Chapter 8
Lending a Hand

The Structure of Everyday Cooperation

In conclusion of the second part of this book, I shall come back to a topic that has been lurking in the background of all previous chapters of this section: the question of altruism. I have argued above that the concept of altruism is insufficient for capturing the motivational structure of cooperation (Chapter 5), coordination (Chapter 6) and commitment (Chapter 7). It is now time to address the question concerning the structure of altruism. Over the past decade, the concept of altruism has come to play an increasingly important role in social science. This is particularly true in experimental economics, where altruism is routinely quoted when it comes to explaining the vast discrepancies between the observed behavior in the experiments, and the predictions based on the standard economic model of human behavior. In this debate, as well as in some other contexts, altruism usually means having 'pro-social' or 'other-directed' preferences. It is observed that people are not always egoistic in terms of the somewhat narrow conception of self-interest that is still at work in much of economic theory. The conclusion that is usually drawn is to drop the assumption that individuals are only interested in what they can get for themselves in favor of a wider conception that extends to such preferences as benevolent desires, preferences for reciprocity and fair dealing, and the inclination to punish transgressions against the norms of fairness even if one is not directly affected by that transgression, and if punishment is costly to the punisher (cf., e.g., Fehr and Fischbacher 2003; Henrich et al. 2004).

Thus the term 'altruism' has become something of an indicator for what one might call the *defensive strategy* in recent economic theory. Proponents of this strategy acknowledge systematic deviations between the economic model and actual human behavior, but tend to believe that it is possible to correct this shortcoming simply by widening the class of human preferences. Not all participants in the debate, however, believe that such amendments to the model will do. Authors such as Amartya K. Sen have voiced serious doubts concerning this defensive strategy (Sen 1977, [1985] 2002; Peter and Schmid [eds] 2007). These authors favor what one might label the *critical strategy*, claiming that much more radical conceptual changes than a simple expansion of our view of human desires will be needed in order to come to an adequate understanding of human action, changes that affect not only our notion of motivation, but our concepts of the agent's identity and the nature of choice, too.

H.B. Schmid, *Plural Action: Essays in Philosophy and Social Science*, Contributions to Phenomenology 58,

Put very bluntly, the general thrust of this paper is to reclaim the term 'altruism' for the critical camp. The main claim is that limiting the extension of the term to such phenomena as benevolent desires or other-regarding preferences, as it is done within the defensive venture, means pulling the concept's teeth. An adequate theory of altruism has to go beyond pro-social and other-directed preferences, and necessitates far-reaching revisions in our outlook on human behavior.

The concept of altruism is approached from the perspective of action theory. The point of departure (in §26) is what I call the paradox of altruistic action, i.e. a conflict between the concept of action as used in much of current social science on the one hand, and our intuitive notion of altruism on the other. It seems that some of our pre-theoretic intuitions concerning the structure of altruism do not square with our standard theory of action. In the current literature, it is usually believed that this paradox is due to an overstrained notion of altruism, and that a more relaxed view of altruism can easily be accommodated in the standard theory of action via other-directed desires. The next section voices doubts concerning this solution to the paradox. A class of behavior is discussed which does indeed seem to be altruistic in the strong intuitive sense, and does not seem to fit other-directed desire explanations. The distinguishing feature of this class of altruistic behavior is the peculiar way in which the benefactor's behavior is linked to the beneficiary's pro-attitudes. (In the following, the term 'benefactor' refers to the altruist, 'beneficiary' stands for the individual or group of individuals profiting from the action in question.) The paradigmatic case for this class of behavior is the case of spontaneous, low-cost and transitory supportive behavior among strangers, such as moving aside to facilitate another person's passage, or holding a door for another person in a railway station. In everyday folk psychological explanations, we tend to explain such behavior in terms of the beneficiary's pro-attitudes rather than in terms of any of the benefactor's own. I call such explanations heterodox. Heterodox explanations go against the grain of a basic action theoretic assumption according to which behavior always has to be explained in terms of the agent's own pro-attitudes. This classic view is expressed most clearly in Donald Davidson's action theory (the *locus classicus* being Davidson 1963), and it is a basic feature of the "Humean" model of action, and there seems to be no version of intentional and rational choice explanations that does not rest on this assumption (cf., e.g., Elster 1985). If heterodox explanations of cooperative everyday behavior are literally true, much of received action theory is wrong. At the same time, the heterodox view is not limited to folk psychology and everyday talk. It seems to receive some support from two more sides at least. On the one hand, it is in tune with some old and venerable strands in the theory of empathy (Lipps 1903). On the other hand, there are some psychological theories – one regarding the link between empathy and altruism (Batson 1994), the other one regarding cooperative behavior in early childhood (Tomasello 1998) – that seem to lend some support to heterodox explanations of the behavior in question (or so I shall argue). Therefore I shall conclude that we should treat the interpretation of the behavior in question as an open issue, and grant heterodox explanations the benefit of doubt.

The final section comes back to the paradox of altruism, and addresses the hypothetical question: *if* the heterodox explanation is correct, could the behavior in question still be interpreted as the benefactor's own action? Or would the benefactor's behavior have to be attributed to the beneficiary's agency, just as metaphors such as the colloquial expression "lending a hand" seem to suggest? I argue that heterodox explanations need not displace the benefactor's own agency. The argument uses an important distinction, which could be useful for further refinement in action theory even if heterodox explanations should turn out to be wrong. The distinction is between individual intentional autonomy – i.e. the claim that each individual's behavior instantiates his or her own action – and individual motivational autarky, i.e. the claim that the explanation of each individual's behavior has to bottom out in that individual's own pro-attitudes. I argue that to reject the latter assumption does not entail rejecting the former.

§26 The Paradox of Altruistic Action

Let me start by mentioning some basic features of the concept of action that I take to be fairly uncontroversial, or at least acceptable for most analytical philosophers, social scientists, as well as for most competent speakers of ordinary language. First, for there to be action there has to be some kind of *agent*. The role of the agent is at least threefold: in most paradigmatic cases of action, the agent functions as a *source of initiative*; in all cases, he exerts some degree of *control* over the action, and functions as the entity to whom the action is *attributed*, and who can be held responsible for its consequences according to the set of social norms that constitutes or regulates the practices in which the agent participates. Typically – but not necessarily – the agent is an individual. Second, there has to be something that the agent initiates and controls, and for which she can be held responsible. Action requires some kind of *behavior*, which is basically some kind of event or series of events in the world of which the agent is in a degree of control. In the case of external action, the basic events are bodily movements which the agent does not perform by doing something else. The control need not be total; people can act even with trembling hands, just as long as there is some control involved. Third, some *goal* is needed. There has to be something the agent's behavior is all *about.* Here, some qualification is needed. As I use the term in the following, goals are not simply states of affairs the agent *wishes*, or *wants*, or *desires* to exist, or has any other kind of pro-attitude about. Rather, they are whatever has to be the case for somebody to *have done* what he or she *intended* to do. Thus goals are the conditions of satisfaction of intentions, i.e. states of affairs *as being caused* by an intention (Searle 1983). The difference between a desired state of affairs that is not a goal and a goal is this: if I do not simply desire that the door be closed, but *intend* to close the door, and you pre-empt me and close it for me, I do not achieve my goal, even though the state of the world which I desire is

now realized.[1] Thus it seems that goals are a subset of the class of desired states of affairs: agents need to *want* to do what they intend to do, but it is not the case that each of their wishes or desires gives rise, or motivates, a corresponding intention. Agents can have a desire without intending to do anything about it.

Thus desires play a dual role regarding intention. First, desires *constitute* intentions in that they capture the motivational element that is an essential part of intentions. If we assume a *narrow* concept of desire, it is true that one can intend to do things which one does not *want* or *desire* to do, such as in the case of one's intention to keep one's annual dentist's appointment (Schueler 1995; Searle 2001a). But in the current literature, "desires" extend beyond the class of objects the thought of which tends to induce a positive affective reaction. Most of the current literature uses the term in a wide sense that comprises pro-attitudes of any kind: wishes, interests, projects, commitments, inclinations, and so on (Davidson 1963: 685ff.).

The second role which desires play with regard to intentions is *motivational* rather than constitutive. The intention to see the dentist, which encompasses the (extrinsic) desire to do so, is *motivated* by one's intrinsic desire not to suffer the pain that will result from carious teeth. In this second, *motivational* rather than *constitutive* sense, desires logically *precede* intentions and provide the *motivating reason* for forming an intention which is quoted in the explanation of an intention. It is usually assumed that, in order for there to be an action, a linguistically competent agent has to be able to come up with an answer to the question of *why* she or he wanted to do what she did, and that this answer has to quote some volitional agenda of her own.[2] So, in one sense, desires describe the constitutive motivational *component* of intention (i.e. the fact that intention is a motivation-encompassing attitude), and in another sense, desires describe some necessary *antecedents* of intention, i.e. the motivational base that explain the intention.

For a desire to be able to explain an intention, however, yet another feature has to be in place. The intention (and therefore the complex of events which it causally controls) has to be *minimally rational*, i.e. the agent has to show at least some minimal degree of concern about the intention qua executive plan being conducive to the end (however successful or unsuccessful she might be at this task). To put it somewhat more cautiously: the agent cannot be *entirely indifferent* as to whether or not the induced events are suitable as a means to achieving her goals, and as to whether or not there are better means available to her. She has to *believe* that her behavior is a suitable means to realize her goals. To put it in Davidsonian terms, the *primary reason* that rationalizes the agent's behavior has to include some pro-attitude, and some suitable belief.

---

[1] As is obvious from this terminology, the concept of intention used in this chapter is largely Searlean (Searle 1983). As far as I can see, however, nothing of what I say is in conflict with a view of intention in terms of executive plans along more Bratmanian lines (c.f., e.g., Bratman 1999).

[2] It seems that the more the desire that *constitutes* the intention is *intrinsic* and *general*, the more difficult it becomes to distinguish it from the *motivating* desire that explains the intention: it is difficult to come up with anything else in explanation of one's wanting to lead a meaningful life than just that. But even if there are descriptions under which the distinction collapses, it remains meaningful; in those cases, the constitutive desire *is* the motivating desire.

These four features can all be drawn together in a single phrase. For there to be an action, there has to be some agent-controlled complex of events of which it is possible to make some (minimal) sense in terms of what the agent *wants*.[3]

If this is action, what is the problem with its being altruistic? Why should there be a paradox in the notion of altruistic action? The problem that gives rise to the paradox is this: there seems to be a basic element of *selfishness* built into the very concept of action. And this element, it seems, is at odds with an intuitive idea of genuine altruism, so that a complex of behavior cannot at the same time instantiate an action and a case of genuine altruism. To illustrate this paradox, I borrow from Thomas Nagel's analysis of altruism (1970: 80–81) in the following.

First, the element of selfishness: according to the standard theory, we cannot make sense of a complex of behavior in terms of anything other than whatever the agent happens to want *him-* or *herself*, if we are to interpret that complex of behavior as an action. In other words: behind every action are the agent's *own* desires (in the wide sense of the term explained above). This, however, does not square with an intuitive notion of genuine altruism. According to this notion, genuine altruism should be about the *beneficiary's* interests rather than about any of the benefactor's own, so that the interpretation of *altruistic* behavior should appeal to *other people's interests* rather than to the agent's. This is not altogether implausible: what makes an individual an altruist is precisely that it is possible to make sense of a more or less substantial part of her behavior in terms of the interests of people *other* than herself. And this is precisely how genuine altruists tend to explain their behavior: they did what they did *because other people wanted or needed it to be done*, full stop.[4] Why should we add some of the altruist's own interests to the story, if they don't appear in the altruist's own account?

It might seem that we moved too quickly from pro-attitudes to interests, though. Not all of our pro-attitudes are in our interest, and not all interests are reflected in our pro-attitudes: after all our wanting to breaking our bad habits might be in our interest without us *wanting* to break them, or having some other pro-attitude towards breaking them. And paternalistic cases of altruistic behavior shed a rather sharp light on that fact. In contrast to desires, interests involve the problem of *justification* (which shall be addressed below). But let us disregard for the moment cases in which altruists further their beneficiary's interests against their beneficiary's pro-attitudes. Let us concentrate on those cases in which desires do not collide with interests, and in which it is in the beneficiaries' interest to have their desires fulfilled.

With this in mind, we can now state the paradox. It is a conflict between two propositions that seem plausible at first sight, and from which we seem to be forced

[3] Needless to say, these conditions are *necessary* rather than *sufficient* for the standard concept of action. We do not, however, need to delve deeper into the analysis here as these conditions alone give rise to what I shall call the Paradox of Altruistic Action.

[4] The French phenomenologist Emmanuel Lévinas is famous for making this idea the point of departure of his thinking on interaction and society (cf., e.g., Lévinas 1991). For a lively description of the *immediacy* and of the *unthinking character* of altruistic responses see Craig Taylor's analysis of the structure of sympathy (Taylor 2003). I shall address the objection that such explanations quote *justifying reasons* rather than *motivating reasons* below.

to conclude that there is no such thing as an altruistic action. The two propositions are the following:

(i) A necessary condition for a complex of events to be an action is that, at the basic level of intentional explanation, it can be made sense of in terms of the agent's own pro-attitudes and beliefs.
(ii) If there is no conflict between pro-attitude and interest, a complex of behavior is genuinely altruistic to the degree that, at the basic level of explanation, it is to be made sense of in terms of the beneficiary's desires rather than in terms of any of the altruist's own.

Proposition (i) captures the "Humean" model of action, while proposition (ii) captures an intuitive notion of genuine altruism. To the degree to which we grant both propositions some plausibility, we might see ourselves pushed towards the conclusion that the very notion of altruistic action is an oxymoron. It seems that from a standard action theoretic perspective, a genuine altruist's behavior simply cannot instantiate the altruist's own action: his behavior would have to be attributed to the other's agency rather than to her own, since it is in terms of the other's 'desires' rather than her own that sense can be made of her behavior. If the altruist's behavior is to be taken to instantiate the beneficiary's action rather than the altruist's, it isn't altruistic, since it is rationalized by the agent's own pro-attitudes. Genuine altruism and agency are, it seems, conceptually incompatible. Either an individual is a genuine altruist, or she is an agent.

This paradox is clearly of the Zenonian kind. Since cases of altruistic action abound in real life, there *must* be something wrong with this way of putting things. The question then is: which one of the two conflicting sides is at fault? Is the theory of action with its element of selfishness to blame, or is the intuitive notion of altruism simply skewed? Should we relax proposition (i) or rather proposition (ii)?

Where this question is addressed at all in the current literature, the recommendation is unanimous: proposition (ii) is at fault. It is believed that the apparent conflict is due to an overstrained notion of altruism, which should not be taken seriously. Here are two important examples for this view. Eliot Sober and David Sloan Wilson (1998: 223) claim that to define selfishness in terms of "whatever people want" is simply short-circuited and grossly biased, leading, as they put it, to an utterly "spurious" view of altruism, because this definition leaves no room for motifs that are the agent's own, but are not aimed at the agent's own welfare. Philip Kitcher's remarks on the matter are even harsher. According to Kitcher, the whole paradox is something of a scam anyway, and only non-philosophers could ever be so naïve as to think that there is more to the problem than simple conceptual confusion (Kitcher 1998: 291). According to the line followed by Kitcher and a great many other authors, it is a mistake to think that, just because an interpretation of altruistic behavior has to appeal to other people's interests, it cannot be interpreted in terms of the altruist's own desires. There is, as they point out, such a thing as *altruistic* (or other-directed) *desires.* The view that is often claimed to be commonsensical and even folk psychological is this: altruists are normal agents, just that they have nicer desires. As with any other agents, they do *what they want*, but what they want

happens to be to further other people's interests. Thus there seems to be a way to accommodate the central feature of the intuitive notion of altruism within standard theory of action. Other people's desires *do* guide (and thus to some degree explain) the altruist's behavior, but they do not thereby *displace* the altruist's *own agency*. Rather, the link between the altruist's own behavior and the other's desire is precisely what the altruist *wants.* This link is the condition of satisfaction of her altruistic intention. Thus the respective behavior can be altruistic, and instantiate a case of the altruist's own action at the same time. According to this view, there is no paradox of altruistic action.[5] All that is needed is a clear understanding of the concept of other-directed desires. Other-directed desires are a particular kind of second-order desires whose content is the promotion of other people's interests.

§27 The Structure of Everyday Altruism

I do not dispute the existence and importance of other-directed desires. The claim I wish to defend here is the following. While the paradox of altruistic action can be resolved by relaxing proposition (ii), i.e. by appeal to other-directed desires in a *wide range* of cases, there is another class of apparently altruistic behavior, which cannot easily be fitted into this view. For this class of behavior, we have to find another solution to the paradox, or so I shall argue.

The class of behavior I have in mind here is different from action based on other-directed desires in the following respect. Whereas actions based on other-directed desires are relatively *complex,* requiring second-order desires, some degree of deliberation, and a clear understanding of other people's pro-attitudes, the behavior in question here is very simple, it is unthinking, linked to other people's immediate *goals* rather than to their desires, and in many cases it might appear more like mere reflex behavior than like a proper choice of a course of behavior.

Let me make a short detour to introduce the phenomenon. The idea for this chapter goes back to a session of the Economic Science Association at the ASSA-meeting in Chicago early in 2006. The economist and behavioral scientist Herbert Gintis opened his talk with a simple case of everyday behavior. If I recall correctly, the anecdote went something like this. When he couldn't find out how to open the door to the conference building, some passer-by who observed the scene took it upon herself to quickly press the open-door button for him, immediately leaving the scene after having helped without even waiting to be thanked. I certainly do not want to underestimate either the psychological costs of the sight of Herbert Gintis' plight, or the rewards of his grateful smile, but it seems plausible that such behavior is indeed both genuinely altruistic, i.e. not motivated by any psychological reward

[5] For the purposes of this chapter, I label this solution the *other-directed desires explanation* of altruistic behavior; I will call "other-directed" such desires (in the wide sense) as the desire to help concrete others, or groups of others, as well as such desires as the desire to conform to social norms or rules of conduct.

or cost, and quite pervasive in social life. On my way back home from the meeting, I observed cases such as the following: People holding doors for strangers carrying suitcases; people helping strangers move baby carriages in and out of trains; people moving aside on their benches so that other people could have a seat, too; people facilitating other people's passage by moving out of their way; people lifting other passenger's suitcases to and from carry-on luggage trays; a person picking up an umbrella that had slipped out of the elderly owner's hand; and, as a last example, I overheard a person trying to finish another person's sentence who got stuck in the middle of her question for directions.

I assume that such behavior permeates our entire social lives, but it is in the public sphere that its genuinely altruistic nature becomes most obvious. The observed interactions occurred among complete strangers, and typically, people didn't even wait to be thanked, but simply moved on immediately after having helped. There's genuine altruism at work here, it seems, and not just a hunt for grateful smiles.

Let's first have a closer look at how such behavior is *different* from those cases of altruism which are in the focus of much of the current literature on the topic. Compare such behavior with the paradigmatic case of altruism, which is donating to charity. At least three distinctive features immediately hit the eye. First: while it is essential to paradigmatic cases of altruism that the benefactor incur some costs (of whatever nature they might be), this does not seem to be central in the case of the behavior in question. Indeed, some degree of *indifference* seems to be the hallmark of the observed behavior. The costs incurred by the altruist are minimal, and they certainly play no role in the benefactor's own perception of the situation. This is very different from the donor's case, where some degree of self-sacrifice is part and parcel of the matter.[6] Second, something like *non-deliberativeness:* Paradigmatic acts of altruism typically involve some *care* or *concern* for the people being supported.[7] This entails that, in these cases, the benefactors are *conscious* of the beneficiary's needs, often going as far as to develop an understanding of the beneficiary's needs that differs from the beneficiary's own perception thereof (leading to vicarious, patronizing or patriarchal forms of altruism). This is very different from most of the cases listed above. There doesn't even appear to be enough *thinking* involved for any talk of care or concern to make proper sense. Many of these acts resemble unthinking, spontaneous, perhaps even impulsive behavior much more than deliberate choices. And third, the main difference: *other-goal orientation*. The behavior in question is not directed towards any of the other people's deeper needs, or well-being. Rather, it is just about other people's immediate *goals*. The benefactors in the above example support the beneficiary's in their pursuit of their immediate aims, independent of any evaluation of these goals. By contrast to paradigmatic cases of altruism, the behavior in question is more a matter of manners than a matter of morals. In short, the phenomenon is this: altruistic, low-cost, more or less spontaneous and non-deliberate behavior in pursuit of other people's goals.

[6] The people in the examples I mentioned before wouldn't risk missing their flight; serious cases of people in need of help are left to professional altruists.

[7] Often, this is even made an element of the very definition of the term "altruism".

In a next step, I use the contrast between paradigmatic cases of altruism and Gintis-class behavior to try to raise some doubts concerning the other-directed desire model's explanatory capacity. I have to admit from the outset, however, that I do not have conclusive evidence that this behavior does not fit the model. All I have are two reasons for doubt.

The first clue draws on everyday intuitions, and on ordinary language. I call it the argument from folk psychology. We usually talk about paradigmatic cases of altruism differently than we talk about the kind of behavior in question here. When social psychologists asked people who donated to charity or did volunteer work why they did so, they answered that they "wanted to do something useful" or that they "wanted to do good deeds for others", or something along these lines (Reddy 1980, quoted in Sober and Wilson 1998: 252). These self-reports are perfectly in tune with other-directed desires explanations, quoting those altruistic goals which the altruists themselves wanted to achieve. I do not know if any such study has ever been carried out, but I think it is not implausible to assume that if asked a similar question, a substantial number of the helpers in our cases of everyday altruism would have given answers of quite a different type. If asked why she pushed the open-door button, Herbert Gintis' helper might have replied something like "because (I saw that) he couldn't find it." It seems that this would have been much more natural a reply than something along the lines of "because I wanted to help him (pass the door/enter the building)". Similarly, if one asked the person on the park bench why she had moved aside a little, she would probably say "because he wanted to sit down, too" rather than "because I wanted that he could sit down beside me", "because I wanted to be nice to him", "because I wanted to avoid a conflict with him", or anything of that sort. The decisive difference is this: in explaining the behavior in question, these reports quote *other people's pro-attitudes* rather than any of the agent's own. As opposed to donators, everyday altruists are more likely to explain their behavior in terms of what *other* people want rather than in terms of any of their own other-directed desires. In other words: it seems somewhat artificial to fit this behavior into standard action theory by postulating an additional set of desires which the benefactors do not seem to know of themselves.[8] And if this is true, it seems that what we have here is a case of genuine altruism in the strong intuitive sense mentioned above.

The question, of course, is: how literally should we take such manners of speaking? Self-reports and ordinary language are not the ultimate source of authority in action theory, let alone mere conjectures concerning *possible* self-reports. But it remains a remarkable fact that, in everyday life, people quite often *do* refer to other people's pro-attitudes rather than to any of their own when they are asked to explain their behavior. Especially with Ockham's razor in mind, the question should be asked: what reasons do we have to assume more desires to explain these people's behavior than those they quote themselves when they give an account of what they do? What reasons, that is, besides the fact that this happens to be what standard action theory requires us to do?

[8] It is true that not all desires need be conscious. But it is plausible to assume that desires cannot be *inaccessible to consciousness* (cf. Searle 1981), so that under suitable circumstances, agents are conscious of their desires.

Yet there are at least two obvious objections to the line of argument developed so far. The first objection is that the behavior in question should be interpreted in terms of some desire that one's behavior be guided by the rules of politeness, or some such norm-oriented motivational states from the altruist's part. The person moving aside on the park bench might not have the desire to have the other person sit beside her. But surely, it might seem, she will have the desire to be polite. The reason why I'm not convinced by this objection is the following: whereas most forms of the type of behavior to which our cases of everyday altruism belong are sustained by social norms of politeness and propriety, we can easily find cases in which such behavior actually *violates* social norms. I shall come back to this below.

The second objection is more fundamental. It is this. It might seem that the whole problem with the paradox of altruism arises from an equivocal use of the term "explanation" in the exposition of the paradox of altruistic action above. Whereas in proposition (i) above, "explanation" refers to *motivating reasons* of the behavior in question, it is about *justifying reasons* in proposition (ii). Justifying reasons differ from motivating reasons in the following respect. Whereas motivating reasons *rationalize* a given complex of behavior, *justifying reasons* are facts that seem to make a given goal worth pursuing (cf., e.g., Pettit and Smith 2004: 270). It seems plausible to assume that the two kinds or reasons are mutually independent, and that they may be different kinds of entities. Whereas beliefs and pro-attitudes seem to be the only plausible candidate for *motivating* reasons, it might appear that *all sorts of facts* may act as justifying reasons. Thus it seems that when explaining their behavior in terms of the needs or pro-attitudes of *other people*, altruists refer to the *justifying reasons* for their actions rather to their motivating reasons; they answer the question of why it was *right* or *morally required* to act the way they did rather than the question of what kind of pro-attitude rationalizes their behavior.

In order to assess the validity of this objection, it seems useful to modify some of the above examples so that it can be excluded that the "why"-question is answered in terms of justification rather than motivation. The following thought experiment was suggested to me.[9] Take again the case of Herbert Gintis standing in front of the closed door with his helper observing the scene. But now suppose that there is another person on the scene, the helper's colleague, who is much closer to the button and whom the helper knows to be familiar with the opening mechanism, the helper sees that his colleague observes that Gintis cannot find the button. But the colleague remains inactive, so the helper steps in and pushes the button.

In this situation, the question "why did you push the button" acquires a different meaning (Garfinkel 1981: chap. 1): the question is not "why did you push the button *rather than doing nothing*", but "why did *you rather than he* push the button". As both the helper and her inactive colleague seem to have the same *justifying reason* for action, it seems clear that the explanation for the difference between the helper's and her colleague's response has to be given in *motivational* rather than justificatory terms.

[9] I wish to thank an anonymous referee for Economics and Philosophy for this suggestion.

The question now is: what would the helper's response be? Would she say something along the lines of "Because I want to be helpful to others when I can, whereas he (the non-helping colleague) lacks any such desire", thereby quoting one of her own pro-attitudes? I'm not convinced at all. If the question is simply "why did you push the button" in the presence of the colleague, the most natural reply seems to be "because he (Gintis) wanted to enter the building and my colleague didn't help him". If the question is "why did you *rather than he* push the button", the most plausible reply seems to be something along the lines of "because he (the colleague) was inattentive/wasn't paying heed/didn't bother". None of these answers quote any of the helper's own pro-attitudes.[10]

I conclude that ordinary language and folk psychology does not support the claim that where justifying reasons are not the issue, explanations can only be given in terms of the agent's own pro-attitudes. And I argue that pro-attitudes are not needed to explain the difference between everyday altruists and apparent everyday egoists. It seems to me that ordinary language simply does not square nicely with the view that the only things that move us are our own beliefs and desires.

The second argument for the view that everyday altruism requires us to relax proposition (i) rather than (ii) rests on conceptual analyses and some scattered empirical observations. Recent research in developmental psychology has shown how very basic the understanding of the purposiveness of other people's behavior is in human cognition. Toddlers are experts in identifying other people's goals, and they are exceptionally successful at this task long before they develop any theory of mind (Tomasello 1998). Infants can grasp what another person's behavior is supposed to achieve long before they pass false belief tests, i.e. long before they have an idea of the other agent as acting on beliefs which may differ from the infant's own.

Thus it seems that the understanding of other people's *intentions* (in terms of what this person tries to achieve) is basic to the conception of another person, not the other way around. Now there seems to be some empirical evidence that links the understanding of another person's intentions to action tendencies of the observer. I will not go into the controversy that revolves around simulation theory and the role of mirror neurons here; instead, let me only point out that a tight conceptual link between understanding on the one hand and acting on the other is at the very origin of the history of the concept of empathy. Theodor Lipps (who can be considered the father of the concept even though others used it before him) observed such phenomena as people in an audience who, sitting in their seats and watching a tightrope walker, seemed to compensate the acrobat's imbalances with movements of their own bodies. Empathy is, Lipps claims, "internal co-action" ("innerliches Mittun";

[10] We should be careful here not to be too quick to jump from such descriptions to ascriptions of pro- (or rather: con-) attitudes. A person might well have the desire to help others, but be helplessly unperceptive or slow on the uptake as far as other people's goals are concerned. Also, it seems obvious that individuals who *do* perceive other people's intentions are not entirely *passive* with regard to the degree to which they let other people's attitude guide their behavior. One might well direct oneself to be more or less accommodating towards other people's intentions – a training which might be guided by one's desire.

Lipps 1903).[11] Understanding isn't originally motivationally neutral. There is an action impulse that flows *directly* from the very understanding of the other agent's behavior, and it is aimed towards the same goal. Thus the very act of understanding provides some motivational steam in and of itself, and it seems that much of the behavior described above operates under this steam. It seems that in the act of understanding, the cognitive (or theoretical) and the conative (or practical) components are internally intertwined.[12] This is shrouded by the fact that it is possible to disentangle the two elements. It is possible to *suppress* the sympathetic components involved in empathy, and to understand other people's behavior without any cooperative impulse, or even to combine empathy with *antipathy*, such as in the notorious case of the cruel person who gloats over his victim's suffering. Some authors have taken this case to prove that there is no internal link between empathy and sympathy at all (Scheler [1912] 1954). Entirely un-sympathetic empathy, however, is *derivative*. Take the case of somebody getting stuck in the middle of a sentence. There is an immediate impulse in the listener to finish the speaker's phrase, perhaps even against his or her own wishes. Similarly for the case of an elderly person struggling to lift her suitcase on the luggage tray. To understand what she is struggling to achieve already means to have an *impulse* to lend her a hand. It does not seem necessary to assume any extra desire to be a helpful person, or a desire to be kind to other people, or some such antecedent motivational state on the altruist's part.[13]

From these (admittedly weak) clues emerges a picture of the intentional structure of the behavior in question that is very different from other-directed desires explanations. It seems that the perception of the beneficiary's intentions is much more closely linked to the benefactor's behavior than the other-directed desires explanation has it. If the suggested line of interpretation is correct, it seems that everyday altruism does not require a particular desire from the benefactor's part for the purpose of a motivational explanation of her supportive behavior.[14] If a person moves

[11] For a reconstruction of Lipps' account in the context of the current debate on the topic see Stueber 2006.

[12] The current discussion on "besires" (a word that combines "belief" and "desire" to refer to mental states that seem to have both world-to-mind and mind-to-world direction of fit, but are different from declarations) seems to come close to the phenomenon at issue here. The paradigmatic case in the besires literature, however, is moral judgment, which is very different in structure from empathy.

[13] In current research on altruism, the assumption that there is a tight link between the understanding of other people's intentions and action tendencies receives strong support from Dan Batson's research. Based on his experimental work, Batson has developed the "altruism-empathy hypothesis" that states that empathy and altruistic action go hand in hand (cf., e.g., Batson 1994).

[14] What is important, however, is the difference between the agent's *not having a desire* to x in terms of the agents not having a pro-attitude towards x, on the one hand, and the agent's having what one might call a "con-attitude" towards x, on the other. For this point, and for the most compelling case for other-motivated behavior in the received literature, cf. Paprzycka (2002). If there is a con-attitude, or some other conflicting desire from the agent's part, other-motivated behavior will not ensue. But this does not mean that there needs to be an additional pro-attitude in addition to the perception of the other's intention for the agent to behave as she does, if the con-attitude is lacking.

aside on a park bench, she might do so simply because she *sees* that the other person wants to sit down (cf. Paprzycka 2002). The motivation is in the perception of the other's desire rather than in anything *she* wants. In this sense, the intentional structure of the behavior in question would indeed not be rooted in the agent's own desires, but in the other's. Insofar as this is the case, we may label everyday altruistic behavior *other-motivated.*

§28 Another Solution to the Paradox

I am well aware that I have not presented any conclusive evidence for the assumption that there really *is* such a thing as other-motivated behavior. So far there is no proof that any of the cases I have discussed *cannot* be explained in terms of other-directed desires; all I have are reasons for doubt, coming from folk psychology, the theory of empathy, and, as we shall see in the following, from some strands in psychological research. I think, however, that these clues are strong enough to give the heterodox interpretation of the behavior in question – i.e. the interpretation that takes the behavior in question to be other-motivated rather than other-directed desire motivated – the benefit of the doubt. I will not go further into this issue here, but instead come back to the initial problem. In this third and concluding part, I will address a purely hypothetical question: *if* other-motivated behavior really existed, could it be altruistic *action*? If we reject other-directed desires explanations, how, then, can the paradox of altruistic action be resolved?

Let me start by stating again the basic problem, which is to show how an individual's other-motivated behavior could instantiate his or her own action. This is an important problem to solve, because one reason why most philosophers do not even think of the possibility of other-motivated behavior seems to be the belief that this problem cannot be solved. If a benefactor's behavior were indeed to be other-motivated, it seems that it would be altruistic in the naïve, strong, intuitive sense mentioned at the beginning of this chapter, leading right back into the paradox of altruistic action.

The argument for this view is the following. According to the orthodox view that is most famously expressed by Davidson, "*R* is a primary reason why an agent performed the action A under the description *d* only if *R* consists of a pro attitude of the agent toward actions with a certain property, and a belief of the agent that *A*, under the description *d*, has that property" (Davidson 1963: 687). Again, this does not mean that no other pro-attitudes than the agent's own can shape his behavior. Rather, the claim is that other people's pro-attitudes are taken into account only insofar as this is what the agent *wants*, i.e. insofar as there are other-directed desires in which the motivational explanation of an action bottoms out. This is at odds with other-motivational behavior. Since an adequate intentional interpretation of the behavior in question would not bottom out in any of the *benefactor's* desires, but in the *beneficiary's*, it seems that, according to the standard notion of action, the behavior in question would have to be attributed to the *beneficiary* rather to the benefactor

himself, since it is the beneficiary's pro-attitudes in terms of which sense can be made of the benefactor's behavior, not any of the benefactor's own. This, however, is in conflict with the *deep-seated notion that, under normal circumstances,*[15] *each single individual's behavior constitutes that individual's own action, i.e. that each individual is an agent*. This is a notion that should not be dropped light-heartedly. It seems it cannot be dropped without thereby excluding some members of the class of agents, and that does not seem right. The idea of individual agency is basic for our most elementary practices of mutual score-keeping in ascribing commitments and entitlements, rights and responsibilities. The importance of this notion can be further emphasized by pointing out its connection to core normative notions. Dropping the idea of universal individual agency (or individual intentional autonomy, as I shall call it) may result in taking those individuals who play subordinate roles in social life to be their superior's extended body, rather than agents' of their own – remember that Aristotle's slave is defined by the fact that he or she is his or her master's instrument. In this case, all sorts of authoritarian, patriarchal or even worse ideas about which individuals do and which do not count as agents in their own right seem to be licensed. The question of who counts as an agent becomes a matter of societal power distribution. And this simply does not seem right. Even convinced Foucaultians are reluctant to deny the powerless a claim to their own agency.

This is one of the reasons why we should not let go of the idea that each individual's behavior instantiates his or her own action. But let's focus for a moment on why this basic and deep-seated assumption seems to be at risk here. If other-motivated behavior were to exist, it seems that the expression "lending one's hand" would have to be taken quite literally. It might appear that whoever lends his or her hand would thereby cease to be the agent behind his or her hand's behavior; that behavior would then have to be attributed to the agency of the person to whom it is lent out, as it were. The role left to the benefactor would indeed be no more than that of a mere *instrument*, or *organ*, of the beneficiary's will, not that of an agent in his or her own right. And this seems utterly implausible, even more so than in the case of submission to power. It is simply not true that people lending their hands lose their status as the agent behind their hand's behavior. They are still held responsible, and even from an internal perspective, it is implausible to assume that lent hands become parts of the other's extended body. Lent hands do not move on the other's remote control. So there seems to be no reason not to hold on to the idea that, insofar as an individual's behavior instantiates any case of action at all, it has to be *the individual's own action*. If we hold on to this assumption, however, the verdict against other-motivated behavior seems to be spoken, for it appears that an individual's behavior can be an action only if it is possible to make sense of it in terms of that individual's *own* intentions and desires. Thus it seems that there can be no other-motivated behavior.

I think that this line of reasoning is flawed. There is a way to reconcile the benefactor's own agency with the possibility of his or her behavior's being

[15] "Normal circumstances" exclude such cases as reflex behavior, coughing, sighing, blinking, where such behavior is not purposefully caused by the agent as a part of his action.

other-motivated. The apparent verdict against other-motivated behavior is due to a confusion in our standard conception of agency that needs to be sorted out. It is necessary to distinguish the following two claims, which are usually lumped together:

1. Individual Intentional Autonomy: *Under normal circumstances (barring certain cases of reflex behavior and pathological dissociations between will and action*[16]*), an individual's behavior is to be interpreted as his or her* own *action.*
   I take it that this assumption is at the heart of our standard conception of agency, and I suggest that for some of the reasons I mentioned above, we should hold on to it. The problem, however, is that on a regular basis the claim of individual intentional *autonomy* is lumped together with a further and much stronger claim:
2. Individual Motivational Autarky: *Any motivational explanation of an individual's behavior has to bottom out in some of that individual's own desires*.

Whereas intentional autonomy is a thesis on which agency is instantiated by a given complex of behavior, *autarky* is a thesis about the motivational resources on which agents may draw. As the thesis states that the only resources are the agent's *own*, it claims that individuals are something like closed intentional economies. Therefore I label this thesis "motivational autarky".

Let's have a closer look at this second assumption before examining how it relates to the first. What does "bottoming out" mean in this context? The assumption of individual motivational autarky does not mean that other people's pro-attitudes couldn't play any role in the explanation of an individual's behavior. As stated above, nobody denies that people do sometimes act on other people's desires or intentions, but it is claimed that they do so if and only if they have *a desire of some sort to do so*, i.e. *on the basis of* an other-directed desire *of their own* in which an intentional explanation of their behavior has to be based. So individual motivational autarky is compatible with altruism in terms of other-directed desires-explanations. It is not, however, with heterodox explanations. For the whole point of these explanations is that they appeal to the beneficiary's pro-attitude rather than to the benefactor's own.

In orthodox explanations of altruistic actions, it is always true that benefactors do what they want because they want it. Orthodox explanations need not thereby deny that sense can be made of the benefactor's behavior in terms of the beneficiary's pro-attitudes. But, in this view, this is only true because there is yet another underlying pro-attitude on the benefactor's part that sustains that link. One can easily imagine more sophisticated cases, in which the boundary between orthodox and heterodox explanations seems to blur. Imagine somebody explaining his altruistic actions with his other-directed desires, but then giving an account of these other-directed desires in terms of a third party's will: "I want to help you because He ordains me to do so."

[16] These include such cases as the *alien hand syndrome,* in which the patient's hand seems to follow an agenda of its own, which even may include the murder of its owner. Other forms of dissociation between will and behavior include *echopraxy* in which patients compulsively imitate the behavior which they observe.

Even though this explanation includes references to other-directed desires, it is heterodox, because it bottoms out in another person's ("His") will (note that, in this case, that other person is not the immediate beneficiary). According to the orthodox view, we have to assume yet another, more basic desire to make sense of this case, e.g. the benefactor's tacit desire to want to do what He wants him to do, or some such additional desire.

So the difference between orthodox and heterodox explanation of the benefactor's behavior is not really a question of whether there are other-directed desires around or not. Rather, the question is where the chain of pro-attitudes quoted in the motivational explanation of the altruistic behavior in question ends (this is what "bottoming out" means in this context): in the benefactor's own pro-attitudes (orthodox explanation), or rather in some other individual's (heterodox). Of course, the simplest case is the most important, where the difference between the two views becomes particularly obvious.

Heterodox explanations are incompatible with individual motivational autarky. Does that mean that they have to fly in the face of the assumption of individual intentional autonomy, too? Let's now have a closer look at the relation between the two assumptions. My thesis is the following: while 2 implies 1, the converse is not the case. Thus there can be individual intentional autonomy without motivational autarky. While heterodox explanations are incompatible with individual motivational autarky, they are in tune with individual intentional autonomy. In other words, the benefactor's behavior need not be taken to be based in any of his or her own pro-attitudes to be interpreted as instantiating his or her own action.

How is this possible? How can a behavior be interpreted as an individual's action without the interpretation bottoming out in that individual's own desires? The answer is this. If one acts on another individual's pro-attitudes, one can form an *intention* to do whatever is necessary so that the other's goal is achieved without having a particular *desire* to do so. In this case, it is true that the benefactor does what he or she *intends*, but it isn't true that he or she does what he or she *wants.* So the *intention* in terms of which the benefactor's behavior is to be made sense of is the benefactor's own, but not the *desire* in which it is motivationally based. Thus there is still a sense in which we *want* to do what we intend to do; intention is a motivation-encompassing attitude (Mele 2003), and it remains so. The constitutive desire is ours. But not so for the motivating desire. On occasion, it is not the case that our constitutive "wanting to A" is motivated by any of our own desires. Our wanting to A may well be motivationally explained by other people's desires. This does not displace our own agency. Thus the benefactor is not intentionally autarkical, but he is intentionally autonomous. It's not that the beneficiary acts *directly* through the benefactor's behavior, as if on the other's remote control.[17] The benefactor still does what he or she intends (and thus constitutively wants) to do *him- or herself*. This case shows how one's behavior can still be interpreted as *one's own action*, even though *the intentional interpretation of one's behavior does not bottom out in one's own individual pro-attitudes*.

[17] For a different heterodox view on the matter cf. Paprzycka (2002).

In other words, and to add a new species to the philosophical literature: I'm not the other's *motivational zombie* if I move aside to make room for him on the park bench without having any corresponding other-directed desire of my own. Motivational zombies are individuals whose behavior does *not* constitute *their own* actions, but rather another individual's, or a group's.[18]

Intentional zombies abound in Sci-Fi, in early accounts of hypnosis and mass-suggestion, in self-reports of schizophrenics, and in some of Al Mele's recent books. But do intentional zombies exist in the real world? It is clear that people can be manipulated into doing all sorts of things; but it is important that only the strongest type of manipulation would amount to intentional zombieism.[19] It is sometimes said that under hypnosis something closely approaching intentional zombieism can be effectuated.[20] I have no knowledge of such cases, and it seems that if they occur, they are limited to very short behavioral sequences. It is clear, however, that everyday altruists are not motivational zombies. My moving aside is still *my own action,* but the intentional resources going into it – the desires motivating my behavior – extend beyond my own pro-attitudes. My intentions are linked to the other's pro-attitudes in much the same way in which normally, my intentions are linked to my own motivating desires. Just as I normally form the intention to sit down on the basis of my own desire to rest a little, without needing another, yet more basic desire to do what I want to do, I can form the intention to move aside on the base of the other's desire to sit down, without there being an additional desire to do what the other wants. Nothing about this structure is particularly mysterious. And it does not affect the agent's individual intentional autonomy.

Again, I do not claim that this is what's actually happening; all I claim here is that it is not *necessary* to abandon the principle of individual intentional autonomy to accommodate other-motivated behavior. All that needs to be abandoned is the dogma of individual motivational autarky.

---

[18] It might be noted in passing that this type of philosophical zombie seems to be somewhat closer to the voodoo idea of zombieism than the "phenomenal zombies" that abound in the philosophical literature, as the distinguishing feature that marks out zombies from other creatures does not primarily seem to be that zombies do not have a consciousness, but that they do not have a *will.*

[19] If the students in a class agree to keep quiet or show a friendly face when the teacher is on the front left side of the room, and chatter or look angry when she is on the right, and if they don't do this too conspicuously, they will soon find their teacher sticking to the left all the time, probably without having any idea about the scheme. But this does not bypass the teacher's agency. It isn't the case that the teacher didn't *intend* to do what she did: upon questioning, she will probably answer that she "likes it better" to be standing on the left side. The problem is, that she does not know that she wanted to do what she did because other people manipulated her into wanting it. So the teacher is very far from being the class' intentional zombie.

[20] It is claimed that in deep hypnosis people may be instructed to show some nonsensical behavior, and that after having woken up, they do what they have been told without having a memory of the instruction, and without having the slightest clue of why their hand suddenly moves. From the internal perspective, the phenomenon is that of the *alien hand syndrome* mentioned above, only that this time there is another person's will behind the behavior, which amounts to intentional zombieism.

If this is true, if individual motivational autarky is no essential conceptual ingredient of action, the question arises: how come it is always lumped together with the idea of intentional autonomy? *Why do we tend to mix up the idea of being the agents* responsible *for our behavior with the very different idea that, in the last resort, only* our own desires *are fit candidates to make sense of our behavior?*[21] In short, my answer is this: it's because in *our culture*, at least, motivational autarky describes the way people are *supposed to be* (and see themselves). Being the one and only ultimate motivational source of the intentional infrastructure of one's own behavior is not a conceptual feature of agency, but it is a very strong normative ideal.

This claim needs some explanation, especially in view of the thoroughly *positive picture* of other-motivated behavior that I have given so far. Think of the person holding the door for another person, or the person moving aside on the park bench, or the one assisting an elderly person with her luggage. How could all these spontaneous niceties, these acts of kindness ever be *in conflict* with any plausible normative ideal? Given the list of examples at the beginning of the chapter, it might even be tempting to explain other-motivated behavior as a kind of *internalization* of social norms. After all, what all the people in the above-listed examples are doing is just being polite. It is important to see, however, that while in most cases other-motivated behavior can be seen as "pro social", there are other cases in which it goes *against* the norms of proper conduct. Thus we might be required to *suppress* the impulse to finish a sentence for a person who is struggling with stuttering – out of simple respect for that person's integrity. And, in education, it is very often *against the norms of proper conduct* to let oneself be carried away by one's other-motivated impulses, because children need to be given the opportunity to exercise their own agency. In most cases, social norms do favor other-motivated behavior. In some other cases, however, this is not true. In these cases, it is not just important that people's goals are achieved; what's even more important is that people can achieve their goals *themselves.*

While a person's explaining her behavior in terms of another person's intentions is frequent in everyday talk, we tend to press for "deeper" explanations, and even to react *embarrassed*, if a person fails to come up with one of her own desires in explanation of her behavior. *People, we seem to think, shouldn't be doing things just because other people wanted them to be done.* People should be *self-reliant* about their own goals, and not be a motivational pawn in other people's play. Thus motivational autarky seems to be part and parcel of our idea of full-blown selfhood and personal identity.

A vivid illustration for the value of motivational autarky and the dangers of other-motivational behavior is provided by Stanley Milgram's famous experiments. As most readers will remember, Milgram's test subjects – perfectly decent ordinary people from a suburban milieu – proved to be willing to administer potentially deadly electroshocks to innocent others, just because they were told to do so by

[21] This conjunction of the idea of individual intentional autonomy and individual intentional autarky is particularly strong in Philip Pettit's *Common Mind* (1996), where he defends intentional psychology against collectivism under the label "individual autarchy".

some authority figure within an experimental setting. There were neither financial incentives nor sadistic inclinations involved. So how come those people did what they did? Milgram himself explains his results by what he calls an "agentic state" (Milgram 1974). An agentic state, Milgram says, is a condition in which a person sees herself as acting on another person's desires rather than his or her own. In Milgram's view, his test people's perception that the motivational base of their behavior is alien to their own psyche explains why their conscience is outflanked by their behavior: only behavior that is motivationally based on the agent's own desires is subject to moral control. While the behavior of Milgram's test people is very different from other-motivated behavior as characterized above in many respects, his concept of the agentic state captures very nicely the central feature of heterodox explanations, according to which people sometimes do what they do, not because of anything they want themselves, but because of what other people want.

As far as I can see, Milgram does not give a clear answer to the question of whether the agentic state is an actual fact about the motivational structure of agency, or whether it is just the delusional self-image of people acting under the influence of authority. However, he seems to be somewhat biased towards the latter reading when he reproaches his compliant subjects for being unable to keep their own act together and assuming responsibility for what they did by claiming that they acted on none of their own desires (this is particularly obvious in Milgram's discussion of Elinor Rosenblum, which is one of the case studies in Milgram's book). In these passages of Milgram's analysis, the agentic state seems to be no more than the test people's attempt to protect their self-image from what they did by blaming the events on the authority. Also, Milgram depicts the agentic state as an *unusual* condition, one that requires the presence and massive influence of authority. And, as is well understandable from the setting of his experiments, he portrays agentic states as morally utterly condemnable. Thus the normative ideal of motivational autarky becomes very clear in Milgram's depiction of the fatal consequences of agentic states.

By contrast to Milgram, and in light of the above examination of the structure of motivationally non-autarkical behavior, I propose to consider three things: first, it seems worthwhile not to dismiss the possibility that experiences of agentic states might be more than just cover-ups used by agents to keep their self-image clean of their wrongdoings. The alternative is to see the self-perception involved in agentic states as referring to actual matters of fact about the motivational structure of behavior. Second, we should consider the possibility that agentic states might be a rather normal condition that permeates much of our everyday live and need not be limited to the presence of authority figures, and which, third, may lead to morally disastrous consequences under conditions such as those examined by Milgram, but can also be very beneficial under such circumstances as to be found in airports and railway stations, among many other places.

Considering the wide range of behavior at stake here, it seems difficult to answer the hypothetical question of whether or not we should uphold the idea that people shouldn't do things only because other people wanted them to be done if it turns out that motivational autarky is in fact a value and not a conceptual feature of action.

I will not pass any judgment here. What is certain, however, is that we cannot even discuss the question of whether or not motivational autarky is indeed an ideal worthy of defense, if we continue mixing it up with intentional autonomy. Because intentional autonomy is a constituent of *any* action, it is not to be changed. By contrast, motivational autarky might turn out to be a cultural ideal, which we may or may not want to uphold.[22] In either case, it is important to distinguish the two.

I conclude with a brief summary of my line of argument and with a remark on the bearings of my results for the economic model of human behavior. The paradox of altruistic action consists of two propositions, which seem to be intuitively plausible, but mutually exclusive. The first proposition is that, for a complex of behavior to be an action, it has to be based on the agent's pro-attitudes. The second is that, for a complex of behavior to be genuinely altruistic, it has to be made sense of in terms of other people's interests rather than in terms of any of the altruist's own. The received literature tends to solve this paradox by relaxing the second proposition and by allowing altruistic action to be based on a particular type of desires. I argued that, while this solution is convincing for a wide range of cases, there is a particular class of altruistic action with regard to which it does not seem to work well. I defined everyday altruism as spontaneous cooperative behavior in low-cost situations, and I provided some clues that seem to indicate that, in order to accommodate such behavior, we might be forced to relax the first proposition. In the last section, I distinguished a weaker from a stronger reading of the first proposition, and I labeled them individual intentional autonomy and individual motivational autarky. There is no conceptual necessity to assume that agents need be motivationally autarkical.

I left the decisive empirical question of the role of motivational autarky in human interaction open, and I should say a word on how the question could be decided. Throughout the chapter I took desires or pro-attitudes to be whatever rationalizes an agent's behavior, given his beliefs. I take rationalization to be a matter of motivation rather than justification. Even though there are unconscious beliefs, I take a special epistemic authority to lie with the agents' themselves, so that the question of whether or not an agent has a desire is answered by the agent's assent under suitable conditions. Unconscious desires are such that they become conscious under suitable conditions (cf. Searle 1983). In light of this view, the question concerning folk psychology becomes particularly important, at least if we assume, as seems reasonable to do as long as there is no evidence to the contrary, that ordinary language is not permeated by some "false consciousness"; hence the special emphasis on folk psychology in the above.

To wrap up the argument, a short remark on the consequences for the economic model of behavior. In the critique of economic thinking, the idea that the link

[22] My claim that intentional autarky is a cultural ideal does not entail that it is necessarily culturally relative, as there might be – and actually are, I would like to think – values that are upheld in all cultures. Considering the ambivalent role that individual intentional autarky plays in our lives, however, I would expect that it is stronger in some cultures than in others. This, however, is an empirical question.

between people's desires and their choices might not be as tight as is assumed by the orthodox account has played a mayor role. The analysis of phenomena such as weakness of the will have helped to cast serious doubt on the assumption that people's behavior adequately reflects their desires. The argument delineated in this chapter hints at yet another way in which the link between what people want and what they do might be more complex than is normally assumed. It might be that, even where behavior does reflect an individual's motivating desires, the desires in question might not be the *agent's own*, after all. Where agents are not motivationally autarkical, i.e. where people's psychologies are permeable to other people's motivations, and where there are relations of spontaneous cooperation or mutual identification between agents, including some forms of influence, power and authority, it might not be so easy to say whose desires an individual's choice reveals. The argument developed in this paper strongly suggests that people can and should be seen as *agents* even if they are not motivationally autarkical.

In his powerful and trenchant critique of rational choice theory, Amartya Sen has claimed that committed agents may act on other people's goals without making them their own (cf. Sen 1985; Peter and Schmid [eds] 2007). While most of Sen's critical points are widely accepted, this particular and uniquely radical claim has been met with considerable skepticism (cf., e.g., Pettit 2005); it has been argued that any violation of the assumption of self-goal choice would simply *displace* the individual's agency. In an earlier paper, I have argued that Sen's claim does make sense as far as *shared desires* are concerned (Schmid 2005a), at least as far as shared desires are irreducible to interrelated individual desires. In light of the above considerations, I would now tend to go even further and argue that there is a sense in which a desire does not have to be an individual's own, or jointly held with other individuals, in order to motivate that individual's action. Still, an individual needs to have his or her own goals in order to be an agent. But, as far as motivational autarky is not part of the concept of agency, the motivational base of the intention that defines the goal need not be any of the individual's own desires for her to be an agent.

It is tempting to think of the relation between other-motivated action and other-directed desire-motivated action as a matter of some switch of frame of mind, in which, in a given situation, one may either act spontaneously and unthinkingly on other people's desires, or decide to take the time to think about the matter and follow the orthodox route by basing one's decision on one's (egoistic or altruistic) pro-attitudes. I admit that this may very often be the case; but Sen's discussion of the structure of committed action seems to point towards the possibility that other-motivation might not only be a matter of unthinking low-cost cooperative reflex behavior, but extend to fully conscious and deliberate choices in which the stakes are great.