

<https://helda.helsinki.fi>

---

## Understanding the norm of self-interest

Ylikoski, Petri K

Acta Universitatis Upsaliensis  
2022

---

Ylikoski , P K 2022 , Understanding the norm of self-interest . in J Erola , P Naumanen , H Kettunen & V-M Paasivaara (eds) , Norms, Social Structures and Policy Changes : Essays in honor of Hannu Ruonavaara . Uppsala Housing and Urban Studies , vol. 1 , Acta Universitatis Upsaliensis , Uppsala , pp. 21-39 .

---

<http://hdl.handle.net/10138/343478>

---

cc\_by\_nc\_sa  
acceptedVersion

---

*Downloaded from Helda, University of Helsinki institutional repository.*

*This is an electronic reprint of the original article.*

*This reprint may differ from the original in pagination and typographic detail.*

*Please cite the original version.*

# Understanding the norm of self-interest

Petri Ylikoski

This paper discusses social psychologist Dale Miller's (Miller 1999) idea of the norm of self-interest. According to him, there exists a widely accepted pseudo-scientific folk theory about the "deep" motives of human behaviour that generates self-fulfilling prophecies regarding self-interested behaviour. People behave selfishly because they think that other people expect them to behave selfishly. Thus, there is a norm of self-interest. Although this phenomenon may also be called a self-fulfilling ideology (Schwartz 2012), calling it a norm is justified because it is based on shared beliefs and expectations about what is considered appropriate behaviour. The norm suggests that self-interest both is and should be a powerful determinant of behaviour. If people deviate from these expectations, then they will feel sanctioned. The norm of self-interest is thus not only a descriptive but also an injunctive norm. Its influence shows both in how people behave and in how they account for their actions. It is a norm also in the sense that things could be otherwise: if people were not captured by the folk theory about motivation, their behaviour would be significantly different. Thus, the norm is both contingent and conventional.

There are many ways to understand the notion of self-interest (Kim 2014). In Miller's discussion, self-interest refers to concern with one's own material interests. People who employ a more flexible, or vague, notion of self-interest may say that they are discussing narrow self-interest. In any case, Miller's thesis is that a widely shared folk theory about human motivation is based on this notion. According to this theory 1) monetary incentives are important determinants of choice and 2) people's attitudes reflect their material interests. Miller does not deny that people are concerned with their material interests, but rather he claims that because of the norm of self-interest, these selfish concerns play a larger role in social life than they would play if the norm were weaker or non-existent. People who are influenced by self-interest believe that selfish motives are both more prevalent and stronger than they actually are, and the expectations that are fuelled by these beliefs produce more selfish behaviour than would otherwise occur. The norm of self-interest thus amplifies people's selfish behavioural inclinations.

I proceed in the order as follows. In the next section, I outline some philosophical reasons why I find the norm of self-interest an interesting notion. Section 3 summarizes the basic idea of the norm of self-interest as it is presented in Miller's papers. In section 4, I discuss some experimental results that can be used to evaluate the evidential support for the theory and consider some additional theoretical ideas that could be used as additional support for the theory. Section 5 addresses some ambiguities in the notion of the norm of self-interest, and section 6 concludes my discussion.

## Why is this an interesting idea?

There are several theoretical reasons why I find the idea of the norm of self-interest worth consideration. First, this notion challenges a widely accepted idea concerning the explanation of human action in which self-interest is the ultimate motive of human action. Often, this idea is combined with the notion that behaviour is inherently understandable only if it can be seen as rational. These ideas come together in the thinking of many rational choice theorists, who often seem to accept that rational explanations are somehow explanatorily privileged. I have elsewhere criticized this type of intentional fundamentalism as being incompatible with social sciences that are oriented towards a causal explanation of human actions (Ylikoski 2012, Ylikoski and Kuorikoski 2016); therefore, I will not discuss it further here.

Underlying these ideas is the idea of the ideal of natural order, which was coined by philosopher Stephen Toulmin. According to Toulmin:

“Our ‘ideals of natural order’ mark off for us those happenings in the world around us which do require explanation, by contrasting them with ‘the natural course of events’ – i.e., those events which do not. Our definition of the ‘natural course of events’ is therefore given in negative terms: positive complications produce positive effects, and are invoked to account for deviations from the natural ideal, rather than conformity to it.” (Toulmin 1963, 79.)

Rational choice sociologists James Coleman and Thomas Fararo borrow this idea for the explanation of human action.

“... the idea of seeing collective life as irrationally bent ... arises from the postulate of individual action as rational. It does not arise from the empirical generalization that all action is manifestly rational. ... As Toulmin puts it, a theoretical discipline is often founded on a principle of natural order” (Coleman and Fararo 1992, xiv.)

I think that this is a confused idea. It confuses the evaluation with the explanation of human action. To be truly causal, our explanatory practices should be symmetrical. This quality is recognized by Toulmin who contrasts pathology and biochemistry as follows:

“From the pathological point of view, the proliferation of cancer-cells is something ‘unnatural’, being a deviation from the standard physiological function. For the biochemist, however, the task is to identify all the processes going on in the body equally, functional and dysfunctional alike. What we call ‘normal’ physiological processes are a special case, which to the biochemist are neither more nor less ‘natural’ than pathological ones. So, at the biochemical level, the only theory giving complete understanding will be one which treats normal and pathological processes on a basis of complete equality.” (Toulmin 1963, 80-81.)

Thus, we should not privilege the rational explanations of action. There are many reasons for this approach. First, the idea that the human decision-making process consists of a

rational core that is then disturbed by some additional cognitive mechanisms (the biases) is not well-supported by cognitive science (and neurosciences) and would be in fact an evolutionary miracle. Second, privileging rational self-interested action creates quasi-philosophical pseudo-problems. For example, this stance often leads to an unjustified presumption of explanatory reductionism (“*Self-interest explains reciprocity and concern with reputation*”) and empirically suspect *ad hoc* theorizing concerning hedonistic motives (“*Maybe there are psychic incentives*”). The problem is that clearly empirical hypotheses are treated as conceptual implications of the rational choice theory. The wrong ideal of natural order makes substantial empirical possibilities invisible to a researcher. Thus, one implication of the explanatory privilege of rational self-interest is that it makes the role of norms in the generation of selfish behaviour conceptually impossible. This is silly. Even if we end up rejecting Miller’s hypothesis on the causal role of the norm of self-interest in the generation of action, we should do it on the basis of empirical evidence, not based on *a priori* armchair considerations.

This is the core of my interest in the norm of self-interest. However, it also has some additional attractions. For example, this norm also helps to highlight some important biases in the recent rational choice-inspired work on social norms. The first lesson is that we should avoid the use of social norms as an *ad hoc* explanatory resource that is called in when self-interest cannot explain behaviour. Consider the notion of social preferences that is employed by experimental economists who define social preferences as:

“... motives such as altruism, reciprocity, intrinsic pleasure in helping others, inequity aversion, ethical commitments, and other motives that induce people to help others more than would an own-material-payoff maximizing individual.” (Bowles and Polania-Reyes 2012, 370)

Here, social preferences are defined by a violation of the assumption of self-interested action. Thus, social preferences are used to explain the deviance from the ideal of rational self-interested action. The operating question is “What must be added to self-interest to explain people’s choices?” However, social preferences do not reflect a coherent set of psychological motives or mechanisms. The idea is a mere placeholder for something substantial.

The second lesson is related to economists’ assumption that social norms are always *prosocial*. This is a surprising *functionalist* commitment of much rational choice theorizing. Although social norms may solve collective action problems, they do not have to. Social norms also do not have to be optimal in any way. In this context, the norm of self-interest illustrates how social norms can be regarded as a general social mechanism that contributes to the explanation of all human behaviour, including self-interest. The norms do not have to be prosocial or beneficial. Of course, norms do not have to be involved in everything, but they could be an important part of the story regarding our selfish behaviour.

Finally, we can also dispense with much futile philosophizing on ultimate human motives. We can accept as a baseline the pluralistic view that people have hedonistic, self-interested, prosocial and altruistic motivational impulses. The strength and prevalence of these motives is an empirical question, not something that is resolved by folk-psychological intuitions. Similarly, how (and whether) these impulses are turned into

consistent long-term behavioural patterns is an interesting social scientific question, not an explanatory basis for all thinking about human action. If we clear away the speculative psychology that underlies rational choice theory, we can start to study empirically how our competence in rational action is developed and sustained. In this view, rationality is not a natural state that we deviate from when we are children, suffering from mental problems, drunk, or tired. The consistent pursuit of long-term self-interest is far from a trivial achievement. It requires fighting short-term selfish impulses, ranking one's various goals, reasoning strategically, using cognitive tools, and having a predictable institutional environment. The emergence and development of rational and self-interested behaviour can be studied at various time scales (i.e., evolutionary, historical, and individual life-course), and something similar to the norm of self-interest may presumably have a role to play in each time scale. It combines the psychological development of an individual (i.e., the development of cognitive abilities, increasing impulse control, the development of the ability to construct folk-psychological narratives) and social settings (i.e., social norms, practices of self-presentation and justifications of action, institutions) in a way that allows us to consider how our ideas regarding human motivation could be different and how they could make a difference in the social life that we live (Ylikoski and Kuorikoski 2016).

Before discussing the details concerning the norm of self-interest, a brief clarification is in order. Although I have criticized rational choice, it is important to consider that the norm of self-interest does not challenge rational choice theory as a whole but only particular psychological interpretations of it. Many social scientists employ rational choice theory as a loose scheme to analyse social situations. In this "logic of situation" approach (Hedström, Swedberg and Udéhn 1998, Bengtsson and Hertting 2014), what is crucial is the analysis of social situations in terms of the constraints and opportunities that they provide to agents to pursue their goals. One can also discuss "interest analysis" (Ylikoski 2001). In this approach that was inspired by Max Weber and Karl Popper, *recurring social situations* are analysed from the point of view of their favourability to an agent's goals. In this approach, all discussion regarding interests is built on assumptions concerning an agent's goals. There is no need to assume that these goals are by definition self-regarding. Similarly, the relevant rationality assumption is that agents can learn to determine their "interests" in a situation. This approach provides a useful conceptual tool for social analysis. The existence of the norm of self-interest does not challenge the viability of this approach; on the contrary, it amplifies the approach's applicability because it narrows down the set of possible goals that agents can have and incentivizes them to pay attention to their situation.

### **How does the norm of self-interest work?**

The basic element of Miller's theory is the idea of a folk theory about motivation. According to this theory, monetary incentives strongly motivate people, which is also reflected in their attitudes. The distinctiveness of this theory is that it mostly involves the motivations of others. People believe that self-interest is a strong motivator for other people even in situations where they think that their own attitudes and behaviour are not influenced by it (Miller and Ratner 1998). For example, people believe that offering compensation for

blood donation will significantly increase the number of volunteers but that this incentive will have very little effect on their personal behaviour. Therefore, we have here a case of pluralistic ignorance: people do not regard their own non-selfish motives as evidence for the belief that other people could be similarly motivated. People thus overestimate how much other people are motivated by their material interests. According to Miller, self-interest is actually a much weaker predictor of people's attitudes and actions than this intuitive theory suggests.

Kruger and Gilovich (1999) provide supporting evidence for this observation. They studied how people attribute responsibility and how they expect others to attribute it. They found that people expected others (spouses, team members, etc.) to take more credit for desirable than undesirable outcomes, but this did not in fact occur. People were indeed egocentrically biased in their attributions, but they were not cynically biased as the folk theory predicts. The participants were thus not that self-interested as agents, but they were rather ardent self-interest theorists. The fact that people overestimate their responsibility for both positive and negative outcomes suggests that the source of the bias is caused by asymmetries in the observability and recall of behaviours, not by the motivational factors that are presumed by the folk theory. In any case, this type of naïve cynicism is a key component of the norm of self-interest.

What makes people adopt this strange theory of motivation if it is not based on their own experience? According to Miller, people adopt this theory similarly to the way that they adopt other forms of social knowledge, namely, through instruction and experience. Because of a lack of other types of research, Miller cites studies (Frank, Gilovich and Regan 1993) that argue that education in (neoclassical) microeconomics alters students' conceptions of the appropriateness of acting in a self-interested manner. The rational choice theory seems to persuade the students that the motivation behind people's actions, whatever it appears to be, is often blunt self-interest. Miller suggests that our everyday life is saturated with evidence that seems to support the belief in the importance of self-interested motivation. For example, numerous sayings ("Ethics pay") and slogans ("Ask yourself, are you better off today than you were four years ago?") communicate that the pursuit of self-interest is both prudent and good. Political ideologies may even suggest that self-interested action is the best way to achieve the collective good (e.g., the notion of the invisible hand).

The second important component of Miller's theory is the idea that this theory on human motivation can serve as a basis for a self-fulfilling prophesy in social action. Even if a person is willing to cooperate but believes that others will not cooperate, she will not cooperate if it conflicts with her self-interest. After all, what is the point of cooperation if the realization of the collective goal requires that most people contribute? This non-cooperative behaviour then provides evidence to other people concerning the lack of cooperativeness in the group. Although public goods games do not perfectly model collaborative situations such as this, they are close enough. Numerous experiments show that in repeated games, even the people who are hopeful and start by cooperating soon learn the futility of their attempts and give up cooperation. At least two factors are responsible this pattern. First, non-cooperation establishes itself as a descriptive norm. Thus, if people are sensitive to what most other people do and expect others to also be sensitive, they will find it appealing to conform to the norm. Second, being ready to

benefit other people when they do not reciprocate is a recipe for being exploited. Nobody wants to be a sucker – or a well-meaning fool – in the eyes of other people. The same observation applies to doing chores that are pointless if other people are not doing their part.

The folk theory of self-interested action also provides a basis for a self-fulfilling prophecy in institutional design (Miller 1999; Schwartz 2012). When you suspect that others are strongly motivated by their self-interests, the emphasis on institutional design becomes the avoidance of situations where dishonest individuals can exploit other people or the institution. Absent from the agenda in this mindset is developing and supporting arrangements that could help people to express and cultivate their non-selfish motivations. This absence is the key to a vicious circle: if most of the situations that we face are framed in terms of self-interested action, this behaviour seems natural, which, in turn, provides further confirmation for the intuitive theory of self-interested action.

When both expectations and institutional design provide support for the descriptive norm of self-interested action, it becomes easy to add a normative element to the belief: people start to believe that they *should* be selfish. Consider the justifications that people provide for their actions. This is an important dimension of norm-related behaviour that is not addressed in economic experiments. People are constantly providing accounts of their behaviour to both other people and themselves. The format and vocabulary of these accounts are based on culturally shared views concerning plausible and acceptable causes of action. Miller's thesis is that if people do not have a suitable vocabulary and explanatory schemes for their behaviour, they will find it difficult to account for this behaviour. How would they answer for the confusion that is created in other people by their unexplained behaviour? How can they set a limit on their non-selfish behaviour because other people may want to exploit it? These difficulties lead people to avoid behaviours that raise suspicions in other people. According to Miller, this is the case with non-selfish behaviours: the norm of self-interest influences what people find credible that, in turn, shapes people's behaviour.

Usually, people find the accounts that appeal to self-interested motives to be honest and credible, but non-selfish accounts are often treated with suspicion. After all, a person who cites self-interested motives does not have any reason to lie, while the claimed non-selfish motives may be deceptive and hide self-interested ulterior motives. Miller claims that this asymmetry provides important support for the norm of self-interest. Because people do not want to raise suspicions about their motivations – or they just do not want to make an extra effort to justify themselves – they will attempt to normalize their non-selfish behaviours by presenting them as self-interested. Thus, we may say that we want to help because it is fun or it is in our long-term interest (reputation gain) instead of simply saying that we want to help. Miller argues that this asymmetry constrains our unselfish behaviour: we tend to act on our non-selfish impulses only when we can find a self-interested frame to account for our behaviour. We feel that without a relevant interest, it is inappropriate for us to act on a public issue (Ratner and Miller 2001). Moreover, if we do not have a credible justification, we will abstain from the behaviour.

This idea is illustrated by Holmes, Miller and Lerner's (2002) concept of *the exchange fiction*. It is often observed that offering products to potential charity donors in exchange for their donations elicits more donations than appeals to charity alone. According to

Holmes and his associates, this observation occurs because donors can frame donations in terms of exchange. The fiction of exchange provides them an opportunity to act on their non-selfish impulses without having to reveal, or even to recognize, their real motive. The candles (and other gifts) that are received in exchange for donations provide a psychological cover for expressing compassion. The donor does not really need the candles, and they would not buy them from a store. Thus, receiving the gift provides very little extra motivation for donating; however, it *legitimizes* the act. It disinhibits charitable giving by providing a ready justification for it.

Ratner and Miller (2001) utilize the same idea to explain why people's interests and actions are more strongly correlated than their attitudes and interests. They argue that an absence of a stake or an interest hinders the actions that support political and social causes. Action without this justification is not regarded as appropriate or legitimate, because people believe that "it is not their place to" take a stance or to advance a cause. According to this view, people need a psychological standing on the issue, and if the influence of the norm of self-interest is strong, they cannot act without a vested interest. This requirement is reflected in many institutional practices. For example, to qualify as having legal standing for judicial review in the US legal system, a person must show that he or she has suffered or will suffer some injury, either economic or otherwise.

Accounts of action that do not match expectations create disbelief and suspicion, which can be regarded as strong negative sanctions. Who wishes to appear silly or open to exploitation? These categorizations affect both how we think about ourselves and how others think about us as potential collaborators. The strong influence of these labels provides an important path for the norm of self-interest to shape our behaviour. The folk theory concerning our selfish motivations denies other motivational factors as credible grounds for action. When other motives are not credible – or if making them credible requires considerable effort – we tend to avoid the behaviours that lack a self-oriented justification. This tendency is the self-fulfilling prophesy in action. Small-scale, non-selfish impulses may be tolerated or even praised, but in important matters, we regard such motivations with suspicion.

## Expanding and evaluating the idea

The previous section described Miller's theory as it appears in his main publications regarding the norm of self-interest (Miller 1999, Ratner and Miller 2001). Later, some studies attempted to test some elements of this theory, and some studies can also be used to expand it. I next briefly review some of these later developments.

We begin with the exchange fiction hypothesis. According to this hypothesis, a bargain-price product that is received in exchange for a donation creates an "exchange fiction" that permits the benefactors to justify their (larger) contributions as consistent with their self-interest. This fiction permits donors to act on their impulse to help without committing themselves to a difficult-to-live-up-to psychological contract of aiding a victim or other similar victims in the future. As Holmes, Miller and Lerner indicate, "the exchange fiction provides the mask under which the altruist can express her compassion and concerns with justice without having to reveal, or even recognize, her motives—after



all, she is merely engaging in an economic transaction” (2002, 145-146).

The empirical results that underlie this theory are confirmed by further studies. However, these studies challenge the exchange fiction explanation for the observations. Simpson, Irwin and Lawrence (2006) argue that the same observations can be explained more simply in terms of cognitive dissonance theory. They also provide a variant of the original experiments that seems to support dissonance rather than the exchange fiction explanation. Similarly, Briers, Pandelaere and Warlop (2007) argue that a gift may not necessarily work so much as a psychological cover but may instead provide information regarding the appropriate size of a donation. To support this claim, they devised experiments where a potential donor received a suggestion concerning the size of a donation rather than a gift, and the results were very similar to the original results by Holmes and his associates. These alternative explanations, and the empirical evidence that supports them, show that the norm of self-interest model is not the only possible explanation of the experimental results. However, as the authors of these papers admit, their results do not refute the idea of the norm of self-interest. Rather, their results show that there should be alternative tests for this theory.

Another element of Miller’s theory is the idea that people think that others are more selfishly motivated than themselves. Epley and Dunning (2000) challenge our confidence in this observation. They attempted to test whether people feel "holier than thou" because they harbour overly cynical views of their peers (but accurate impressions of themselves) or overly charitable views of themselves (and accurate impressions of their peers). They asked participants to predict how they and their peers would behave in situations with moral or altruistic overtones. These predictions were compared to the behaviour of the participants who actually faced the dilemma to determine whether self-serving comparisons were produced by mistaken predictions of one’s own or other people’s behaviour. The results suggest that people find it more difficult to accurately predict their own behaviour in moral dilemmas than to predict other people’s behaviour. The subjects systematically overestimated the influence of moral concerns on their own behaviour, while they were relatively accurate in predicting how others would behave. Interestingly, they also had similar difficulties in predicting the behaviour of particular (rather than generic) people. Thus, the bias here may not be egocentric but related to the difficulties in making case-based predictions. These results seem to challenge the claims by Miller and Ratner (1998). However, Epley and Dunning indicate that the results do not need to be contradictory. They argue that “whether people overestimate the impact of self-interest or predict it correctly will depend largely on whether the predicted situation involves hedonic consequences that are immediate or remote. In our studies, hedonic consequences were immediate” (Epley and Dunning 2000, 873). When hedonic concerns are more remote, such as in the studies by Miller and Ratner, self-interest may be less influential, and people may be more likely to overestimate its impact. This conclusion may be correct, but it is still prudent to avoid strong assumptions concerning the relations between attitudes and actual behaviours. People may overestimate the role of external incentives in other people’s behaviour (Heath 1999), but we should always pay attention to the context of attribution.

Another source of conflicting evidence is studies concerning the effects of economics education on students’ non-selfish behaviour. The original findings that the study of

game-theory fosters self-interest (Frank, Gilovich and Regan 1993) have led to an extensive cottage industry in economics (Frank, Gilovich and Regan 1996; Kirchgässner 2005; Rubinstein 2006; Pietro Cipriani, Lubian and Zago 2009; Etzioni 2015). Economists have used various methodologies and definitions of non-selfish behaviour. Often but not in all settings, economics students indeed stand apart from other students in terms of their selfish behaviour, but it still remains unclear whether this difference is due to the effects of economics instruction or due to the self-selection of selfish people into economics courses. It is possible that both mechanisms are at work. In any case, the interpretation of the results of these studies is quite difficult. Different studies operationalize non-selfish behaviour in incompatible ways, such as non-cooperation in social dilemmas, dishonesty, donating less to charities, etc. There is no reason to assume that all these measures tap into the same property, and it is also unclear regarding why the effects of economics education should be similar on all of these measures. Second, generalizations from experimental situations or survey answers to behaviour in real-life situations are tricky. It is plausible that economists' behaviour is as sensitive to framing and social context as other people's behaviour. Thus, it is difficult to claim that economists' behaviour is more selfish *sui generis*. Third, the actual effect of learning game-theory (or other economic theories) is unclear. Although learning game-theory may strengthen students' folk theory regarding the importance of selfish motivations, it could also make students just better at recognizing social dilemma situations or make them more capable of strategic reasoning. Although economists may also overestimate the prevalence of social dilemma situations, accept market outcomes as fair or associate selfishness with rationality, the actual mechanisms that underlie these effects require much more study. It is also crucial to recognize that despite all of these possible effects, economics students do not behave according to the predictions of rational choice theory. Thus, it is unclear that one can justify the claim that rational choice theory is self-fulfilling or performative as has been suggested in the literature (Ferraro, Pfeffer and Sutton 2005).

Now, let us discuss the ideas that provide additional support for Miller's theory. Three theoretical ideas can easily be used to strengthen the case for the norm of self-interest. A crucial element of Miller's theory was the wish to avoid being exploited by other people. The wish to avoid being exploited does not reduce to the anticipation of economic loss. Most crucially, we want to avoid the emotional consequences of being exploited by others. Vohs, Baumeister and Chin (2007) call this phenomenon *sugrophobia*. They characterize sugrophobia as a highly aversive feeling that motivates people to avoid it and to analyse what may have caused it to learn how to avoid it re-occurring. Being duped leads not only to anger towards the person(s) who is seen to be at fault but also to self-blame. Self-blame is a self-conscious emotion such as shame, guilt, pride, and embarrassment. Its intensity can vary among individuals, but it can be highly motivating (Effron and Miller 2011). Self-blame is an experience that a person is very eager to avoid experiencing again; thus, it raises the threshold for future cooperation. An analysis of sugrophobia can easily be used to partly explain the mechanisms that support the norm of self-interest.

The second additional mechanism comes from the research in economics and psychology: *motivational crowding out* (Frey and Jegen 2001; Gneezy, Meier and Rey-Biel 2011; Bowles and Polania-Reyes 2012; Festré and Garrouste 2015). The usual assumption in economics is that non-economic motives are complements of economic preferences. This

means that non-economic motives remain constant when economic incentives change because they are completely independent from economic motives. This constancy justifies disregarding non-economic concerns in economic analysis; the economist is interested in how changes in incentives (or information) change behaviour, and he does not have to address other possible motives because their effects can be assumed to remain stable. The research on the crowding out (and crowding in) effects challenges this assumption. The crowding out effect can be defined as a weakening or reduction of the influence of non-economic motives due to the use of economic incentives. This idea is highly compatible with the norm of self-interest. The difference is that in the crowding out effect, the focus is on how incentives change the way that people frame the situation in the relative local context, not people's general ideas concerning human motivation. The idea is that situational factors determine the motives (or impulses) on which people act. People react to both the existence and the size of incentives. The incentives provide information on the goals and beliefs of the principal who provides the incentives. If the motives of the principal seem to be selfish, this influences the moral framing of the situation. Similarly, the framing of a situation as market interaction easily creates a detachment from moral concerns and a strong focus on material gain. Naturally, these changes can also cause endogenous changes in people's motivations. Thus, the extensive (and exclusive) use of external incentives may influence the prevalence of non-material motives in the community.

All of this is highly compatible with the basic ideas of the norm of self-interest theory. The postulated effects are similar, and one could say that the above analysis of the crowding out effects highlights the importance of social norms. The original focus of Miller's theory is probably too centred on people's theorizing about ultimate human motives rather than the more situation-specific effects of the expectations about proper behaviour. The idea of *homo economicus* may have a role in the justification of incentive-based practices, and it may make some people blind to the influence of non-selfish motives, but its greatest impact may be through situational factors that influence what types of concerns are treated as appropriate, prudent, or legitimate in particular social situations. Of course, the crowding out research has not really focused yet on an analysis of these norms, but this is a promising avenue for future research, for example, when analysing how the neoliberal political discourse on New Public Management influences social institutions and organizations.

Third, an additional mechanism that supports the norm of self-interest could be experience sampling bias (Denrell 2005). In *experience sampling bias*, people interact with one another depending on their evaluations of potential partners. Negative experiences decrease the probability of interaction, and the evaluations of negatively evaluated individuals will thus not be updated based on new experiences. This fact implies that negative initial impressions are more stable than positive impressions that will be updated with further interactions. A consequence of these processes is that people tend to overestimate both the frequency and the strength of negative attributes.

As Fetchenhauer and Dunning (2010) suggest, this fact may explain the often-confirmed observation that in trust-games, people severely underestimate other participants' trustworthiness. Typically, 80% to 90% of trustees honour trust, but trusters on average estimate that this rate will be only 45% to 60%. This cynicism has real consequences,

because people refrain from trusting and pass up potentially very beneficial opportunities. The experience sampling explanation for this fact is when people trust and this trust is exploited, they become more cynical. After all, the experience of being exploited is quite motivating. However, mistakenly failing to trust a trustworthy person does not provide vivid feedback because one does not know what could have occurred. Because these mistakes are not corrected, the overall estimation of the trustworthiness of other people will remain biased. Asymmetric feedback can thus explain why people overestimate the role of self-interest in human motivation. People tend to be exposed to instances of surprising self-interest (e.g., when they trust and this trust is violated), but they rarely face instances of surprising selflessness.

The idea of experience sampling bias can be expanded. In the above simple case, the bias is driven by the ability of people to choose with whom to interact. However, we can have similar dynamics in cases where the agents cannot choose with whom to interact, but they can influence the form of this interaction (Denrell 2005, 967-968). Therefore, if a person cannot avoid the interaction, then she may be able to use some safeguards. For example, formal contracts and other types of incentives that protect against the adverse consequences of opportunistic behaviour are quite common in real life outside behavioural economics labs. These safeguards typically make it probable that individuals will be punished for opportunistic behaviour, which implies that it is in their self-interest to behave as if they were trustworthy. As a consequence, the absence of opportunistic behaviour does not provide much evidence regarding the real trustworthiness of these people. For example, a binding contract may lead partners to attribute cooperative behaviour to the restrictions that are imposed by the contract rather than to the non-selfish motivations of the participants. This attribution implies that the initial impression of untrustworthiness will not be revised considering the observed cooperative behaviour. Consequently, the cynical view of human motivation remains unchallenged. Only in social settings in which safeguards are not used or considered to be inappropriate, there is a possibility to revise one's false beliefs concerning untrustworthiness. Maybe this is a partial explanation for the higher or at least more accurate levels of trust in families and some other social settings where individuals can neither avoid trusting others nor use formal safeguards.

What makes experience sampling bias theoretically interesting is that it is based on a social rather than a cognitive mechanism. Thus, even if people were fully rational in their belief formation, they would still be susceptible to this bias. However, there is no reason to see social mechanisms, such as experience sampling bias, as competing with cognitive mechanisms. For example, in the case of the folk theory about human motivation, the social mechanisms tend to amplify one another. The *attributional cynicism* that is observed by Critcher and Dunning (2011) is an example. Although the experience sampling bias affects what type of evidence is available to people, this cognitive mechanism describes how people respond to this evidence. The claim by Critcher and Dunning is that people treat the evidence of selfless and selfish acts differently. Selfish behaviour is usually taken at face value, while evidence of apparently unselfish behaviour is often reinterpreted so that the belief in the norm of self-interest is not challenged. Given the ambiguity of the notion of self-interest, it is usually quite easy to construct a scenario that would make only the unselfishness of the behaviour apparent. This construction makes it possible

to “save” the belief in the norm of self-interest from contrary evidence. Thus, one can stubbornly hold on to one’s beliefs and still consider oneself rational.

The general point here is that if people are victims of both experience sampling bias and attributional cynicism, these mechanisms would amplify one another. Even if these mechanisms are individually insufficient to support the beliefs that underlie the norm of self-interest, they may be able to do it jointly. Moreover, if there are many such mechanisms and only very few counteracting mechanisms, eradicating the norm of self-interest may be very difficult. Well-designed interventions may well reduce the influence of one mechanism, but it is improbable that they would do the same for all relevant mechanisms. Cutting only several threads in a rope may be insufficient to make it break.

### Some ambiguities

Thus far, I have mostly discussed the norm of self-interest as an interesting theoretical idea. *Even if it is wrong, this discussion has important theoretical implications.* It shows that we should not define social norms as beneficial or prosocial. It also suggests that we should not treat self-interested action as self-explanatory. Finally, it demonstrates that we should treat human motivation as an empirical question rather than as a conceptual consequence of rational choice theory. Before finishing my discussion, I want to address an important feature of Miller’s hypothesis that affects its empirical testing: ambiguity.

Miller suggests that the norm of self-interest “exists in Western cultures” but does not say much more. Does “Western” here refer to the US, North America, or maybe a combination of North America and Europe? Alternatively, does the norm exist only among US students, who are the principal subjects of the studies that Miller cites in support of his theory? Regardless, what does “exists” mean in this context? Does it imply that the norm does not exist elsewhere or does it just say that the norm exists at least in Western cultures? Given the little comparative evidence that we have, the interpretation that the norm exists at least in Western cultures is much safer. This interpretation allows us to be somewhat more flexible with the meaning of Western culture. Thus, it is possible that the force of the norm is even stronger somewhere else. This may be an advantage, because recent cross-cultural studies seem to suggest that Westerners are in fact the most prosocial among human cultures, at least when measured in terms of behaviour in anonymous game-theoretical settings (Henrich et al. 2005). Perhaps Miller is wrong in assuming that the norm of self-interest is strongest in advanced capitalist societies (if this is what he means by Western culture).

We should also ask what it means for the norm of self-interest to *not* exist in some communities. Is it possible that in some communities, there is a contrary norm that includes an expectation of unselfish behaviour and a general suspicion regarding selfish motives? This idea sounds too fanciful. I think a better way to consider the norm of self-interest is to regard its existence as a matter of degree. We should thus abandon the binary question concerning existence and consider how strong the norm of self-interest is in a given community or social situation. Here, we would probably find multiple dimensions, such as how strongly people believe in the theory on selfish human motivation, in which (and how many) contexts the norm affects people’s behaviour, how many of the

supporting mechanisms are in place, and how rich (or impoverished) is the community members' vocabulary of justification? Establishing these dimensions of measurement requires considerable work, but it makes more sense than fantasizing about opposite norms.

These measures of the norm of self-interest also suggest some predictions. Although the theory is too complex to be tested at once, we can test its elements. Thus, we should predict that if the theory is correct, we should see, for example, an *increase in non-selfish behaviour* as a consequence of the following experimental interventions:

- increase people' belief in the existence and importance of non-selfish motives;
- reduce sugrophobia by decreasing the self-blame that is involved;
- expand the credible vocabulary for justifying non-selfish actions;
- expand the basis for taking a psychological stand on social issues; and
- reduce experience sampling bias by reducing agents' freedom to choose with whom to interact based on a short or non-existent interaction history.

The problem is that these are very difficult experimental interventions. Furthermore, given the overlapping mechanisms that underlie the norm, the working of other mechanisms may prevent the predicted effects of the interventions.

## Conclusion

In this paper, I attempted to provide a systematic discussion of Dale Miller's idea of the norm of self-interest. Although an empirical evaluation of the theory is still underway, I found it interesting from a theoretical point of view. First, in contrast to much of the current theorizing about norms in rational choice theory, this theory does not regard norm-influenced behaviour as a residual category that includes only the cases that cannot be rationalized by self-interested motives. Rather than treating the influence from social norms as an *ad hoc* explanatory resource, self-interest is regarded as (one) basic mechanism that influences social behaviour irrespective of whether it is "rational". Naturally, self-interest is only one mechanism that influences social behaviour. Furthermore, there is no reason to suppose that the norm of self-interest is the full story concerning self-interested behaviour. However, the crucial matter is that no *a priori* ideal of natural order limits its explanatory potential.

The second theoretically interesting feature is that the theory does not presuppose any fundamental theory of human motivation. The discussion regarding "ultimate motivations" may not even make sense. Although many people seem to take seriously the idea that there are ultimate motives of human behaviour, we should not automatically assume that this type of theorizing has many scientific credentials. While it is an undeniable part of our folk psychological practice to consider hierarchical relations among reasons and to regard the motivations that are strong, ever-present, or general as more "fundamental", it is not obvious that the postulated "ultimate motives" are psychologically real (Ylikoski

and Kuorikoski 2016). In any case, it is very doubtful that such ultimate springs of action can be discovered by armchair speculation or by simple economic experiments. Thus, there is no reason to take ultimate motives as a default assumption. A more plausible research strategy would be to follow empirical findings without prejudice and at least tentatively assume that people are prone to act on all types of impulses, whether they are rational or irrational, selfish or non-selfish, or something that cannot be described by using these terms. However, I hope that I have made a case for studying the possibility that people appear to behave selfishly because they are sensitive to the norm of self-interest.<sup>1</sup>

---

<sup>1</sup>This research has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) / ERC grant agreement no. 324233, Riksbankens Jubileumsfond (DNR M12-0301:1) and the Swedish Research Council (DNR 445-2013-7681 and DNR 340-2013-5460).

## References

- Bengtsson, B. and Hertting, N. (2014). Generalization by mechanism: Thin rationality and ideal-type analysis in case study research. *Philosophy of the Social Sciences*, 44(6):707–732.
- Bowles, S. and Polania-Reyes, S. (2012). Economic incentives and social preferences: substitutes or complements? *Journal of Economic Literature*, 50(2):368–425.
- Briers, B., Pandelaere, M., and Warlop, L. (2007). Adding exchange to charity: a reference price explanation. *Journal of Economic Psychology*, 28(1):15–30.
- Cipriani, G. P., Lubian, D., and Zago, A. (2009). Natural born economists? *Journal of Economic Psychology*, 30(3):455–468.
- Coleman, J. and Fararo, T. (1992). Introduction. In Coleman, J. and Fararo, T., editors, *Rational choice theory: advocacy and critique*. Sage, Thousand Oaks.
- Critcher, C. R. and Dunning, D. (2011). No good deed goes unquestioned: Cynical reconstruals maintain belief in the power of self-interest. *Journal of Experimental Social Psychology*, 47(6):1207–1213.
- Denrell, J. (2005). Why most people disapprove of me: experience sampling in impression formation. *Psychological review*, 112(4):951.
- Effron, D. A. and Miller, D. T. (2011). Reducing exposure to trust-related risks to avoid self-blame. *Personality and Social Psychology Bulletin*, 37(2):181–192.
- Epley, N. and Dunning, D. (2000). Feeling "holier than thou": are self-serving assessments produced by errors in self-or social prediction? *Journal of personality and social psychology*, 79(6):861.
- Etzioni, A. (2015). The moral effects of economic teaching. In *Sociological Forum*, volume 30, pages 228–233. Wiley Online Library.
- Ferraro, F., Pfeffer, J., and Sutton, R. I. (2005). Economics language and assumptions: How theories can become self-fulfilling. *Academy of Management review*, 30(1):8–24.
- Festré, A. and Garrouste, P. (2015). Theory and evidence in psychology and economics about motivation crowding out: A possible convergence? *Journal of Economic Surveys*, 29(2):339–356.
- Fetchenhauer, D. and Dunning, D. (2010). Why so cynical? Asymmetric feedback underlies misguided skepticism regarding the trustworthiness of others. *Psychological Science*, 21(2):189–193.
- Frank, R. H., Gilovich, T., and Regan, D. T. (1993). Does studying economics inhibit cooperation? *Journal of Economic Perspectives*, 7(2):159–171.
- Frank, R. H., Gilovich, T. D., and Regan, D. T. (1996). Do economists make bad citizens? *Journal of Economic Perspectives*, 10(1):187–192.



- Frey, B. S. and Jegen, R. (2001). Motivation crowding theory. *Journal of economic surveys*, 15(5):589–611.
- Gneezy, U., Meier, S., and Rey-Biel, P. (2011). When and why incentives (don't) work to modify behavior. *Journal of Economic Perspectives*, 25(4):191–210.
- Heath, C. (1999). On the social psychology of agency relationships: Lay theories of motivation overemphasize extrinsic incentives. *Organizational Behavior and Human Decision Processes*, 78(1):25–62.
- Hedström, P., Swedberg, R., and Udéhn, L. (1998). Popper's situational analysis and contemporary sociology. *Philosophy of the Social Sciences*, 28(3):339–364.
- Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H., McElreath, R., Alvard, M., Barr, A., and Ensminger, J. (2005). "Economic man" in cross-cultural perspective: Behavioral experiments in 15 small-scale societies. *Behavioral and Brain Sciences*, 28(6):795–815.
- Holmes, J. G., Miller, D. T., and Lerner, M. J. (2002). Committing altruism under the cloak of self-interest: The exchange fiction. *Journal of Experimental Social Psychology*, 38(2):144–151.
- Kim, A. (2014). The curious case of self-interest: Inconsistent effects and ambivalence toward a widely accepted construct. *Journal for the Theory of Social Behaviour*, 44(1):99–122.
- Kirchgässner, G. (2005). (Why) are economists different? *European Journal of Political Economy*, 21(3):543–562.
- Kruger, J. and Gilovich, T. (1999). "Naive cynicism" in everyday theories of responsibility assessment: On biased assumptions of bias. *Journal of Personality and Social Psychology*, 76(5):743.
- Miller, D. T. (1999). The norm of self-interest. *American Psychologist*, 54(12):1053–1060.
- Miller, D. T. and Ratner, R. K. (1998). The disparity between the actual and assumed power of self-interest. *Journal of Personality and Social Psychology*, 74(1):53.
- Ratner, R. and Miller, D. (2001). The norm of self-interest and its effects on social action. *Journal of Personality and Social Psychology*, 81(1):5–16.
- Rubinstein, A. (2006). A sceptic's comment on the study of economics. *The Economic Journal*, 116(510):C1–C9.
- Schwartz, B. (2012). Crowding out morality: How the ideology of self-interest can be self-fulfilling. In Hanson, J., editor, *Ideology, psychology, and law*, pages 160–184. Oxford University Press, Oxford.
- Simpson, B., Irwin, K., and Lawrence, P. (2006). Does a "norm of self-interest" discourage prosocial behavior? Rationality and quid pro quo in charitable giving. *Social Psychology Quarterly*, 69(3):296–306.

- Toulmin, S. (1963). *Foresight and understanding*. Hutchinson, London.
- Vohs, K. D., Baumeister, R. F., and Chin, J. (2007). Feeling duped: Emotional, motivational, and cognitive aspects of being exploited by others. *Review of General Psychology*, 11(2):127–141.
- Ylikoski, P. (2001). *Understanding interests and causal explanation*. Ph.D. thesis, University of Helsinki. Available from <http://ethesis.helsinki.fi/julkaisut/val/kayta/vk/ylikoski/>.
- Ylikoski, P. (2012). Micro, macro, and mechanisms. In Kincaid, H., editor, *The Oxford handbook of philosophy of the social sciences*, pages 21–45. Oxford University Press, Oxford.
- Ylikoski, P. and Kuorikoski, J. (2016). Self-interest, norms, and explanation. In Risjord, M., editor, *Normativity and naturalism in the philosophy of the social sciences*, pages 212–229. Routledge, New York.