^{**} Ragnar Frisch Centre for Economic Research, Gaustadalléen 21, 0349 Oslo, Norway. E-mail: snorre.kverndokk@frisch.uio.no.

^{***} Corresponding author: Ragnar Frisch Centre for Economic Research, Gaustadalléen 21, 0349 Oslo, Norway. E-mail: karine.nyborg@frisch.uio.no.

¹ The Norwegian Research Council under the SAMRAM program supported this work financially. We are indebted to comments on earlier drafts from Rich Howarth, Michael Hoel, various conference and seminar participants, and two anonymous referees. The research reported in this paper was conducted while Brekke and Nyborg were employed by Statistics Norway.

I Introduction

Voluntary contributions to public goods are a part of everyday life. We vote to preserve democracy, recycle our household waste, send money checks to the World Wildlife Fund, volunteer at the local school, and carry our dirty trash all the way back home when hiking. Nevertheless, from the point of view of neoclassical economic theory, this behavior seems hard to explain.² Our purpose in the present paper is to provide a framework for formal analysis of the relationship between moral motivation, economic incentives, public policy and actual consumer choice. The focus is on moral, not social, norms. The desire to gain social acceptance has been analyzed by several authors (see, for example, Holländer [1990], Lindbeck [1997], and Nyborg and Rege [2000]), but will be entirely disregarded here.

Assuming that people truly care about each other's welfare is not enough to explain the large contributions to public goods that can be observed. Andreoni [1988] demonstrated that with pure altruism, average contributions to public goods will be close to zero in large economies. He pointed out that the existence of Red Cross, the Salvation Army, and the American Public Broadcasting thus appeared to be inconsistent with standard consumer theory. Andreoni's solution to the apparent paradox was to introduce a private benefit of contributing, "the warm glow of giving" [Andreoni, 1990].

While this *impure altruism model* was a major step towards an economic understanding of voluntary contributions, it cannot explain the observation that price incentives sometimes appear to have unexpected effects on voluntary contributions. For example, Frey and Oberholzer-Gee [1997] found that respondents' willingness to accept a hazardous waste treatment plant in their local neighborhood *decreased* if monetary compensation was offered. Gneezy and Rustichini [2000a] report experiments where the effect of monetary compensation on performance was not monotonic; introducing

² For discussions of this, see for example Sen [1977], Sudgen [1984], and Bernheim [1986].

compensation reduced performance, while increasing compensation increased performance when compensation already was introduced. A similar phenomenon appears in the evidence reported below: In Norway, many clubs and organizations arrange an event called *dugnad* once or twice a year, in which members meet on a voluntary basis to do practical work for the organization (painting the club house, setting up a new playground, etc). Many respondents to an interview survey we conducted said that if a fee for not showing up at the *dugnad* were introduced, they would participate *less frequently*.

The impure altruism model cannot explain such behavior: A fee on non-participation corresponds to a subsidy on voluntary contributions, which *increases* contributions.³ However, it is consistent with Bruno Frey's [1997] crowding theory, which is based on cognitive evaluation theory from social psychology (Deci and Ryan [1985]): Frey argues that although external incentives such as fees increase economic incentives to contribute, they can reduce individuals' internal or intrinsic motivation. Nevertheless, Frey's theory does not explain the presence of intrinsic motivation to contribute in the first place.

Andreoni's impure altruist obtains more "warm glow of giving" simply by contributing more; no complex moral reasoning on issues such as individual obligations or social values is required. If she is informed that recycling is less environmentally beneficial than previously believed, for example, it is not clear that she would change her recycling behavior. In contrast, the approach of dual or multiple preferences (see Harsanyi [1955]; Sen [1977]; Margolis [1982]) acknowledges individuals' ability to make sophisticated moral reasoning. In the present paper, we integrate the dual preferences framework of Nyborg [2000] with traditional neoclassical consumer theory. For this purpose we use the concept of identity, inspired by the identity philosophy of Mead [1913] (see also Brekke and Howarth [2000]; Akerlof and Kranton [2000]).⁴

³ In the pure altruism model, the subsidy will have no effect on contributions [Andreoni, 1990].

March and Olsen [1995] contrast two different modern views of governance. The first is what they call an exchange perspective, corresponding nicely to the classical Homo Economicus model. The exchange perspective assumes that individual action depends on the answers to three questions: "What are the alternatives? What are the consequences that will follow from each alternative? What is the value, in terms of preferences of the decision maker, of the consequences?" (p. 7). The second view is an institutional perspective, "(...) built around ideas of identities and conceptions of appropriate behavior. It assumes that individual action depends on the answers to three different questions: What kind of a person am I? What kind of situation is this? What does a person such as I do in a situation such as this?" (p. 7).

Our model integrates these two perspectives. We assume that people want to think of themselves as socially responsible ("What kind of a person am I?"). We further assume that individuals' perceived social responsibility varies with external conditions and public policy ("What kind of situation is this?"). Based on the first two questions, the individual considers a third one, slightly different from March and Olsen's: "What should a person such as I *ideally* do in a situation such as this?" However, when it comes to *actual* choices, we assume that individuals make trade-offs between the wish to be socially responsible and the desire for consumption and leisure, based on costs and benefits, corresponding to the above three questions of the exchange perspective.

In our model, individuals first determine their morally ideal effort by asking themselves the following question: "What would the consequences for social welfare be if everybody acted like me?" In the second step, individuals maximize utility, trading the benefits of maintaining a self-image as a socially responsible person against the costs. Self-image is determined by a comparison of one's actual effort to the morally ideal effort. This line of thought implies that contributions are motivated by a private good similar to Andreoni's "warm glow"; i.e., the benefits of a good self-image. This may be taken to

⁴ The focus on identity in explaining voluntary contributions is also found in the expressive preferences approach of Brennan and Lomasky [1993] and in Brennan and Hamlin [2000]. In contrast to these authors, however, we endogenize the ideal behavior individuals strive towards.

indicate a rather cynical view on moral behavior as simply serving the private interest of having good conscience. However, in our model, this private good is essentially linked to true moral reasoning: Improving one's self-image as a socially responsible person is only possible by making one's own effort come closer to one's own morally ideal effort. Thus, a better self-image can only be obtained by doing what one truly believes to be morally right.

In section II, the formal model is presented. Then we analyze the effects of changed information or increased productivity of effort in section III, while the impacts of economic incentives are discussed in section IV. Section V reports some results from an interview survey with about 1200 respondents conducted in 1999. Section VI concludes.

II The model

Assume that the society consists of N identical individuals with utility functions given by

$$(1) \quad U_i = u(x_i, l_i, G, I_i)$$

where x_i is individual i 's consumption of private goods, l_i is leisure, G is a pure public good, and I_i is a measure of the individual's self-image as a socially (or morally) responsible person.⁵ The utility functions are assumed to be increasing and strictly quasi-concave. For simplicity, we assume that labor supply and income are exogenously fixed, since our focus will be on the choice of leisure versus spending time to contribute to the public good.⁶

The time constraint of each individual is

⁵ Allowing heterogeneous consumers is possible, but would complicate the reasoning considerably. This would require a better specification of the Kantian rule discussed below, more specific assumptions concerning subjective interpersonal utility comparisons, and a clarification of whose normative views are reflected in the social welfare function.

$$(2) \quad l_i + e_i = T$$

where e_i is the individual's effort, measured in units of time, to contribute to increased supply of the public good (for example through recycling efforts). T is each individual's exogenous total time constraint minus the individual's exogenous labor supply, which is assumed to be equal for all. Again to simplify, we disregard monetary contributions to the public good.

The total amount of the public good depends on public provision G_p , and private provision $\sum_i g_i$:

$$(3) \quad G = G_p + \sum_i g_i$$

where

$$(4) \quad g_i = \gamma(e_i, \theta)$$

is individual i 's production function for the public good. Here, θ is an efficiency parameter exogenous to consumers, reflecting institutional or technical issues. We assume that $\gamma(0, \theta) = 0$, as well as $\gamma_e > 0$, $\gamma_{ee} < 0$, $\gamma_\theta > 0$, and $\gamma_{e\theta} > 0$; i.e., increases in the efficiency parameter increases the output of public good resulting from marginal efforts (here, subscripts denote derivatives).⁶

An individual's self-image as a socially (morally) responsible individual is determined by the relationship between her actions and her morally ideal actions. With regard to a positive self-image, the best one can do must be to contribute the morally ideal effort (this is embedded in the word “ideal”). Thus, we will specify self-image as follows, where e_i^* denotes the morally ideal effort from individual i :

⁶ This assumption may also be justified by the applications on recycling of household waste and participation in *dugnads* given below.

$$(5) \quad I_i = f(e_i, e_i^*) = -a(e_i - e_i^*)^2, \quad a > 0.$$

With the chosen functional form, the individual's self-image has a global maximum equal to zero at $e_i = e_i^*$. We see that the function is concave in e_i : It is increasing in e_i for $e_i < e_i^*$, and falling in e_i for $e_i > e_i^*$. Further, self-image is decreasing in e_i^* for $e_i < e_i^*$, and increasing in e_i^* for $e_i > e_i^*$.⁷

Before deciding her actual behavior, the individual considers what is her morally ideal effort. Assume that all individuals share a utilitarian moral philosophy, in the sense that they think the interests of every individual should count equally in evaluations of social welfare:⁸

$$(6) \quad W = u_I + \dots + u_N$$

To find the morally ideal effort e_i^* , the individual asks herself: "Which action would maximize social welfare, given that everyone acted like me?": a simple version of Immanuel Kant's Categorical Imperative.^{9,10} The answer to this question is found by maximizing W with respect to e_i , subject to (1)-(5) and $e_i = e_j$ for all $j \neq i, j = 1, \dots, N$.¹¹ Thus, in our model, there is no conflict between the utilitarian moral philosophy and the Kantian Categorical Imperative: Utilitarianism (or any other social welfare

⁷ In Brekke et al. (2000), we used a more general function for self-image: $I_i = f(e_i, e_i^*)$, where $f_{e_i} > 0$ if $e_i < e_i^*$; $f_{e_i} = 0$ if $e_i = e_i^*$, while $f_{e_i} < 0$ if $e_i > e_i^*$. Further, $f_{e_i^*} < 0$ for $e_i < e_i^*$, $f_{e_i^*} > 0$ for $e_i > e_i^*$ and $f_{e_i^*} = 0$ for $e_i = e_i^*$. f was assumed to be twice continuously differentiable and concave in e_i , so $f_{e_i e_i} < 0$. For any given e_i^* , I_i was assumed to have a global maximum at $f(e_i^*, e_i^*) = K$, where K is a constant. However, since this general function gives little further insight into the problems analyzed as compared to the more specific self-image function in (5), we will use the latter to simplify calculations.

⁸ This particular social welfare function is chosen only for simplicity. Replacing it with some other welfare function may change certain properties of the model predictions, but would not alter the main logic of our argument.

⁹ In a model with heterogeneous consumers, the adequate question would be: "Which general rule of action would maximize social welfare, as I perceive it, given that everyone acted according to the same general rule as I?" The morally ideal action would then be a function of one's own individual characteristics, for example income and/or preferences.

¹⁰ In the survey reported below, individuals who reported to sort household waste for recycling were asked why they did this. Of 1102 respondents, 73 percent agreed or partially agreed that one of their motives was a wish to regard themselves as a responsible person. 88 percent agreed or partially agreed that one motive was that they should do themselves what they would like others to do (for details, see Bruvold, Halvorsen and Nyborg [2000]).

¹¹ In social welfare optimum, $e_i = e_i^*$. Thus, the inclusion of the identity variable does not affect the optimization exercise above compared to a similar exercise without an identity variable, since $f(e_i^*, e_i^*) = 0$ regardless of the value of e_i^* .

function) defines what a good society is, while the Kantian rule determines the relevant assumptions to make about others' behavior when determining one's own morally ideal action.¹²

As all consumers are identical, subscripts denoting individuals can be omitted. This yields the first order condition

$$(7) \quad u_l = N u_G \gamma_e$$

where, as before, subscripts denote derivatives.¹³ Thus, if everyone contributed the morally optimal effort, the marginal utility of leisure for any individual would exactly equal the social benefits of the public good produced by the marginal effort. The maximization exercise implies that social efficiency is an important factor when considering the morally ideal action.

After solving the social welfare maximization problem, the individual maximizes utility (equation 1) subject to her time constraint, the public good functions, and the identity function (equations 2 - 5), taking the effort of all other individuals as given. This determines her actual behavior. In this latter problem, e^* is regarded as exogenous. The outcome gives a Nash equilibrium characterized by the first order condition

$$(8) \quad u_l = u_G \gamma_e + u_l(-2a(e_i - e_i^*))$$

evaluated at the Nash equilibrium level of the public good, $G = G_p + N g'$, where g' is the equilibrium contribution from the representative individual. Note that as all individuals are equal, they contribute the same amount of time.

¹² Thus, the perception of a good society is based on teleological or consequentialistic reasoning, while the view of one's own role in striving for a good society can be regarded as deontological.

(7) and (8) demonstrate that maximization of individual utility cannot yield the socially optimal allocation, even with preferences for maintaining self-identity as morally responsible. For this to be the case, (7) and (8) must be equivalent, which would be the case if $(N-1) u_G \gamma_e = 0$, since in the determination of the ideal effort, the latter term in (8) equals zero. Hence, this would require that $N = 1$ or that effort does not have any social benefits, i.e., $u_G = 0$ and/or $\gamma_e = 0$, which is not consistent with our assumptions. The intuitive explanation is that since effort is costly, individuals will only increase effort until a further increase will not pay off in terms of a better self-image. In the social optimum, the benefit in terms of a better self-image is, on the margin, zero. We will never have oversupply of individual provisions, i.e., it will never be individually optimal to provide more effort than $e = e^*$; in that case, the marginal effect on self-image would be negative. We can thus conclude that even with preferences for a positive self-image, we get underprovision of public goods.

This conclusion parallels that of the impure altruism model [Andreoni, 1990], but the reason for underprovision is different in the two models. The inclusion of a warm glow effect in Andreoni's model induces consumers to contribute more than otherwise, but it also increases the socially optimal provision, since the warm glow counts in the social welfare optimum. Underprovision is due to the public good element, and the underprovision will increase with N . In contrast, the inclusion of I in the utility function does *not* affect the social optimum in our model: By assumption, self-image is always at its maximum, i.e., $I = 0$, in the social optimum. The preference for a positive self-image, though otherwise similar to a warm glow effect, thus only affects the private utility maximization. However, we still get underprovision, because the marginal improvement in one's self-image goes to zero as effort approaches the ideal e^* .

¹³ Note that the term $u_i(-2a(e_i - e_i^*))$, disappears from the first order condition because in the welfare optimum above, everybody contributes the morally ideal effort, i.e., $e_i = e_i^*$.

III Increased efficiency

Assume now that an exogenous shift in the efficiency parameter θ occurs. For example, the government may introduce a curbside pick-up system for recycled household waste, implying that households can deliver the same amount of recycled waste as before, but still spend less time on recycling activities. In this case, e can be interpreted as time spent on recycling activities, while g is the resulting reduction of waste for end treatment as compared to a base case without recycling. To simplify further, we assume that the utility function is additively separable in (x, l) , G and I , and that the marginal utility of I is 1. The latter assumption is without loss of generality. The utility function thus becomes

$$(9) \quad U = u(x, l) + v(G) + I$$

where u and v are increasing and concave. The effect on the morally ideal *effort* can be found by differentiating the first order condition (7)¹⁴ from the welfare maximization with respect to θ .

$$(10) \quad e_{\theta}^* = N \frac{v_G \gamma_{e\theta} + N v_{GG} \gamma_e \gamma_{\theta}}{-u_{ll} - N^2 v_{GG} \gamma_e^2 - N v_{G} \gamma_{ee}}$$

As seen from (10), the denominator is positive. However, in the numerator there are two opposing effects: First, when θ increases, the marginal productivity of effort increases (as $\gamma_{e\theta} > 0$ by assumption), implying that the public good has become cheaper to produce on the margin. *Ceteris paribus*, this effect implies higher ideal effort. Secondly, for a given effort, an increase in θ will increase individual contributions g and thereby also the total public good provision G . This implies a reduction in the marginal utility of the public good, and the partial effect of this is that the ideal effort decreases. Thus, the ideal effort will generally change when θ increases, but we cannot tell a priori in

which direction. However, individual ideal *contributions*, i.e., $g^* = \gamma(e^*, \theta)$, will always be increasing in θ , as the second effect mentioned above disappears, and there is an additional positive effect on contributions from increased efficiency at a given level of effort ($\gamma_\theta > 0$). This is seen by differentiating (4) and using (10):

$$(11) \quad g_\theta^* = \frac{Nv_G \gamma_{e\theta} \gamma_e + \gamma_\theta (-u_{ll} - Nv_G \gamma_{ee})}{-u_{ll} - N^2 v_{GG} \gamma_e^2 - Nv_G \gamma_{ee}}$$

To study the effect of changed technology on *actual* effort, we differentiate the first order condition (8) with respect to θ . Assume that any single individual is unable to perceive the difference in environmental quality from a change in her own actual effort; i.e., $u_G \gamma_e = 0$. The first order condition for Nash equilibrium is then $u_l = -2a(e - e^*)$, i.e. individuals provide more effort until the benefit (in terms of a better self-image) obtained by a marginal increase in effort equals the marginal utility of leisure. The marginal effect on actual effort from a technology change is then given by

$$(12) \quad e_\theta = \frac{2ae^*_\theta}{-u_{ll} + 2a}$$

With standard concavity assumptions, the denominator is positive, and hence the effect on effort of increased efficiency depends on the sign of e^*_θ . As demonstrated above, e^*_θ may be positive or negative. Note that with an exogenously given ideal effort, there would be no effect on actual equilibrium efforts of a shift in θ .

New information, for example indicating that recycling has a higher environmental impact than previously thought, gives similar effects. In this case, let effort e_i denote the time spent on recycling,

¹⁴ The first order condition defines e as a function of θ . $u_l(x, T - e(\theta)) = Nv_G(N\gamma(e(\theta), \theta)) \cdot \gamma_e(e(\theta), \theta)$.

Differentiating with respect to θ and rearranging gives (10).

while g_i can be interpreted as the actual environmental impact caused by the individual's effort (note that this implies a slight reinterpretation of g as compared to the case of increased recycling efficiency discussed above). Effort is observable, while the environmental effect of a given individual's effort is unobservable. This may be the case even if total public good supply G is observable. A positive shift in θ may then be interpreted as higher *perceived* effect of private effort.

The effect on individual utility of a marginal increase in θ , evaluated at the Nash equilibrium characterized by (8), is given by

$$(13) \quad U_\theta = v_G \gamma_\theta + (N-1)v_G(\gamma_e e_\theta + \gamma_\theta) + 2a(e - e^*)e_\theta^*$$

Assume that $e_\theta > 0$ and $e^*_\theta > 0$, i.e., increased efficiency increases both actual and ideal effort.

Then the two first terms on the right hand side are positive, reflecting that increased efficiency will increase utility via a higher G . However, the last term is negative, and in general we cannot know which effect dominates.

If consumers did not have a preference for a good self-image, or if the ideal effort were exogenously given, increased efficiency of voluntary effort would change neither the ideal nor actual behavior. In Andreoni's [1990] impure altruism model, there is no explicit modeling of why there is a "warm glow of giving", and hence no modeling of how increased efficiency will affect the "glow".¹⁵ Perhaps the most striking difference between our model and one with an exogenous ideal is that in the latter case, increased efficiency could never lead to a decrease in individual utility. Increased efficiency would then simply increase the consumer's opportunity set, implying that the consumer might choose to behave just like before and thus get the same utility as before. However, in our model, increased

¹⁵ One might argue that the warm glow model is valid only when the public good is important, i.e., there is no warm glow if giving has no importance for the society. Also, increased efficiency or new information could be modeled as an increase in U_G .

efficiency may *decrease* utility, because increased efficiency can imply a higher morally ideal effort e^* , and thus the utility loss caused by deviations from the ideal increases.¹⁶

IV Responsibility and economic incentives

We will now introduce economic incentives into the model. Let us do this by means of a simple example. Assume an organization, for example a sports club, where an important task which will benefit all members (such as constructing a new soccer field) is to be done. In accordance with the Norwegian custom of *dugnad* mentioned in the introduction, the organization decides to perform the task on a volunteering do-it-yourself basis. Each individual chooses to show up at the *dugnad* ($e = 1$) or not ($e = 0$). The quality G of the task completed depends on the number of people that participate. Assume, as above, that the utility function can be written as $U = u(x, l) + v(G) + I$. Let g^p denote the individual contribution when participating.

The morally ideal effort e^* is found by considering whether participation or non-participation from every individual yields the highest social welfare. Since individuals are identical, the welfare maximization exercise implies that $e^* = 1$ if

$$(14) \quad u(m, T-1) + v(Ng^p) > u(m, T) + v(0)$$

i.e., if the increased public good supply has a higher value for the members than their lost leisure time.

Here, m is the individual's exogenous income measured in consumption units, which is all consumed (when a fee is introduced below, consumption x and income m may differ).

¹⁶ Within a wide range of plausible functional forms, utility will be non-decreasing in θ ; but an example where utility declines with increasing efficiency can be obtained from the authors. However, in a model with a heterogeneous population, declining utility can occur under fairly standard assumptions. For example, suppose that there are two equally large groups, both of N individuals. Individuals in group 1 have utility $\ln(l) + \ln(G) + (e-e^*)^2$, while those in group 2 have utility $\ln(l) + (e-e^*)^2$. For both groups, $l + e = 1$ and $G = \sum_i \theta e_i + G^0$. It is easily seen that the ideal effort (the effort maximizing $W = \sum_i U_i$) is $e^* = \frac{1}{3} - \frac{G_0}{3N\theta}$ for both groups. Thus, the ideal effort is increasing in efficiency. Clearly, increased efficiency is then a disadvantage for group 2, who have no direct benefits from the public good. A thorough analysis of moral motivation in a heterogeneous population is a subject for further research, however.

If (14) holds, individuals can improve their self-image through participation, but at a cost of less leisure. Then, full participation is a Nash equilibrium if

$$(15) \quad u(m, T-1) + v(Ng^p) > u(m, T) + v((N-1)g^p) - a$$

i.e., if the utility loss due to less leisure is outweighed by the benefits in terms of increased public good supply and a better self-image. Similarly, no participation is a Nash equilibrium if

$$(16) \quad u(m, T) + v(0) - a > u(m, T-1) + v(g^p)$$

Due to the concavity assumptions, (15) and (16) cannot hold simultaneously, and there will be a unique Nash equilibrium.

Now, assume that a fee t for non-participation is introduced. Suppose first that $t = c$, where c is the cost of buying g^p units of the public good in the market, implying that the collected fees are exactly *sufficient* for the organization to increase the public good by Ng^p even without participation. Both t and c are measured in consumption units. Assume that each individual knows that $t = c$, and that everyone also assumes that the purpose of the fee is to buy services in the market to ensure that $G = Ng^p$. Then, $G = Ng^p$ will be exogenously fixed when individuals determine e^* , since no matter what one chooses to do, the public good provision equals Ng^p . One interpretation of this situation is that individuals leave the responsibility of ensuring a sufficient public good supply to the organization's leadership.

Non-participation ($e^* = 0$) is now the morally ideal action if

$$(17) \quad u(m, T-1) + v(Ng^p) < u(m-c, T) + v(Ng^p)$$

i.e., if a unit of leisure is more valuable to individuals than the consumption equivalent of the fee. In this case, individual utility maximization will yield non-participation too. Full participation, $e = 1$, will be a Nash equilibrium only if

$$(18) \quad u(m, T-1) + v(Ng^p) - a > u(m - c, T) + v(Ng^p)$$

which is inconsistent with (17). Thus, $e = 0$ if $e^* = 0$.

It is perfectly possible that both (14) and (17) hold, i.e., that the project is socially desirable, but that members would prefer market provision, financed by a fee, to participation in the *dugnad*. In this case, the introduction of a fee will weakly reduce participation: If full participation is the unique Nash equilibrium initially, no participation may become the unique Nash equilibrium when a fee is imposed, i.e., both (15) and (17) hold. Thus, the introduction of a fee may have the unexpected effects mentioned in the introduction.

On the other hand, suppose that the fee is merely *symbolic*; intended to serve as an incentive to participate, but not nearly sufficient to finance external provision: $t < c$. The tax income is used to buy the public good in the market, but the services bought cannot replace all the work that was supposed to be done voluntarily. Hence, if people do not show up, the quality of the public good will be affected.

Then, $e^* = 1$ if

$$(19) \quad u(m, T-1) + v(Ng^p) > u(m-t, T) + v(Ng^p t/c)$$

i.e., if the loss of leisure by participation counts less for the individual than the fee payment and the lower public good supply resulting if nobody participates.

In this case, $e = 1$, i.e., participation is the Nash equilibrium if:

$$(20) \quad u(m, T-1) + v(Ng^p) > u(m-t, T) + v((N-1)g^p + g^p t/c) - a$$

Assume that both (14) and (19) hold. In this case, the fee does not change the morally ideal effort $e^* = 1$. Further, a reasonable assumption seems to be that $(u(m-t, T) + v(g^p t/c))$ is decreasing in t ; implying that in a situation with no initial public good supply, the individual cannot increase her utility by "imposing a fee on herself" and buy g^p units of the public good individually in the market. Under these assumptions, the right hand side of (20) is less than the right hand side of (15). Thus, a symbolic fee weakly *increases* participation. At the same time, however, a more substantial fee, which is sufficient to cover costs, may *decrease* participation.¹⁷

Note that the essential feature of the fee is not its size as such, but whether it is *perceived* to be symbolic or sufficient. If perceived as sufficient to buy the required services in the marketplace, the fee gives moral justification for not showing up. This implies no problem for the organization if perceptions are correct. However, if members incorrectly perceive the fee to be sufficient to finance external supply, and introduction of a fee changes the Nash equilibrium from full participation to no participation, reduced welfare will be the result.

Note also that the reason for the different effects of a symbolic versus a sufficient fee is that the symbolic fee leaves the responsibility for public good supply with the individual, while a sufficient fee shifts this responsibility to the organization. Within social psychology, the cognitive evaluation theory [Deci and Ryan, 1985] maintains that external interventions (such as a monetary reward) reduce intrinsic motivation if the intervention is perceived as controlling, but may increase intrinsic motivation if perceived as acknowledging. Our context differs somewhat from that of Deci and Ryan; in particular, their discussion is related to motivation to perform interesting tasks, not moral

¹⁷ A similar result is obtained in the symbolic fee case if we assume that the planned public good supply Ng^p is realized through market supply if members do not show up, and that the excess costs $((c-t)$ times the number of non-participants) are distributed equally among members.

motivation as such. Nevertheless, in both approaches the individual's feeling of own responsibility clearly plays a key role for behavioral effects.

V Empirical evidence

To test the empirical relevance of the assumptions and the theoretical implications of the model, we used Statistic Norway's Omnibus survey for November-December 1999. This is an interview survey which is conducted 4 times a year. A large number of background variables are included routinely, while additional questions are included in the survey questionnaire on a payment-per-minute basis. The November-December survey of 1999 had a gross representative sample of 1982 individuals of age 16-79, and a response rate of 59 percent.¹⁸ The main sampling method was in-person interviews conducted at home visits (76%), although some interviews were conducted by telephone (24%). Our survey questions were concerned with recycling of household waste and participation in *dugnads*.

A. Utility and increased efficiency

As explained above, increased efficiency of voluntary contributions may *decrease* utility in our model. Thus, if local authorities arrange for easier delivery of recycled household waste, so that the same amount of waste can be recycled with less effort (for example by introducing or extending curbside collection), we would like to know whether at least some consumers consider this a personal disadvantage. Table I reports our results on this issue.

(Table I about here)

¹⁸ The net sample has a slight overrepresentation of individuals aged 25 to 44 and underrepresentation of those between 67 and 79. There is also a slight underrepresentation of people from the Eastern part of Norway and overrepresentation from the less urban Southern and Western parts.

Within the framework of a traditional warm glow or pure altruism model, it is hard to see why 26 percent of respondents regard an extended collection system as a disadvantage for themselves.¹⁹

Within our model, this is not unexpected.

B Effects of information

Secondly, in our model, consumers care not only about the size of their contribution per se, but on the contribution's effects on social welfare. Thus, information indicating that voluntary contributions are more important for social welfare than previously thought would be expected to change the morally ideal effort, and thus also, in general, actual effort.

As discussed in section III, information that recycling is more important for the environment has similar effects as an increase in the efficiency parameter θ . Provided that $e^*_\theta > 0$, the prediction of our model is that actual recycling will increase.²⁰

One half of the sample were asked the following question: *If you got new information that recycling is more important to the environment than earlier assumed, would you recycle as before, more, or less than before?* 46 percent would recycle more, while only 2 percent would recycle less. The other half of the sample were asked the same question, but with "more important" replaced by "less important". Correspondingly, 29 percent would reduce their recycling, while only 3 percent would increase it.

These results are intuitively very unsurprising. Nevertheless, as discussed above, information effects are not explicitly modeled in warm glow models. If we for example replace (5) with $I_i = f(e_i)$, in which case our model may be interpreted as one version of Andreoni's [1990] impure altruism model, such effects would not occur, since warm glow is then connected to contributions as such, not to their social importance.

¹⁹ Costs of the public collection system is not considered here; this would of course have to be taken into account in a less partial model.

C The effects of economic incentives

Third, according to our theory, a fee for non-participation in voluntary community work may *decrease* participation: The fee will add an incentive to participate if it is perceived as symbolic, but give moral justification for not showing up if it is perceived as sufficient to buy the required services in the marketplace. The impure altruism model cannot explain such behavior: A fee on non-participation corresponds to a subsidy on voluntary contributions, which *increases* contributions in Andreoni's [1990] model. The described behavior may seem more consistent with Frey's [1997] crowding theory. However, in the crowding theory, there is no presumption that effects should depend on whether the fee is sufficient to cover costs or not.

In our survey, 69 per cent of the sampled population were members of organizations relying on *dugnads* (voluntary community work), and only 12 per cent of these respondents reported that they never participate. All members of organizations using *dugnads* received the question reported in Table II (responses in percentages):

(Table II about here)

All respondents received both the "enough" and "not enough" versions.²¹ Note in particular that as much as 15 per cent would *reduce* their participation when the fee is enough. This seems hard to explain within the framework of alternative theories, but fits nicely with the predictions of our model.

Our data concerns hypothetical rather than actual choices, which may be considered a weakness. Frey and Goette [2001], using data on actual volunteer work in Switzerland, find that the *incidence* of monetary rewards reduce volunteering, while the *size* of rewards increase volunteering (see also Gneezy and Rustichini [2000a]). When a monetary reward is in use, this might possibly lead

²⁰ In this particular setting, the formulation "recycling is more important" makes it reasonable to assume that $e^*_\theta > 0$.

²¹ The order of these questions was reversed for half of the sample, but this did not appear to affect the results substantially.

individuals to think that the organization can reach its desired level of public good supply through adjustments of the reward, in which case the self-image benefit obtained by volunteering is reduced or eliminated. However, a larger reward provides a stronger economic incentive to volunteer. A field study of an introduction of a fine is given in Gneezy and Rustichini [2000b]. Here, the effect of imposing a fine on parents arriving late to collect their children at day care is studied. The introduction of a fine actually increased the number of late-coming parents, and it remained stable after the fine was cancelled. For a survey of empirical evidence on crowding-out effects of monetary incentives, see Frey and Jegen [2001].

VI Conclusions

The simple model presented in this paper illustrates some important points. First, true moral motivation is not incompatible with utility maximization; and secondly, public policy may have indirect effects on behavior through its effects on moral motivation.

In our model, consumers have a preference for thinking of themselves as socially responsible. In this respect our model has similarities to "warm glow" models, since seemingly unselfish behavior is ultimately motivated by a private good. However, there are important differences. In the present model, a consumer's self-image can only be improved by striving towards what she truly believes to be morally right. This implies that consumers are morally sophisticated, in the sense that they do elaborate on the social welfare consequences of alternative rules of behavior, and they do take their conclusions on this into account when making actual choices.

Still, the benefits of a good self-image will be traded against its costs. As in Andreoni's [1990] impure altruism model, our model implies that there will always be underprovision of public goods, although the reasons for underprovision are different. In contrast to Andreoni's model, however, our theory implies that economic incentives for voluntary contributions may have adverse effects on

contributions. Public policy affects behavior not only through its effect on relative prices and budget and/or time constraints, but also through the policy's effects on individuals' perception of the morally ideal action. Survey data on recycling and voluntary community work presented above is consistent with these ideas.

Our prediction that monetary incentives may decrease morally motivated contributions is shared with Frey's (1997) crowding theory. According to crowding theory, monetary rewards will reduce moral motivation if perceived as controlling rather than acknowledging. In our model, such adverse effects may occur if the monetary reward shifts the perceived responsibility for public good supply from the individual to the governmental level.

References

- Akerlof, G. A., and Kranton, R. E., 2000. Economics and identity, *The Quarterly Journal of Economics*, CXV (3), 715-753.
- Andreoni, J., 1988. Privately Provided Public Goods in a Large Economy: The Limits of Altruism, *Journal of Public Economics* 35, 57-73.
- Andreoni, J., 1990. Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving, *The Economic Journal* 100, 464-477.
- Bernheim, B. D., 1986. On the Voluntary and Involuntary Provision of Public Goods, *American Economic Review* 76 (4), 789-793.
- Brekke, K. A., and Howarth, R. B., 2000. The Social Contingency of Wants, *Land Economics* 76, 493-503.
- Brekke, K. A., Kverndokk S., and Nyborg K., 2000. An Economic Model of Moral Motivation, Discussion Paper no. 290. Statistics Norway, Oslo.
- Brennan, G., and Hamlin A., 2000. *Democratic Devices and Desires. Theories of Institutional Design* series. Cambridge University Press, Cambridge.

- Brennan, G., and Lomasky, L., 1993. *Democracy and Decision. The Pure Theory of Electoral Preference*. Cambridge University Press, Cambridge.
- Bruvoll, A., Halvorsen, B., and Nyborg, K., 2000. Household Sorting of Waste at Source, *Economic Survey* 4/2000, 26-35. Statistics Norway, Oslo (available at <http://www.ssb.no>).
- Deci, E. L., and Ryan, R. M., 1985. *Intrinsic Motivation and Self-Determination in Human Behavior*, Plenum Press, New York.
- Frey, B. S., 1997. *Not Just for the Money. An Economic Theory of Personal Motivation*. Edward Elgar, Cheltenham, UK.
- Frey, B. S., and Goette, L., 2001. Does Pay Motivate Volunteers? *Review of Economics and Statistics*, forthcoming.
- Frey, B. S., and Jegen, R., 2001. Motivation Crowding Theory: A Survey of Empirical Evidence. *Journal of Economic Surveys*, forthcoming.
- Frey, B. S., and Oberholzer-Gee, F., 1997. The Cost of Price Incentives: An Empirical Analysis of Motivation Crowding-Out, *American Economic Review* 87 (4), 746-755.
- Gneezy, U., and Rustichini, A., 2000a. Pay enough or don't pay at all, *Quarterly Journal of Economics*, CXV (3), 791-810.
- Gneezy, U., and Rustichini, A., 2000b. A Fine is a Price, *The Journal of Legal Studies*, 29(31), part 1, 1-17.
- Harsanyi, J. C., 1955. Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility, *Journal of Political Economy* 63, 309-321.
- Holländer, H., 1990. A Social Exchange Approach to Voluntary Cooperation, *American Economic Review* 80, 1157-1167.
- Lindbeck, A., 1997. Incentives and Social Norms in Household Behavior, *American Economic Review* 87 (2), 370-377.
- March, J. G., and Olsen, J. P., 1995. *Democratic Governance*. Free Press, New York.
- Margolis, H., 1982. *Selfishness, Altruism, and Rationality*. University of Chicago Press, Chicago.
- Mead, G. H., 1913. The social self, *Journal of Philosophy* 10, 374-380.

- Nyborg, K., 2000. Homo Economicus and Homo Politicus: Interpretation and Aggregation of Environmental Values, *Journal of Economic Behavior and Organization* 42, 305-322.
- Nyborg, K., and Rege, M., 2000. The Evolution of Considerate Smoking Behavior, Discussion Papers 279. Statistics Norway, Oslo.
- Sen, A. K., 1977. Rational Fools: A Critique of the Behavioural Foundations of Economic Theory, *Philosophy and Public Affairs* 6, 317-344.
- Sugden, R., 1984. Reciprocity: The Supply of Public Goods Through Voluntary Contributions, *The Economic Journal* 94, 772-787.

Table I. Personal utility effects of extended recycling arrangements. Number of respondents = 990.²²

<i>Assume that the municipality arranges for more recycling in homes.²³ Recycling is voluntary. Which of the following statements do you mostly agree on?</i>	Responses, percent
It is good that the environment is taken more into account, but for me personally it is a disadvantage that more effort is expected	26.0
It is good that the environment is taken more into account, and for me personally it is an advantage that I now can increase my effort ²⁴	39.7
It would not mean anything to me	33.7
Do not know	0.6

²² This question was posed only to those who responded "yes" to the following: "Do you think recycling contributes to a better environment?".

²³ In retrospect, we see that "arranges for more recycling" is not a very precise phrase. We still believe the most reasonable interpretation of the question is that voluntary efforts are facilitated through extended collection systems.

²⁴ Note that the Norwegian term used for "effort" ("*innsats*") can also be taken to mean "contribution". The same word was used in both response alternatives.

Table II. Effects on *dugnad* participation (voluntary joint community work) of a fee for non-participation. Percent. Number of respondents = 802²⁵

<i>"Suppose that you have to pay NOK²⁶ 100 as an extra subscription if you do not participate in the dugnad. This is enough (not enough) to pay professionals to do the job. Would you participate more often, more seldom, or would it not affect your participation in the dugnad?"</i>		
	The fee is enough to pay professionals	The fee is not enough to pay professionals
I would participate more often	10	19
I would participate more seldom	15	3
It would not affect my participation	75	77
Do not know	1	1

²⁵ The question was not posed to those who said they were not members of organizations using *dugnads* (31 percent).

²⁶ 100 NOK = 10,80 USD (Dec 3rd, 2000).