CHAPTER

20

Neuroeconomics of Charitable Giving and Philanthropy

Ulrich Mayr, William T. Harbaugh, and Dharol Tankersley

OUTLINE

INTRODUCTION

This chapter examines what neuroscience can tell us about people's motivations to pay for the provision of public goods and services that benefit society as a whole. The focus is on voluntary decisions about charitable giving. However, since most funding for public goods comes from taxation, we also consider the motives that lead people to vote for tax systems that compel themselves and others to pay jointly for public goods. We concentrate on research that adopts the usual economic assumptions of stable preferences and rational choices, but also consider the consequences of alternative, weaker assumptions.

The chapter starts with standard neoclassical utility theory and the implications of the assumption of selfish preferences, or what is rather confusingly called Pure Altruism. We show that the implications of this model are at odds with many real-world and

laboratory behaviors involving voluntary contributions to public goods. Most glaringly, while this model predicts free-riding – benefiting from public goods without contributing to them – by all but a few donors, in reality charitable giving is common, if not quite universal. This suggests that the Pure Altruism model does not provide a complete account of behavior in public good situations – it does not accurately reflect the constraints that people face, or it does not accurately incorporate people's fundamental preferences, or people are making systematically irrational decisions.

Given this conflict, we then explore some alternatives to the Pure Altruism model, starting with the possibility that repeated interactions or attempts to signal one's wealth or character might explain contributions, before considering "warm-glow" models of altruism, where people derive a benefit from the act of giving. We show how these changes in constraints and preferences imply behaviors that are substantially different from those in the Pure Altruism model, and argue that these alternative models are useful for explaining a substantial amount of observed giving behavior.

The second part of the chapter looks at current neural evidence that addresses the reasons why people sometimes show altruistic behavior. After showing that the few studies that have directly dealt with motives indicate evidence of both Pure Altruism and warm-glow motives, we point to a number of open issues that deserve further study.

One question worth discussing at the outset is why neural evidence should be useful for testing models of public good decisions. One simple reason is that neural data provide a way of testing economic models that is very different from the empirical and experimental data that have typically been used. This presents the possibility of independent evidence in favor of, or against, various models of behavior. Another reason is that the neural evidence builds on a rich base of knowledge from animal physiology and from human brain-imaging work about the functional relevance of specific neuroanatomical areas. This means that there is the potential for generating interesting new hypotheses about motivations from observed patterns of neural activations. Moreover, results of neuroeconomics studies can feed back into the neuroscience literature, and further increase our general understanding of how the brain makes decisions and experiences their consequences.

Of course, these benefits are not specific to the domain of public good decisions, but hold for any economic question. More specific to public good decisions is the fact that neural data can provide information that it is very hard to obtain with other methods. Most existing economic studies on the reasons for giving are concerned with behavior, which provides only an indirect and one-dimensional window onto the mechanisms supporting altruistic motives. More direct ways of assessing motives, such as questionnaires, are often even less satisfying. People's self-reports on why they give – or don't – are suspect, since responses may reflect social demand effects, or the desire to report what they believe is socially acceptable or will please the experimenter, rather than their actual motives. In contrast, it should be quite hard for a person to "simulate" specific patterns of brain activity. Thus, neural evidence holds the promise of providing direct and physical evidence of the "hidden motives" behind altruistic behavior.

SOME ECONOMICS OF PUBLIC GOODS AND CHARITABLE GIVING

Introduction

We start by defining public goods and explaining why, given standard economic assumptions about people's preferences, private markets may not provide these goods adequately. After discussing how charitable giving and taxation can be used to provide these goods, we turn to the issue of voluntary contributions in more detail, focusing on the question of what motivates the contributions. This leads to discussion of how neuroscience can address those motivations.

Public Goods Defined

Most charity involves what economists call *public goods*. The definition of a public good begins with a classification system that distinguishes goods or services on two dimensions: whether or not one person's use of the good diminishes its value for others (rival or non-rival), and whether or not a person who doesn't help to pay for the good can be prevented from using it (excludable or non-excludable). Goods that are rival and excludable, such as food, housing, and medical care, are called *private goods*; goods that are both non-rival and non-excludable are called *public goods*. Classic examples of the latter are national defense, environmental quality, and knowledge. The fact that one person benefits from these things does not diminish their value to other people, and people can't be prevented from enjoying the benefits these goods produce.

Not all goods fit nicely into the private/pure public categories. Some goods pass one test but not the other. *Congestion goods* are rival, but non-excludable – for example, the use of a public road during rush hour. Since technological improvements tend to reduce the cost of exclusion, congestible goods can, with time, become private. Barbed wire is a classic example of a technology that made it feasible to create private property rights to grazing land. GPS toll-collecting technology may do the same with roads. *Club goods* are goods that are non-rival but excludable. Examples include concerts and classroom education – at least until there is crowding, one more user doesn't diminish the benefits others receive. However, people who don't pay can easily be kept from consuming these goods. Population increases tend to make non-rival goods rival – for example, as roads become crowded.

Aid to the poor, say for food and shelter, is another sort of good that doesn't fit perfectly into the classification scheme. Money spent to help a poor person creates a private good to the recipient; however, as Kolm (1969) points out, it also provides a non-rival and non-excludable benefit to all who want to live in a world where poor people are well fed and housed. Everyone who has this preference benefits from greater welfare expenditures, even if they haven't helped to pay for them. In this sense, the welfare of the poor is a non-rival and non-excludable public good.

Some voluntary giving is directed toward non-public goods. For example, monetary transfers between family members are common and often large. These gifts don't provide non-rival benefits (outside of other family members, at least), and are probably better studied using models that incorporate voluntary exchange and reciprocity. Similarly, gifts to education, the arts, and churches are often used to subsidize excludable goods. Gifts to provide congestible goods (such as scholarships to college) are also common. These are often thought of as charitable gifts as well, and to some extent most of the discussion in this chapter will apply.

Pure Altruism and the Optimal Level of Public Goods

Neoclassical economic models begin with the assumption that people choose from the available options so as to maximize their utility, or satisfaction. The simplest version of this that works for our purposes comes from Samuelson (1954). We start with the assumption that each member of society cares about his or her consumption of a private good x (usually thought of as total dollars spent on private consumption) and a public good G. We say that people maximize a utility function of the form $u_i = u_i(x_i, G)$ where i indexes each individual. Note that this is a model of Pure Altruism: by assumption, people care about how much G there is, but they do not care at all about how that level is achieved, except insofar as it affects their consumption of x – for example, when they make a contribution or pay a tax that reduces their x_i, but increases G. In other words, the standard model assumes that people care about the level of the public good only insofar as it contributes to their own consumption.

The fact that a public good is non-rival means that, in determining the socially optimal or efficient quantity of the good, we must consider the benefits that *everyone* will get from a given unit of the good. This is why G appears without a subscript and is in contrast to private goods, where only one person can consume the good and so only one person's benefit counts.

To keep things simple, we will assume two potential donors, person 1 and person 2, and measure utility in dollars. We will use the welfare of the poor as the public good; this benefits all people who care about the poor, so the benefit to "society" from an additional increment must include the benefits to both 1 and 2. (For simplicity, the private benefit to the poor is ignored here.) As a numerical example, assume that each person has an endowment m of \$5, and each unit of welfare for the poor of G costs \$4 to provide. Each unit of G benefits person 1 by \$3 and person 2 by \$6, so we are assuming that person 2 derives the larger benefit from seeing the poor taken care of. Each person can either keep all their \$5 or give \$4 to buy one unit of welfare. The total level of G is the sum of the amounts provided by persons 1 and 2, or $G = g_1 + g_2$. We could write these utility functions as $u_1 = x_1 + 3G$ and $u_2 = x_2 + 6G$, and the constraints as $x_i + g_i <= m_i$. Thus, person 1's utility is his starting \$5 minus \$4 if he donates, plus \$3 times the number of units of welfare G. Person 2's utility is his starting \$5 minus \$4 if he donates, plus \$6 times G. Each unit of the public good benefits both persons 1 and 2, so the combined utility for the two is given by $u_1 + u_2 = 5 + 5 - 4G + (3 + 6)G$. Obviously this is increasing G, since the cost of a unit of G is 4, while the combined benefit to persons 1 and 2 is $3 + 6 = 9$, so it's socially optimal to provide the maximum possible 2 units of G. This principle, that in finding the optimal level of G we need to weigh the costs against the benefit to *both* persons 1 and 2, follows directly from the non-rival nature of such goods, or from the fact that each unit brings benefits to more than one person. This optimality rule is known as the *Samuelson Condition*.

		Person 2	
		Keep	Give
Person 1	Keep	5, 5	8, 7
	Give	4, 11	7, 13

FIGURE 20.1 Modeling charitable giving as a strategic game. Each person starts with \$5 and can buy one unit of the public good *G* for \$4. Person 1 gets \$3 in benefit from *G*, person 2 gets \$6. Each cell shows the net payoffs to person 1, person 2 from that outcome. The total payoff is highest when both give, but person 1 does better for himself by keeping his money.

The Pure Altruism Model Implies Insufficient Levels of Public Goods

Since public goods are non-excludable, it is clear that a society that relies on voluntary action is going to have a problem in achieving the optimal level of *G* (Becker, 1974; Bergstrom *et al.*, 1986). A simple version of the argument using the example from above is shown in Figure 20.1. Each person can either keep all his \$5, or give \$4 to buy a unit of the public good. There are two choices for each of the two people, so four possible outcomes. The numbers in each cell show the payoff to person 1, then to person 2.

This situation is a modified version of the prisoners' dilemma from strategic game theory. Person 1 gets a higher payoff from keeping than giving, regardless of what 2 does, since $5 > 4$, and $8 > 7$. The twist is that person 2, because of the high benefit he gets from the public good, would prefer to give regardless of what person 1 does, since $7 > 5$ and $13 > 11$. In total, the two are better off if they both give, but only person 2 will do so voluntarily, thus we will end up with one unit of *G*, rather than two. Economists say that person 1 will *free-ride* on the donation of person 2, meaning that he will consume the *G* which 2 paid for, without making a contribution himself.

Mandatory Provision and Taxation

The usual solution to this free-riding and the inefficiently low level of *G* is some sort of political mechanism to determine how much *G* society wants, coupled with mandatory taxes. Funding public goods with taxes raises some issues, however. First, the process that determines the level of the public good is necessarily imperfect. People will generally have incentives to misrepresent their most desired level of the good if others will be helping to pay for it. Second, taxation generally means that even people who don't want the good have to pay for it, leading to arguments about fairness and coercion. Third, raising revenue by taxation is generally inefficient. Income tax, for example, distorts people's work effort away from the most efficient level. Last, governments and bureaucracies rarely have an incentive to provide public goods in the cheapest possible manner, since they don't get to keep the surplus that would result if they did.

Thus, the funding of public goods is a balancing act – both voluntary and involuntary mechanisms present serious problems. In general, modern societies rely much more on taxation than on voluntary giving to provide public goods. However, for specific goods (such as the arts, some kinds of medical research, some social welfare programs, and higher education) voluntary giving can be quite important. In the US, about 3% of GDP enters the "voluntary sector" through charitable giving. Interestingly, most other modern economies have much lower rates of giving.

Other Arguments Against the Pure Altruism Model

The free-riding result from the above model of Pure Altruism, where people care about the level of *G* and not how it is achieved, is straightforward and, perhaps unfortunately, very intuitive even to people without formal training in game theory. However, clearly not everyone always free-rides when given the opportunity, even when the benefits from another unit of *G* are less than the cost. In fact, by some measures, the amount of free-riding is shockingly low. Data from the Independent Sector estimate that, in 2003, 89% of US households gave something, with the average amount among those giving being \$1620. Results in economic experiments designed to simulate the main aspects of contributions towards a public good also find large numbers of people giving substantial amounts of money (Isaac and Walker, 1988; Andreoni, 1995). In addition to free-riding, the Pure Altruism model makes other strong predictions about the voluntary provision of public goods. We start by examining those predictions, then discuss the evidence concerning them, before concluding this section with a discussion of alternative models of voluntary giving that are more consistent with the evidence.

Evidence About Who Gives and How People Respond to Changes in Levels of Public Goods

One interesting prediction of the Pure Altruism model is that government expenditures on a public

good will "crowd out" or reduce private contributions towards that good – potentially, dollar for dollar (Bergstrom *et al*., 1986). The logic for this is straightforward. Donors give until they just balance the extra utility they get from consuming the private good with the extra utility they get from consuming the public good. If the government takes a dollar in taxes from them and uses it to provide more of the public good, donors can get back to their original, optimal situation by reducing their giving by the dollar amount of the tax and spending the money on the private good. So, in theory, the only way government provision can increase the level of the public good is if the taxes are collected from non-donors. In practice, this is not what happens. Government expenditures do crowd out some private giving, but it's usually substantially less than dollar for dollar. Kingma and McClelland (1995) re-analyzed data from an earlier study involving donations to National Public Radio, and found almost no crowd-out effect.

Another significant challenge to this model comes from Andreoni (1989). He shows that, under very general assumptions (compared with the restrictive ones used in Figure 20.1), the Pure Altruism model predicts complete free-riding by all but a few people with the highest benefits from the public good, even in groups as small as 20. As the number of people benefiting from the good increases, the predicted number of donors shrinks to one, and the amount of the good provided stops growing. Significantly, this theoretical result holds even when people have very strong preferences in favor of the public good. A simple version of this can be seen in Figure 20.1. Person 1 cares quite a bit about the public good – but person 2 cares still more, and person 1 does not care enough to "top off" person 2's contribution. This shows the difficulty in using behavioral information to make inferences about preferences, when it comes to public goods.

Again, this predicted free-riding is not what actually happens. Participation in charitable giving is remarkably widespread, and even low-income people give, despite the fact that they would seem to benefit more from spending money on their own consumption than from the small increases in the level of public goods that their gifts can provide. In the US, for example, data from the independent sector for 1995 show that nearly half of households with incomes below $10,000 gave, with the average donation from those that gave exceeding $300 (Andreoni, 2006).

So, the Pure Altruism model fails three straightforward empirical tests: there is considerably less free-riding than the model predicts; people don't respond to government-financed changes in the amount of the public good by reducing their contributions dollar for dollar; and people who might not be expected to give, such as those with low incomes, actually do give. This does not mean that purely altruistic motives do not exist, but it does mean that we need something else to explain why people give their money and resources to charitable causes. Here, three broad classes of explanations are considered. First, we examine explanations that don't require altruistic preferences of any sort. Behavior might look altruistic because we don't understand the true constraints and structure of the game – what seems on the surface to be altruism might actually just be giving to signal wealth, or reciprocation in a repeated game. Next, we turn to the possibility that the Pure Altruism model of preferences is wrong and, in particular, consider models where the amount of a person's gift – as distinct from the level of the public good – is in the utility function. Both these explanations assume that people are rational. Finally, we consider the possibility that people are not rational, and turn to non-economic explanations for altruism.

Alternative Models of Altruism

The Pure Altruism model described above is sparse – maybe too sparse. One possibility is that the constraints or the rules of the game are unrealistic. For example, suppose that a reputation as a giver endears you to people and makes them willing to pay more for the products you sell. In this case, even a selfish person might give. The argument might be that in these circumstances it doesn't cost you a dollar to give a dollar to charity; instead, it actually pays off in more than a dollar in profits. This behavior is not really motivated by altruistic concerns, and could be modeled with a straightforward change in the budget constraints faced by the giver. Another possibility is that our model of preferences is wrong. Perhaps people not only care about the social problem, but also feel particularly good when they personally take voluntary action to help with the solution. As we will show, this can be modeled with a change in the utility function, as is done in "warm-glow" models of altruistic behavior.

In practice, economists sometimes use a "reduced-form" model to deal with the first possibility above – that our model of constraints is too simple. Reduced-form models subsume elements of the actual model structure into a simplified overall representation. This can be done by writing out a utility function that incorporates the additional constraints implicitly, using the argument that the behavioral predictions of the model are the same as would be obtained with simpler preferences and explicit constraints

(Harbaugh, 1998a, 1998b). But one of the major advantages of neuroeconomics is the ability to obtain direct readings of the neural activity associated with preferences, and so we argue that it is essential to be explicit about modeling changes in the conditions of the situation as changes in constraints, and likewise with changes in preferences.

Constraints

Charitable decisions typically involve repeated interactions between people who know each other. A basic result from game theory is that optimal behavior in a situation where people interact repeatedly can be very different from that in a one-shot setting. We can explore this by modeling charitable giving towards a public good as a simple game with n players. Assume that each player has m units of money, which they can keep or contribute to the public good. The sum of the individual contributions determines the level of G, and an individual's payoff is $m - g + \alpha G$. Note that this is a simple form of the Pure Altruism model $u = u(x, G)$ where $u = x + \alpha G$. By design, $\alpha < 1$ so it's optimal for each individual to give nothing, regardless of what the other players do, while $\alpha n > 1$, so the group does better if everyone gives all of their endowment. If $n = 2$ and α is set to, say, 2/3, this game is the classic prisoners' dilemma. (This set-up differs from the game in Figure 20.1 in that α is the same for everyone and is always less than 1.) In a one-shot game of this sort, *not* contributing is the optimal strategy for each player.

However, many giving decisions involve repeated decisions by people who can identify each other and who have some memory of what happened last time. Repeated interactions (with an indeterminate last round) change the game and the optimal strategy. Computer simulations with repetition show that Rapaport's tit-for-tat strategy (Axelrod, 1984), in which players start by contributing and then reciprocate their partner's previous choices, does substantially better at maximizing individual payouts than does the don't-contribute strategy.

Experiments with people, in either two-person or multi-person set-ups, produce very similar results. People frequently play tit-for-tat or variations on it, and the end result is considerably more voluntary giving – and higher individual payoffs – than the one-shot model predicts (see Andreoni *et al*., 2008, for a review). We argue that giving in these settings should not be called altruistic, and we think that studies which incorporate neural evidence are well suited for further investigation of this issue, because they offer the potential for revealing distinct physiological mechanisms supporting giving behavior in situations where they stand to benefit from contraints of reciprocity versus in situations where there is no personal material gain to be derived from giving.

Repetition of the same game is not the only interesting situation involving complicated constraints. Another sort of model involves the possibility of incomplete information. For example, people might not know each other's ability, wealth, or preferences. You might only want to do business with people who are by nature somewhat altruistic, because they are less likely to cheat you, thereby creating a need to both acquire information about others, and signal information about yourself to potential partners. Talk is cheap, but charitable giving is a way to credibly signal your altruistic nature. Giving is cheaper the more you care about others, and this makes it hard for selfish people to mimic altruistic ones (Ariely *et al*., 2007). A similar argument can be made about giving as a signal of wealth (Glazer and Konrad, 1996) or (for the arts) taste and sophistication.

Warm-glow and Impure Models of Altruism

The Pure Altruism model used so far assumes $u = u_i(x_i, G)$. While people care about the level of the public good, G, they don't care how that level is achieved. But people often say that "they want to make a difference," or that they get a good feeling from making a sacrifice that benefits others.

One way to incorporate this preference for a personal contribution to G is with the warm-glow model of giving (Andreoni, 1989, 1990). The most general form assumes preferences are given by $u = u(x_i, g_i, G)$. People still care about the overall level of the public good, G, but now they also care about their own contribution to the good, g_i. This is known as imPure Altruism, since it contains both the pure and the warm-glow motives. (Note that while Pure Altruismis the benefit a person gets from the level of the public good, warm-glow altruism is, in a certain sense elaborated on below, more selfish. If the public good is welfare for the poor, then warm-glow altruists give because of the good feeling *they* get from giving, while pure altruists give to make the poor better off.)

This apparently minor change in the assumptions about preferences has significant effects on giving behavior. In particular, none of the three empirical facts noted above – incomplete free-riding, frequent contributions from low-income/low-demand donors, and incomplete crowding out of private donations by

government provision of a public good – are implied by the impure model of altruism. This follows from the assumption that charitable contributions provide a direct private benefit to the giver: the only way to get this benefit is to give. Tests using data on charitable giving (Ribar and Wilhelm, 2002) and from economic experiments (Palfrey and Prisbrey, 1997; Goeree *et al*., 2002) provide support for this impure model, but with differing degrees of weight on the pure and the warm-glow elements. Later in the chapter, we show how neural data can also be used to test this model.

It seems strange to argue that people get a private benefit from giving up money, so the warm-glow model immediately brings to mind the question of where the preference for g_i originates. As mentioned above, one view of warm-glow giving is that it is essentially a simplified or "reduced-form" way of modeling the effect of repeated play or incomplete information. The argument is that people are primarily selfish, but they interact in an environment where reciprocation or signaling is rewarded. Rather than modeling the whole game explicitly, we just add g_i to the utility function. This is the argument Harbaugh (1998a) makes for including the "prestige motive" for giving, by which a person's well-being increases with the amount that other people believe he has given to charity, in utility, rather than modeling it as a way of increasing income. Other explanations for the warm-glow motive are discussed in the neuroscience section below.

In which model are donations less selfish? It's not obvious. Consider as extremes person 1 with Pure Altruism, $u_1(x_1, G)$, and person 2 with warm-glow altruism, $u_2(x_2, g_2)$. Both donate only because doing so increases their utility, and in that sense both types are selfish. Now consider a given donation – say, $100 to fund a bridge. As Andreoni (1989) has shown, in a reasonably large economy a purely altruistic person would only give if he had a very strong preference for the good. So, this donation seems selfish in the ordinary sense of the word. Now consider person 2, the warm-glow giver. Again, he is only giving because he likes the feeling of giving. It seems difficult to establish that there is any difference in the degree of altruism the two people exhibit on the basis of this $100 gift.

But now suppose that the public good G is the well-being of the poor. A $100 donation from the Pure Altruism donor 1 will only occur if he has a very strong preference for the welfare of the poor – strong enough to override the tendency to free-ride. Such a donation certainly should be called selfless. And what about the warm-glow donor 2, who gives the same $100 to aid the poor? This donation is motivated not by the donor's concern for the poor, but by his concern that he himself gives (or is seen to be giving) to improve the welfare of the poor. It seems entirely reasonable to argue that a purely altruistic donor who gives is "more altruistic" than a warm-glow giver who gives the same amount. ImPure Altruism would be somewhere in between. In short, while it seems difficult to label a person as more or less selfless on the basis of whether his preferences are defined by the Pure Altruism or the warm-glow model, it does seem reasonable to say that a given gift is more altruistic when it comes from a donor who is motivated by Pure Altruism. Ribar and Wilhelm (2002) address a related issue in the context of the Impure Altruism model and show that, even if people have a strong pure altruism motive, as societies get larger the donations we do observe will almost always be attributable to the warm-glow motive.

While at first glance it might seem unlikely that models based on utility maximization can be consistent with giving money away, we have shown that actually such a result is quite straightforward. All the above models make the standard economic assumption that people have well-defined preferences, and that they choose purposively to maximize their well-being given those preferences. We have shown that these assumptions lead to testable predictions, and that the results of empirical tests have generally supported the imPure Altruism model. However, all these tests leave substantial amounts of unexplained variance. In the next section, we will consider the counterargument to rational choice. Perhaps choices about when to give, and how much, are dominated by transient emotions and easily manipulable frames, or are so ingrained by morals or religion that changes in economic constraints have little effect.

NEURAL-LEVEL EVIDENCE

Neuroeconomic studies of altruism and charitable giving are just beginning. Here, we review the small set of works that deal directly with altruistic giving, and discuss a broader range of studies that have at least an indirect bearing on how experiments that produce neural data can help to answer the question, why do people give? Rather than reviewing the literature study by study, this section is organized in terms of what is known, and what neuroeconomics may be able to discover, regarding key questions. The focus is on the broad topics of whether or not giving decisions are rational, and what motivates giving.

Are Giving Decisions Really Rational Choices?

A basic assumption of all economic models of altruistic behavior is that giving decisions are rational responses to people's preferences and constraints. An alternative view is that while economic models may readily account for self-interested behavior, they cannot adequately explain behavior that is constrained by moral codes or that involves strong emotional responses (Greene *et al*., 2001). However, we believe that the rational-choice model has held up well in the neuroeconomic tests that have been conducted so far.

First, we note that there is substantial evidence, collected from animals over decades, that specific brain areas – the ventral striatum and the nucleus accumbens, and (with some qualifications) the insula and the orbitofrontal cortex – respond to the fulfillment of basic needs such as food, shelter, and social contact. Functional imaging work has demonstrated that, in humans, these areas also respond to abstract rewards such as money (O'Doherty, 2004; Knutson and Cooper, 2005). New work is now beginning to extend these results to rewards in settings involving social exchange. King-Casas and colleagues (2005) studied the development of trust during reciprocal economic exchanges. They found that activity in the head of the caudate increased in response to benevolent transactions received from a partner. Because of animal and human work implicating the caudate in reward-based learning, the authors interpreted this caudate response as a neural reward signal. Later in the sequence of exchanges, enhanced caudate activity began to appear prior to transactions with that same partner. The authors interpreted this as an indication of an intention to commit a benevolent transaction. In other words, the caudate reward signal indicated a decision to give away money when there was an expectation of reward by a trusted opponent player. While this situation does involve gifts of money, it involves reciprocity, where the returns are private and do not have any public good component. As we argued earlier, optimal behavior in this situation could easily involve "giving" even by an entirely selfish person. Nevertheless, this study was among the first to show that neural activity in reward-processing mid-brain areas correlates with economic exchanges involving social behavior.

More direct evidence regarding the argument that decisions in charitable giving situations can be modeled as rational behavior comes from a study by Moll *et al*. (2006). These authors used fMRI to look at brain activity during decisions to give money to a range of different charities. They found increased activity in the ventral striatum both when subjects received money for themselves and when they chose to donate money, suggesting that overlapping areas are involved in computing the utility of receiving money and the utility involved with giving. Further, activity during costly donations in the ventral striatum and a nearby medial frontal area correlated with how much subjects actually gave. This result suggests that the same areas involved in computing the utility associated with receiving or losing money also register the utility resulting from donating money to help provide a public good. While this study does not provide evidence to distinguish between the warm-glow and Pure Altruism models of giving behavior, it does provide compelling evidence that giving decisions use some of the same basic preference computation mechanisms that are also used when no regard for others is implied.

A recent study by Harbaugh *et al*. (2007) looks further into these issues. These authors used a protocol that included a mandatory, tax-like condition in which subjects simply observed more or less costly transactions from their own account to the charity. They also included trials in which the charity received money at no cost to the subject, and trials in which the subject received money at no cost to the charity. They found increased activity in the head of the caudate and the nucleus accumbens when subjects received money for themselves, replicating the results discussed above. A novel finding, however, was that there was also increased activity in these regions when the subjects simply observed the charity receiving money. As further evidence of the importance of these regions for charitable giving decisions, Harbaugh and colleagues showed that the degree of activation in the mandatory condition can be used to predict voluntary giving. The logic for this follows directly from the Pure Altruism model of giving, where $u = u(x, G)$, and is illustrated in Figure 20.2. Here, x denotes money going to the individual subject, and G denotes money going to the public good. The lines show representative *indifference curves*, or combinations of x and G that yield the same utility. The slope of the curves is a measure of a person's willingness to give up one good for the other, which in turn is determined by the relative *marginal utility*, or relative contributions to utility from increased amounts of the goods. A selfish person, with a low (but still positive) marginal utility of G, would only give up a small amount of x in return for a given increase in G, while a more altruistic person would give up a larger amount of x for the same increase in G, giving him steeper indifference curves. Movement towards the upper right – increasing consumption

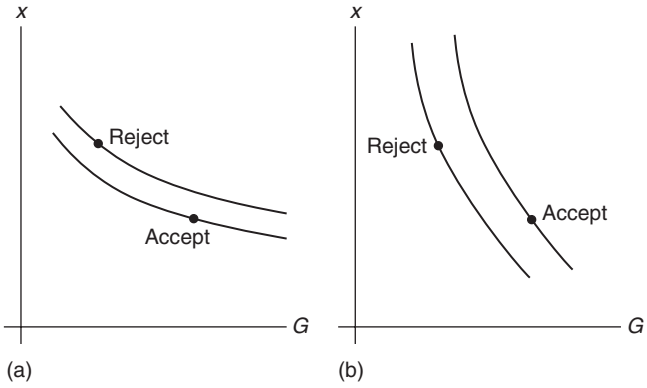

FIGURE 20.2 Altruistic preferences and decisions to accept or reject a charitable transfer: (a) egoistic; (b) altruistic.

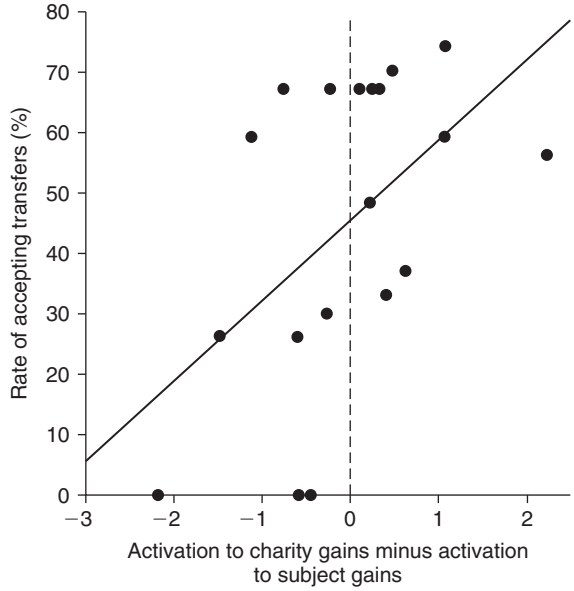

FIGURE 20.3 Neural activity in the mandatory condition predicts voluntary giving. Each point represents one subject. Negative amounts on the horizontal axis indicate that activation in the ventral striatum and insula is larger in response to monetary gains for the subject (x in Figure 20.2) than to gains for the charity (G in Figure 20.2). People on the right or positive side of the dotted line are more "altruistic," and have steeper indifference curves. The vertical axis shows the proportion of costly transfers that the subject accepted. The neural activation data support the Pure Altruism model illustrated in Figure 20.2.

of both goods – would put people on an indifference curve representing a higher level of utility. As Figure 20.2 shows, egoists get higher utility by rejecting transfers that reduce their x and increase G (i.e., they have lower marginal utility for G), while altruists maximize their utility by accepting them (i.e., they have higher marginal utility for G).

It emerges that the predictions of this model are supported by the combination of neural and behavioral data from the Harbaugh *et al*. experiment. The authors used increases in neural activity in the ventral striatum and insula during pure mandatory payments to the subject as an indicator of the subjective marginal utility of money (the private good x), and activity in pure mandatory payments to the charity as an indicator of the marginal utility associated with increases in G. In other words, the magnitude of activation in the ventral striatum in response to receiving money for oneself was taken as the measure of the person's marginal utility for the private good, and the magnitude of activation in response to seeing increases in the money available to the charity was taken as the measure of their marginal utility for G. The Pure Altruism model holds that the difference between these measures predicts people's willingness to give. As Figure 20.3 shows, people with large neural responses to G relative to their responses to x were particularly willing to give away money to the charity in the voluntary parts of the experiment.

Both the Moll *et al*. and the Harbaugh *et al*. studies support the argument that the neural basis of charitable giving decisions is consistent with a rational choice model where people make their giving decisions by comparing the utility they get from spending money on themselves with the utility they get from seeing the charity have more resources to devote to the public good. An open question, however, is whether Pure Altruism or warm-glow provides a better model of the utility of charitable giving. We turn to this question about the motives of giving in the following section.

However, it is also important to note that the existing evidence in favor of utility computations during giving decisions tells us little about the potential limits of the rational choice model. For example, it is at least conceivable that, in a situation for which strong moral codes exist, utility computations can be replaced by a purely rule-based decision process (e.g., "Always give away 10% of your income"). Adherence to such a rule might lead to behavior that is inconsistent with the utility computations for a particular situation. Alternatively, such moral codes might simply feed into the utility computations – for example, by linking rule-incongruent options with a high negative utility. Neural data that trace activity in areas related to goal and rule representations should help to uncover the potential interplay between rule representations and utility computations. More generally, altruistic decisions may be subject to framing effects, or contextual influences on decisions. For example, Andreoni (1995) provides experimental evidence that people end up allocating more money to a public good when their

decision is described as taking money from the group than when it is described as giving. Such differences could be interpreted as irrational, or just as evidence that people rationally use "rules of thumb" to reduce the cost of their choices.

Although neuroscience has not yet addressed the influence of framing effects on charitable giving, several studies have begun to delineate the brain systems responsible for framing effects in moral and social judgment. Greene *et al*. (2001) asked subjects to judge the appropriateness of an action in various moral dilemmas. For example, in one dilemma a trolley is running down a track on which it will kill five people, and subjects must decide whether or not to pull a switch which will re-route the trolley, saving those five people but killing a single person on the other track. Subjects generally respond that it is appropriate to push the button and save the five people at the cost of the one. However, in a slightly different dilemma, subjects must decide whether it is appropriate to push a large person onto the track in order to stop the trolley and save the five people. Despite killing and saving the same number of people in each of the two dilemmas, subjects are significantly less likely to say that pushing the person onto the track is appropriate behavior. Greene and colleagues proposed that the difference in the two scenarios is that the pushing scenario activates emotional systems that are not active in the switch scenario, and that these emotional differences alter people's judgments. The authors found neural evidence to support this theory, insofar as brain regions associated with emotional processing, including posterior cingulate cortex, medial frontal lobes, and posterior parietal lobes, were more active during "personal" moral dilemmas – dilemmas involving emotionally intense behaviors such as direct personal harm. Further evidence for the importance of emotions as determinants of the sorts of social decisions that are involved in charitable giving comes from Koenigs *et al*. (2007), who found that lesions to the orbitofrontal cortex – a region associated with processing of emotions – were associated with an increased tendency to use "rational" utilitarian judgments in decisions involving tradeoffs between one person's life and another's.

Another study, by Delgado *et al*. (2005), suggests a similar framing effect in social judgments. Subjects read biographies of "good," "bad," and "neutral" fictional characters, and then played a trust game with the three characters. Despite the fact that all three characters cooperated at the same level, subjects throughout the experiment continued to trust the "good" partner and distrust the "bad" partner. Further, brain regions involved in reward-learning did not respond differentially to cooperative and defective behaviors of these "good" and "bad" partners, suggesting that prior social information reduces neural responsiveness to good and bad outcomes. Because charitable giving is likely to involve many of the same computations and decision-making mechanisms that support social and moral judgments, charitable behavior is likely to be subject to similar types of emotional and learning biases. We note, however, that not all these biases are irrational – in fact, in the context of repeated interactions discussed in the economics section, many of these biases are likely to represent optimal responses.

Can Neural Evidence Reveal the Motives for Giving?

If we agree that decisions to make charitable donations are at least in part rational, then we can consider the correct way to model the behavioral and neural mechanisms influencing such choices. As explained earlier, economic models have distinguished between two broad classes of motives, one referred to as Pure Altruism and the other as warm-glow, with Impure Altruism models including both. In the context of contributions to charities that provide benefits to the poor, Pure Altruism implies that the donor's utility of giving is directly linked to the utility of the poor. In contrast, a cynical view would be that, under the warm-glow motive, the recipient's well-being is a means to benefit the donor, by signaling wealth, showing off his character as a good person to others, or reducing guilt and making it easier to think about himself as a moral person. A key distinction between these two groups of motives is that a person only gets warm-glow benefits if he makes an active decision to give, while a purely altruistic motive should be satisfied even by passively observing an increase in the public good which is paid for by someone else.

The fMRI study by Harbaugh *et al*. (2007), cited above, contrasted exactly these two situations. In the mandatory trials, the transfers from the subject to the charity were done in a passive, tax-like manner. In the voluntary trials, the subject could make deliberate giving decisions. The fact that reward-center activity occurred even when the charity received money in the form of a mandatory tax provides the first neural evidence showing that exogenous changes in the amount of money a charity receives have a neural effect. However, this still leaves the question: to what extent can this neural evidence be considered a motive that influences actual giving decisions? As discussed above, Harbaugh and colleagues showed that the degree to which reward-center activity

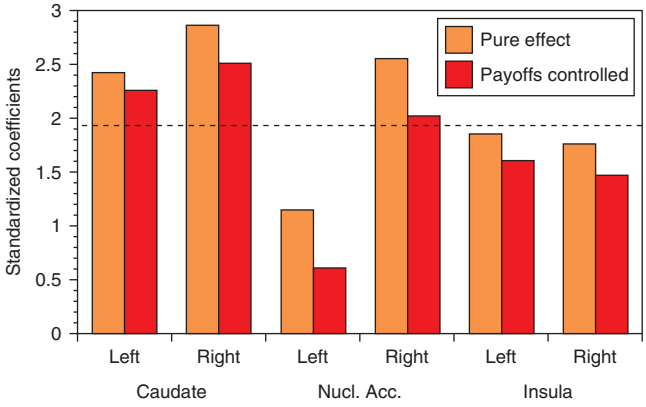

FIGURE 20.4 Higher activation when transfers are voluntary. The figure shows that activation in the indicated areas is generally higher in voluntary than in mandatory transfers.

responded to the pure, mandatory gains by the subject and the charity predicted the participant's "out of treatment" willingness to give in a voluntary manner. This suggests that a Pure Altruism motive exists and influences giving decisions (see also Batson, 1988).

Of course, this pattern of results does not rule out an additional warm-glow motive. As discussed in the economics section, the observed facts of charitable giving behavior show that there must be something more than Pure Altruism behind charity, such as a warm-glow motive. The design in the Harbaugh *et al*. (2007) study provides a way of distinguishing between the two motives, by comparing activity in areas that respond to reward – specifically, the nucleus accumbens, caudate, and insula – in mandatory and voluntary giving situations. As Figure 20.4 shows, activity in these areas was higher in the voluntary treatments. Importantly, the extra neural activation from voluntary giving was still present after controlling for the actual payoffs (i.e., the fact that in voluntary transfers individuals could keep more money for themselves than in mandatory transfers). The fact that the same monetary transfers are more rewarding when subjects can take responsibility for them shows that giving is not only motivated by concern for the recipient, as in Pure Altruism, but also by a person's desire to feel responsible for helping the recipient. This second motive is warm-glow altruism (or, as discussed in the economics section, a more complicated version of constraints, which can be reduced to warm-glow altruism). While the warm-glow model does not specify exactly how a person's voluntary giving enters utility, perhaps the most reasonable argument would be that utility increases with the amount of the gift. Harbaugh and colleagues (unpublished data) were only able to identify an increase in activation during voluntary giving. They did not show that activation increased with the amount given, and this remains a subject for future work.

While the Moll *et al*. and Harbaugh *et al*. studies produced evidence for the existence of Pure Altruism and warm-glow motives, it is clear that additional work regarding the neural basis of warm-glow giving may answer some intriguing questions. For example, one issue involves to what extent the higher activation during voluntary giving is tied to decisions about giving, or whether it is a specific example of the general "free to choose" principle, that more choices make people better off. Another issue concerns the strength of the evidence. While the Harbaugh *et al*. (2007) study found support for the warm-glow effect by demonstrating higher neural activity in the voluntary condition, a more rigorous test would be to demonstrate that activity increases with the payoff to the charity at a faster rate for voluntary than for mandatory transfers. Harbaugh and colleagues could not establish this pattern in a statistically reliable manner in their neural data, although this is perhaps not so surprising, given the small number of subjects. However, another possibility is that the contribution of warm-glow to utility is not always strictly increasing – perhaps people feel that a small contribution is enough for them to think that they have "done their bit."

It also might be the case that, in the Harbaugh *et al.* study, the warm-glow effect was reduced by the minimal nature of the manipulation. In particular, all transfers were anonymous. While even anonymous giving decisions provide some opportunity for the warm-glow feeling (i.e., one can feel good about oneself), outside the lab charitable giving and philanthropy are anything but anonymous – United Way contributions are recorded by employers, supporters of the Opera get their name or even their picture in the program, and people who make large enough contributions to a university can contract to have a building (or at least a brick) named after them.

This raises the point that neither Pure Altruism nor warm-glow are simple one-dimensional motives, and that much research can be done to unpack their constituent components. We believe that neuroeconomic experiments are particularly well suited to this, as is discussed below.

Neural Evidence Regarding the Role of Empathy and Emotion

Evidence from social psychology suggests that when people are in more empathetic, sympathetic, or positive states, they are more likely to give money or volunteer time to charities or needy people.

Batson and colleagues have shown in a number of experiments that inducing empathy in subjects, for example by instructing them to think about how another person feels, increases the rate at which subjects are willing to help another person (Batson *et al*., 1988; Batson, 1991).

The role of empathetic states in giving behavior offers a potential explanation for the identifiable victim effect, in which altruistic behavior is strongly influenced by whether or not givers are exposed to an individual recipient in need (such as a single hungry child). Slovic (2007) shows that presenting multiple victims (starting with two!) usually reduces the altruistic response to a humanitarian disaster. One possible interpretation of these results is that the system that tracks others' bodily or mental states has a severe capacity limitation (Morewedge and Schooler, 2008). Researchers such as Slovic argue that stimuli with more tangible properties, such as photos of a suffering individual, are more capable of engaging emotional systems and thereby evoking the affective states that lead to altruistic behavior than are abstract stimuli such as death-rate statistics. The finding that the presentation of a single face significantly increases donations supports this line of reasoning, since faces are known to be highly arousing and emotional stimuli (Adolphs, 2002).

To the extent that altruistic behavior requires the ability to recognize the needs of other people, we should expect altruism to depend upon the functional integrity of brain systems required for such social perception capacities. Recent neural work by Singer *et al*. (2004; see also Chapter 17 of this volume for additional details) and others has mapped out the neural circuitry associated with empathic responses to observing physical harm done to another. They find that the anterior insula seems to function as a representational system that traces one's own and others' bodily states. Interestingly, neural activity in this region while observing somebody else receiving physical harm showed a considerable correlation with a self-report empathy scale. More cognitive forms of empathy in which subjects are instructed to think about the feelings of another person have implicated a prefrontal-parietal set of brain regions that is thought to be involved in understanding the intentions and predicting the behaviors of other agents (Ruby and Decety, 2004).

Despite this, so far there is no direct neural evidence linking empathy to altruistic behavior. In fact, Harbaugh *et al*. (2007) found no correlation between psychometric empathy scales and either behavioral or neural manifestations of altruistic tendencies (unpublished data), and Tankersley *et al*. (2007) found no correlation between empathy measures and the neural responses that predicted altruistic tendencies. However, aside from the fact that such null-findings are not very conclusive, it is also important to note that the giving situation implemented in these studies was relatively abstract, and did not involve identifiable recipients with specific characteristics or needs.

The Tankersley *et al*. study does suggest, however, that the neural systems that are involved in social perception might serve as precursor functions to empathy. In their experiments, Tankersley and colleagues measured brain activation while subjects either played a simple game or watched as a computer played the game. After scanning, subjects completed a self-report altruism survey. The authors found that the posterior cingulate cortex and bilateral posterior parietal lobes were more active when subjects watched the computer play than when subjects played the game themselves. Interestingly, activity in the right superior temporal cortex was correlated with altruism. More altruistic subjects (as measured by self-reports about behavior outside the lab) showed higher levels of activity in this region, in comparison with less altruistic subjects. Although they did not find significant correlations between measures of empathy and activity in the right superior temporal cortex, this region is thought to be involved in a variety of social perception tasks, including the detection of agency and goals. Indeed, Blair (2005) has proposed a model in which this region supports early analysis of actions, which are in turn transmitted to other systems that facilitate empathic responding.

The Tankersley *et al*. study shows the contribution that neuroscience can make towards understanding the components of altruistic motivations. In their paradigm, right superior temporal cortex activated in response to the flash of a target, indicating that the computer had responded for that trial. The fact that this region responded more to such a simple stimulus for altruists than others suggests that fairly low-level attentional mechanisms are influencing subjects' propensity to engage in helping behavior. Because neuroscientific techniques offer access to motivational mechanisms of which subjects are unaware, they are uniquely suited to reveal findings that individual differences in visual attention influence altruistic tendencies.

In addition to empathy and its perceptual precursors, social psychologists have also shown that simply inducing positive affect – for example, by arranging an experimental setting where the subject "finds" a dime in the phone booth (Levin and Isen, 1975) – increases helping behavior (mailing a stamped addressed letter "accidentally" left at the phone booth). The authors

speculated that the increased willingness to help was an effort to maintain the good mood induced by finding the dime. Although this inference might seem far-fetched, a wealth of psychobiological data supports a connection between prosocial behavior and a variety of positive affective states or dispositions.

Depue and Morrone-Strupinsky (2005) use the term *affiliation* to describe an individual's motivation towards social stimuli, as well as their capacity to enjoy social stimuli. Affiliation is a complex trait related to other traits such as extraversion and sympathy, which are in turn predictive of prosocial behaviors, including altruistic behavior. Depue and Morrone-Strupinsky argue that the dopaminergic system, including the substantia nigra and its subcortical (basal ganglia) and cortical (medial prefrontal cortex) projections, influences an individual's sensitivity and motivation towards social rewards such as friendly vocalizations or facial expressions, as well as a wide range of positive affective traits, from desire to self-efficacy. They argue that, in this case, a second opioid-based system is critically involved in experiencing the pleasure of these social stimuli, and that the dopaminergic and opioidergic systems interact to produce learning and memory of social stimuli and their rewarding values. Further, research suggests that genetic differences significantly influence the functioning of the opioid and dopamine systems, and these genetic differences have been related to significant differences in motivation, including prosocial motivation, in animals (Depue and Collins, 1999).

Recent work on the effects of the neuropeptide oxytocin actually provides some indirect evidence that empathy leads to altruistic behavior (Zak *et al*., 2007). Participants who received either oxytocin or a placebo played both a dictator game (i.e., where the player can split money with an anonymous recipient with no opportunity for reciprocal punishment) and an ultimatum game (i.e., where the recipient can punish the giver for unfair sharing). Subjects who had received oxytocin made substantially large offers, but only in the ultimatum game and not in the dictator game. The authors argue that only the ultimatum game requires emphatic appraisal of the others' mental states. They therefore conclude that oxytocin, which is known to promote bonding and affiliative tendencies, boosts altruistic behavior by increasing empathy, but not by affecting altruism *per se*. The somewhat puzzling aspect is that in this case greater altruism after receiving oxytocin is observed under conditions of strict reciprocity – not a situation that is typically regarded as a good test of Pure Altruism. Thus, it would be interesting to replicate these results with other ways to manipulate demands on empathy. In this context, another recent study (Knafo *et al*., 2007) is of interest; this looked at people who differed in the promoter region of a gene that is involved in the receptor for vasopressin, which chemically is very similar to oxytocin (AVPR1a). In a one-shot dictator game, individuals with long versions of the relevant gene gave substantially more money to the anonymous other player than did those individuals with the short version. Thus there is converging evidence about the relationship between biological mechanisms of social bonding and altruistic behavior in economic games, including initial findings that point to important genetic factors. However, the exact pathway and conditions under which such effects become apparent needs further study.

Aside from the propensity for empathy, another trait that is known to increase altruistic behavior is the complex trait of "resilience," which involves (among other things) the ability to respond in an effective manner to stressful environments, and to adapt to new social situations (Charney, 2004). According to Charney, several of the brain's neurochemical systems, including the dopamine system, influence altruistic behavior by regulating an individual's sensitivity to positive and aversive stimuli in the environment, which in turn influence the individual's affective and motivational state. Thus, there seems to be a cyclical effect in which helping behavior induces positive affect and resilience, which in turn increases the propensity to help.

We know from economic data and experiments that there are large individual differences in charitable giving. Only part of the variation can be explained by differences in demographic variables such as gender, income, or education. Rather, it seems that a substantial part of the remaining variance can be explained by potentially stable individual differences in aspects such as the propensity to show empathy, other-regarding emotions, or hormonal factors. Already, there is evidence that differences between people in altruistic behavior go along with interesting neural-level differences. Future work of this kind should reveal the neural and psychological basis of the individual characteristics that are critical during public good decisions.

Fairness as a Motive for Giving

The Pure Altruism motive in particular might be expected to respond to people's appreciation of the needs of other specific individuals or groups. There is evidence from psychological research that many people entertain a "just world" belief (see, for example, Lerner, 1980; Hafer and Begue, 2005). The experience of innocent victims endangers this belief, and can lead

to altruistic behavior in order to reconstitute the "just world." Economists have proposed so-called altruistic punishment as a way of discouraging defectors and free-riders in the context of cooperative public good situations. In fact, laboratory studies have shown that people are often willing to pay a cost in order to punish unfairness – and indeed to reward fairness (Andreoni *et al*., 2003; see chapter by E. Fehr in this handbook). The fact that this occurs even in non-reciprocal, anonymous situations suggests that punishment is driven by the desire for fairness itself, not by the prospect of later positive consequences for the punisher. In fact, neural evidence suggests that altruistic punishment activates very similar reward activity to that found in actual giving situations, and the strength of this activation correlates with the actual amount invested in order to punish defectors (de Quervain *et al*., 2004).

Empirically, it may be difficult to distinguish between Pure Altruism that is driven by the perception that somebody's needs are fulfilled rather than by the abstract goal of establishing fairness across individuals. However, one interesting question is, to what degree is the willingness to engage in altruistic punishment correlated with the neural Pure Altruism Indicator established by Harbaugh *et al*.? Do Pure Altruism and altruistic punishment go together, or are there alternative ways to promote the provision of public goods? Maybe some people deriving utility from acting altruistically while others derive from punishing those who don't?

Goals and Altruistic Decisions

Computations of utility related to self-interest and other-interest need to integrate information from a variety of sources. One possibly important source is an individual's abstract goals and values. There is indirect neural-level evidence that goal representations can play an important role during altruistic decisions. Moll and colleagues (2006) found that activity in a small, anterior medial frontal region, in situations that pitted self-interest against other-interest, was highly predictive of self-reported engagement in charitable giving. According to the authors' interpretation, this prefrontal area represents and maintains high-level goals. By this logic, only individuals who can establish firm goal representations can withstand the pull towards the self-serving option of keeping money for oneself (see also related work by Knoch *et al*., 2006).

An interesting point raised by this result regards the extent to which goal representations in the prefrontal cortex actually influence decisions, by feeding information into the utility computation. In this case, we should see a tight linkage between prefrontal goal activity and Pure Altruism reward-center activity. The alternative possibility is that representations of rules, goals, or moral codes represented in the prefrontal areas allow for direct control of behavior, bypassing the utility-based decision-making system. From a neuroeconomic perspective, this brings us to the fundamental question regarding the potential limits of the utility-based rational-choice model. In principle, it should be possible to address this question by creating situations in which the relative contributions of prefrontal areas and reward-processing areas in the midbrain to actual altruistic decisions can be determined. If prefrontal goal representations influence behavior by bypassing utility computations, then we should see activity in midbrain areas involved in reward processing and prefrontal activity as independent predictors, both for decisions across individuals and for variability across decisions within individuals.

Warm-glow, Private Versus Public Giving, and the Effect of Religion

Recent economic work on voluntary giving decisions has focused on establishing facets of the warm-glow motive (see above). This work has demonstrated that the prospect of signaling "good character" through one's charitable contributions is an important motivator of seemingly altruistic behavior, as in Harbaugh (1998a, 1998b) and Ariely *et al*. (2007). Such motives should be particularly aroused when the charitable behavior is public. Interestingly, recent psychological work provides at least indirect evidence that it takes only very little to create a "public" situation. Bateson *et al*. (2006) showed that a poster depicting a pair of eyes led to an almost three-fold increase in prosocial behavior (payments into a shared coffee fund) compared with a control situation involving a picture of flowers. While the cause is rather trivial, this is a dramatic effect.

So far, there is no neural-level work that attempts to separate the different aspects of warm-glow. If we go by the working hypothesis that midbrain reward areas are generally involved in utility computations, we would not necessarily expect that warm-glow and Pure Altruism can be distinguished within these areas. After all, both motives can be captured within a utility-based rational-choice model. In fact, Harbaugh *et al*. (2007) found that mandatory transfers to the charity (i.e., Pure Altruism) produced midbrain reward-center activity, and that voluntary transfers (i.e., warm-glow giving) produced additional activity in the same areas.

However, warm-glow giving might very well express itself in a distinctive manner in other neural areas. Most, but not necessarily all, aspects of warm-glow giving involve the consideration of an audience and repeated interactions. Specifically, prestige and signaling effects require that the potential donors have a mental model of the observers that represents what these observers might classify as financially impressive or morally admirable actions. Therefore, one might expect that a signature of warm-glow will be activity in "social cognition areas" such as the anterior rostal, medial prefrontal cortex (including paracingulate gyrus), which has been associated with the consideration of others' mental states (see, for example, Amodio and Frith 2006). So it is interesting to consider to what degree private and public giving lead to differing activity in these prefrontal areas. Complicating matters empirically, similar areas have been implied in the representation of one's own value judgments (e.g., Zysset *et al*. 2002). As a theoretical complication, it is also worth considering the so-called "objective self-awareness" theory (Duval and Wicklund, 1972). Supported by a host of empirical evidence (Silvia and Duval, 2001), this theory states that cues that focus people's attention on themselves (e.g., mirrors, cameras, or onlookers) also bring their own standards into the foreground. From a theoretical perspective, it is a significant difference whether somebody shows an increase in prosocial behavior in order to impress others or because the presence of others brings one's own standards into the foreground. It is too early to tell whether neural evidence can clearly distinguish between the process of activating one's own goals or standards, and the process of thinking about somebody else's mental states. Clearly, such evidence could be very useful to pinpoint the various ways in which private and public altruistic decisions might differ.

One of the most interesting puzzles in charitable giving is the effect of religiosity. In the US, nearly 60% of all giving goes to religious organizations (data from the Independent Sector, analyzed in Andreoni, 2006). In a review of the empirical literature, Bekker and Wiepking (2007) report that "Positive relations of church membership and/or the frequency of church attendance with both secular and religious philanthropy appear in almost any article in which this relation was studied." But they then note that experimental studies, e.g., Eckel and Grossman (2004), do not support this effect for giving to secular charities. Evolutionary theorists have speculated that the promotion of cooperative, prosocial behavior may be the most critical advantage brought by the evolution of a mental capacity for religious thought (Sosis and Alcorta, 2003; Wilson, 2002). Maybe the most straightforward explanation of this phenomenon is that religions typically establish strong standards and goals for giving, as in calls for tithing (Lansdell, 1906). The result in Moll *et al*. (2006) regarding the correlation between goal-area activity and actual giving might be an interesting starting point for looking into this possibility. However, what is so far largely missing is experimental evidence that allows us to establish (1) that religious beliefs and practices actually change constraints and preferences, rather than people with certain preferences selecting into religion, and (2) how religion might change preferences. The few experimental studies that do exist suggest that the evocation of a higher power leads to more prosocial behavior (for example, the mention of the ghost of a dead graduate student haunting the lab increases prosocial behavior; Bering *et al*., 2005). One possible explanation of such results is that a higher power functions as an all-seeing observer. We have already seen that even subtle cues of being observed can have powerful effects on prosocial behavior (Bateson *et al*., 2006). In a certain sense, decisions by religious individuals are always "observed," and so charitable decisions might always involve a larger warm-glow benefit. This hypothesis would suggest that religious individuals, or individuals in which a "god concept" was induced experimentally (Norenzayan and Shariff 2007), might consider how their own decisions are perceived by the "higher power." This leads to the prediction that religious individuals in private-giving situations and non-religious individuals in public giving situations should exhibit similar activity in mentalizing areas (Amodio and Frith 2006).

The Neural Basis of Life-span Differences in Altruistic Behavior

Individuals differ in their altruistic tendencies and, as discussed in the previous sections, the few existing brain-imaging studies have produced interesting insights into the nature of these differences (Moll *et al*., 2006; Harbaugh *et al*., 2007; Tankersley *et al*., 2007). This raises the hope that neural evidence can also be used to explain one of the most powerful predictors of individual's altruistic tendencies – namely, that the percentage of income devoted to charitable giving increases relatively steadily across the life-span (Andreoni, 2006). Since older people are typically wealthier people, this means that the majority of the donations that charitable organizations receive come from older donors.

So far, there is no explanation for this intriguing life-span trend. In fact, based on what we know about

(1) the negative effects of aging on frontal cortex (see, for example, Raz, 2005) and (2) the involvement of prefrontal areas in promoting behavior that goes against immediate self-interest (e.g, Moll *et al*., 2006), it is a puzzle why altruistic tendencies seem to increase rather than decrease. In this context, it would be very revealing to examine prefrontal contributions during altruistic decisions across the life-span. It is very possible that there is actually a weakening of prefrontal influences that would lead to a reduction of altruistic behavior if it were not for additional factors.

What might such additional factors be? Interestingly, there is evidence that, as people's time horizon shrinks, they experience a gradual shift in general motivational orientation, away from information-gathering and achievement and towards meaningful emotional experiences – in particular in the social domain (see, for example, Carstensen, 2006). One aspect of this shift seems to be less regard for potential losses, as indicated by a reduction in loss-related neural responses (Larkin *et al*., 2007). While so far this general perspective has not been applied to public good decisions, there does seem to be a natural connection. In the context of public good decisions, a greater tendency to engage in "meaningful interactions with the world" and a reduced appreciation of losses to oneself might lead to a greater Pure Altruism Response in reward areas.

More likely than not, the story of life-span effects on altruistic giving will be a complicated one. It will probably involve both prefrontal and reward-area contributions that represent the negative cognitive and the more positive socio-emotional effects of aging. Neuroimaging evidence should be useful in disentangling these counteracting forces.

CONCLUSION

We started this chapter by explaining the primary economic models of charitable giving: Pure Altruism and warm-glow altruism. We showed that the bulk of the evidence from empirical and experimental studies supported the existence of a purely altruistic motive for charitable giving, but that in societies of more than 20 or so people this motive lost its force, and could not explain the widespread giving that we observe. We explained the empirical, experimental, and neuroscience studies showing that the combination of Pure Altruism and warm-glow known as Impure Altruism was an appealing alternative model, and also why alternative explanations based on repeated play, reciprocity, and signaling might too be consistent with the data.

The above summary suggests a consensus which might indicate that the marginal benefit from further investigations is low. This would be a mistake. There is a plethora of interesting and important questions, and neuroeconomic techniques are well suited to investigating many of them. In particular, in the sections above we have discussed how neuroeconomic methods can be used to address questions about the determinants and components of the Pure Altruism and warm-glow motives.

There are additional questions to consider. While the focus of this chapter has been on charitable giving, the truth is that public goods are mostly funded by taxation. Taxation raises a set of issues that has not yet been addressed by neuroeconomics. Tax compliance is one example. Most economic studies show that tax cheating is far lower than would be expected if people simply weighed the benefits against the small chances of getting caught and the low subsequent penalties. Thus it seems likely that people pay taxes at least partly out of a sense of obligation, if not exactly altruism. Are the motives involved in tax compliance similar to or different from those in charitable giving and philanthropy? The fact that Harbaugh and colleagues found activity in reward-related areas such as the nucleus accumbens during tax-like transfers suggests some degree of similarity between motives involved when paying taxes and during charitable decisions. However, to what degree this generalizes to more realistic situations is an open issue that can be addressed with analytic techniques similar to those used by Harbaugh and colleagues.

A second and related set of questions involves voting for taxes. Assuming vigorous enforcement, paying taxes is not exactly altruistic. But what about voting to increase taxes that fund public goods? Taxes impose costs on the voters, and in that sense voting for a tax might sometimes be considered altruistic. However, taxation also imposes costs on others. In fact, it is the costs imposed on other people that will fund most of any increase in the public good. How do voters weigh these factors? When preferences on a given public good vary widely, so that some of those paying for a public good will not benefit from it (such as in the case of reductions in timber-cutting to save an endangered species), is the altruistic benefit of voting diluted?

Another interesting question is whether warm-glow benefits should be included in calculations of social welfare. If they are, we might argue that societies should rely more on philanthropy, and less on taxation. The logic would be that more voluntary giving – perhaps promoted by increased tax deductions – would make people better off both through the Pure Altruism preference for more public

goods, and through the warm-glow preference for active giving. This would seem supportable if warm-glow motives arise out of a basic preference for taking action to help provide public goods. It might also hold if the warm-glow motive originates with the need to signal one's trustworthy character to potential partners. But suppose that the warm-glow turns out to be primarily about status or showing off income? More giving might mean greater status for one person, but less for another. In this sort of a zero-sum game, it might actually make sense to discourage charitable contributions and rely instead on higher taxes.

Still other questions involve the interactions between the motives of givers and the goals of the non-profit-making organizations that actually spend the money to increase the public good. Fundraisers often say that "people give because we ask them to give." What effect does the "ask" have on the utility of the giver? Given the lengths that some people will go to in order to avoid having to say no, it seems unlikely that the process that leads to charitable giving always involves increases in the utility of those who contribute. It seems that neuroeconomic techniques are particularly well suited to addressing these sorts of questions.

Broadening the perspective a little, neural-level results such as the ones discussed in this chapter might provide important hints regarding how people develop the motivational tendencies that apply in interpersonal and public good exchanges. Particularly revealing is the finding that activity in midbrain areas can accompany altruistic behavior. After all, in terms of learning, the primary function of a reward is to strengthen behavioral patterns that lead to the reward and weaken those that do not. Altruistic behavior may well be subject to the same type of learning, as are other types of behavior. So, an important question is, to what degree is there an *a priori* link between altruistic behavior and reward, or is engagement in altruistic behavior more an acquired taste? For example, an interesting hypothesis is that a warm-glow related reward can serve as a primary learning reinforcer, which over time generalizes to provide pure altruistic rewards. Praising toddlers for sharing their toys might deliver the warm-glow benefit, which then promotes future altruistic behavior even when no one is looking.

References

Adolphs, R. (2002). Recognizing emotion from facial expressions: psychological and neurological mechanisms. *Behav. Cogn. Neurosci. Rev.* 1, 21.

Amodio, D.M. and Frith, C.D. (2006). Meeting of minds: the medial frontal cortex and social cognition. *Nat. Rev. Neurosci.* 7, 268–277.

Andreoni, J. (1989). Giving with impure altruism: applications to charity and Ricardian equivalence. *J. Political Econ.* 97, 1447–1458.

Andreoni, J. (1990). Impure altruism and donations to public good: a theory of warm-glow giving. *Economic J.* 100, 464–477.

Andreoni, J. (1995). Cooperation in public good experiments: kindness or confusion. *Am. Econ. Rev.* 85, 891–904.

Andreoni, J. (2006). Philanthropy. *Handbook of the Economics of Giving, Reciprocity and Altruism*. Amsterdam: Elsevier, pp. 1201–1269.

Andreoni, J., Harbaugh, W., and Vesterlund, L. (2003). The carrot or the stick: rewards, punishments, and cooperation. *Am. Econ. Rev.* 93, 893–902.

Andreoni, J., Harbaugh, W.T., and Vesterlund, L. (2008). Altruism in experiments. In: S.N. Durlauf and L.E. Blume (eds), *The New Palgrave Encyclopedia of Economics*, 2nd edn. Basingstoke: Palgrave Macmillan, (available online at http://www.dictionaryofeconomics.com/article?id=pde2008_A000240 doi:10.1057/9780230226203.0035)..

Ariely, D., Bracha, A., and Meier, S. (2007). Doing good or doing well? Image motivation and monetary incentives in behaving prosocially. IZA Discussion Paper No. 2968 (available online at SSRN: http://ssrn.com/abstract=1010620).

Axelrod, R.M. (1984). *The Evolution of Cooperation*. New York, NY: Basic Books.

Bateson, M., Nettle, D., and Roberts, G. (2006). Cues of being watched enhance cooperation in real-world setting. *Biol. Letts* 2, 412–414.

Batson, C.D. (1991). Evidence for altruism: toward a pluralism of prosocial motives. *Psychological Inquiry* 2, 107–122.

Batson, C.D., Dyck, J.L., Brandt, J.R. *et al*. (1988). Five studies testing two new egoistic alternatives to the empathy-altruism hypothesis. *J. Pers. Social Psychol.* 55, 52–77.

Becker, G.S. (1974). A theory of social interactions. *J. Political Econ.* 82, 1063–1093.

Bekkers, R. and Wiepking, P. (2007). Generosity and Philanthropy A Literature Review. 2007. Available online at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1015507

Bering, J.M., McLeod, K., and Shackelford, T.K. (2005). Reasoning about dead agents reveals possible adaptive trends. *Hum. Nature* 16, 360–381.

Bergstrom, T., Blume, L., and Varian, H. (1986). On the private provision of public goods. *J. Public Econ.* 29, 25–49.

Blair, R.J. (2005). Responding to the emotions of others: dissociating forms of empathy through the study of typical and psychiatric populations. *Conscious Cogn.* 14, 698–718.

Carstensen, L.L. (2006). The influence of a sense of time on human development. *Science* 30, 1913–1915.

Charney, D.S. (2004). Psychobiological mechanisms of resilience and vulnerability: implications for successful adaptation to extreme stress. *Am. J. Psych.* 161, 216.

de Quervain, D.J.F., Fischbacher, U., Treyer, V. *et al*. (2004). The neural basis of altruistic punishment. *Science* 305, 1254–1258.

Delgado, M.R., Frank, R.H., and Phelps, E.A. (2005). Perceptions of moral character modulate the neural systems of reward during the trust game. *Nat. Neurosci.* 8, 1611–1618.

Depue, R.A. and Collins, P.F. (1999). Neurobiology of the structure of personality: dopamine, facilitation of incentive motivation, and extraversion. *Behav. Brain Sci.* 22, 491–517.

Depue, R.A. and Morrone-Strupinsky, J.V. (2005). A neurobehavioral model of affiliative bonding: Implications for conceptualizing a human trait of affiliation. *Behav. Brain Sci.* 28, 313–350.

Duval, S. and Wicklund, R.A. (1972). *A Theory of Objective Self Awareness*. New York, NY: Academic Press.

Eckel C. , and Grossman P. (2004). Giving to Secular Causes by the Religious and Nonreligious: An Experimental Test of the

Responsiveness of Giving to Subsidies. *Nonprofit and Voluntary Sector Quarterly* 33: 271-89.

Glazer, A. and Konrad, K.A. (1996). A signaling explanation for charity. *Am. Econ. Rev.* 86, 1019–1028.

Goeree, J.K., Holt, C.A., and Laury, S.K. (2002). Private costs and public benefits: unraveling the effects of altruism and noisy behavior. *J. Public Econ.* 83, 255–276.

Greene, J.D., Sommerville, R.B., Nystrom, L.E. *et al*. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science* 14, 2105–2108.

Hafer, C.L. and Begue, L. (2005). Experimental research on just-world theory: problems, developments, and future challenges. *Psychological Bull.* 131, 128–167.

Harbaugh, W.T. (1998a). The prestige motive for making charitable transfers. *Am. Econ. Rev.* 88, 277–282.

Harbaugh, W.T. (1998b). What do donations buy? *J. Public Econ.* 67, 269–284.

Harbaugh, W.T., Mayr, U., and Burghart, D.R. (2007). Neural responses to taxation and voluntary giving reveal motives for charitable donations. *Science* 316, 1622.

Isaac, R.M. and Walker, J.M. (1988). Group size effects in public goods provision: the voluntary contributions mechanism. *Q. J. Economics* 103, 179–199.

King-Casas, B., Tomlin, D., Anen, C. *et al*. (2005). Getting to know you: reputation and trust in a two-person economic exchange. *Science* 308, 78–83.

Kingma, B.R. and McClelland, R. (1995). Public radio stations are really, really not public goods: charitable contributions and impure altruism. *Ann. Public Coop. Econ.* 66, 65–76.

Knafo, A., Israel, S., Darvasi, A. *et al*. (2007). Individual differences in allocation of funds in the dictator game associated with length of the arginine vasopressin 1a receptor RS3 promoter region and correlation between RS3 length and hippocampal mRNA. Available online at doi:10.1111/j.1601-183X.2007.00341.x.

Knoch, D., Pascual-Leone, A., Meyer, K. *et al*. (2006). Diminishing reciprocal fairness by disrupting the right prefrontal cortex. *Science* 314, 829.

Knutson, B. and Cooper, J.C. (2005). Functional magnetic resonance imaging of reward prediction. *Curr. Opin. Neurol.* 18, 411–417.

Koenigs, M., Young, L., Adolphs, R. *et al*. (2007). Damage to the prefrontal cortex increases utilitarian moral judgments. *Nature* 446, 908–911.

Kolm, S.C. (1969). *The Optimal Production of Justice. Public Economics*. London: Macmillan.

Lansdell, H. (1906). *The Sacred Tenth or Studies in Tithe-Giving Ancient and Modern*. Society for Promoting Christian Knowledge.

Larkin, G.R.S., Gibbs, S.E.B., Khanna, K. *et al*. (2007). Anticipation of monetary gain but not loss in healthy older adults. *Nat. Neurosci.* 10, 787–791.

Lerner, M. (1980). *The Belief in a Just World: A Fundamental Delusion*. New York, NY: Plenum Press.

Levin, P.F. and Isen, A.M. (1975). Further studies on the effect of feeling good on helping. *Sociometry* 38, 141–147.

Moll, J., Krueger, F., Zahn, R. *et al*. (2006). Human fronto-mesolimbic networks guide decisions about charitable donation. *PNAS* 103, 15,623.

Morewedge, C.K. and Schooler, J. (2008). Mind diffusion: when attribution of mind is greater to individuals than to group members. Working Paper, Carnegie Mellon University.

Norenzayan, A. and Shariff, A. (2007). God is watching you: priming God concepts increases prosocial behavior in an anonymous economic game. *Psychological Sci.* 18, 803–809.

O'Doherty, J.P. (2004). Reward representations and reward-related learning in the human brain: insights from neuroimaging. *Curr Opin Neurobiol* 14, 769–776.

Palfrey, T.R. and Prisbrey, J.E. (1997). Anomalous behavior in public goods experiments: how much and why. *Am. Econ. Rev.* 87, 829–846.

Raz, N. (2005). The aging brain observed in vivo: differential changes and their modifiers. In: R. Cabeza, L. Nyberg, and D. Park (eds), *Cognitive Neuroscience of Aging: Linking Cognitive and Cerebral Aging*. New York, NY: Oxford University Press, pp. 19–57.

Ribar, D.C. and Wilhelm, M.O. (2002). Altruistic and joy-of-giving motivations in charitable behavior. *J. Political Econ.* 110, 425–457.

Ruby, P. and Decety, J. (2004). How would you feel versus how do you think she would feel? A neuroimaging study of perspective-taking with social emotions. *J. Cogn. Neurosci.* 16, 988–999.

Samuelson, P.A. (1954). The pure theory of public expenditure. *Rev. Econ. Stats* 36, 387–389.

Silvia, P.J. and Duval, T.S. (2001). Objective self-awareness theory: recent progress and enduring problems. *Pers. Social Psychol. Rev.* 5, 230.

Singer, T., Seymour, B., O'Doherty, J. *et al*. (2004). Empathy for pain involves the affective but not sensory components of pain. *Science* 303, 1157–1162.

Slovic, P. (2007). If I look at the mass I will never act. Psychic numbing and genocide. *Judgment Decision-Making* 2, 79–95.

Sosis, R. and Alcorta, C. (2003). Signaling, solidarity, and the sacred: the evolution of religious behavior. *Evol. Anthropol.* 12, 264–274.

Tankersley, D., Stowe, C.J., and Huettel, S.A. (2007). Altruism is associated with an increased neural response to agency. *Nat. Neurosci.* 10, 150–151.

Wilson, D.S. (2002). *Darwin's Cathedral*. Chicago, IL: University of Chicago Press.

Zak, P.J., Stanton, A.A., and Ahmadi, S. (2007). Oxytocin increases generosity in humans. *PLoS ONE* 2(11), e1128, doi:10.1371/journal.pone.0001128.

Zysset, S., Huber, O., Ferstl, E., and von Cramon, D.Y. (2002). The anterior frontomedian cortex and evaluative judgment: an fMRI study. *NeuroImage* 15, 983–991.