Accepted for publication in *Nature Human Behaviour* published by Springer Nature.

Extrinsic rewards and crowding-out of prosocial behaviour

Conny E. Wollbrant[1]*, Mikael Knutsson[2], Peter Martinsson[3,4]

[1] Stirling Management School, Economics Division, Behavioural Science Centre, University of Stirling, Stirling FK9 4LA, Scotland UK

[2] Department of Management and Engineering, Linköping University, 581 83 Linköping, Sweden

[3] Department of Economics, University of Gothenburg, 405 30 Gothenburg, Sweden

[4] Department of Technology, Management and Economics, Technical University of Denmark, 2800 Kongens Lyngby, Denmark

* Corresponding author: Conny E. Wollbrant (conny.wollbrant@stir.ac.uk)

Abstract

The law of supply is a fundamental principle of economics and states that any increase in price will increase the quantity supplied. In the case of prosocial behaviour, however, increasing rewards have reduced supply, posing a challenge to standard economic theory. Attempts to study such "crowding-out" have been limited by their small scale and inherent difficulties posed by calibration of experimental tests. We analyse a large-scale natural experiment in the environmental domain consisting of 20,370 independent observations derived from aggregation of approximately 27 million individual decisions. We find that aggregate supply of prosocial behaviour is "s-shaped", demonstrating how attempts to increase prosocial behaviour using monetary rewards can be counter-productive. Our study shows that results derived from a small set of data points collected from an underlying s-shaped data-generating process is vulnerable to misinterpretation, and that proxy measures of intrinsic motivation ought to be collected to ensure theoretical advance.

Prosocial behaviour among humans is commonplace and permeates every aspect of our lives. We donate blood and money to charities and contribute to public goods in the form of tax payments, teamwork, and recycling (1–3). While prosocial behaviours are ubiquitous, they are often under-supplied and fall short of the socially optimal level (4, 5). This contributes to a range of contemporary global problems such as water shortages, antimicrobial resistance and depletion of natural resources (6–9). If such shortcomings are due to insufficient incentives, the law of supply from standard microeconomic theory offers a simple solution: any subsidy, tax-break or extrinsic monetary reward ought to increase prosocial behaviours. If, however, extrinsic incentives crowd-out existing intrinsic prosocial motivations, such rewards can cause a net reduction of prosocial behaviour, implying limitations of the price mechanism for improving social outcomes (5, 10-12).

Previous research on crowding-out of prosocial behaviour has yielded mixed results. On the one hand, crowding-out has been documented in a range of domains (2, 11–15). In the context of blood donations, Richard Titmuss first suggested that explicit extrinsic rewards might be counter-productive (16), a prediction later corroborated among women (but not men) in a field experiment (17). Fines have also increased undesirable behaviours such as the length of hospital stays (15), increased late kindergarten pick-ups (18), and reduced cooperation (2, 19–21). Moreover, evidence from brain scans shows that monetary sanctions encourage neural activity associated with self-interested decision-making and mitigate activity associated with social reward evaluation (22, 23). On the other hand, a number of studies have failed to detect any crowding-out effect. Examples include lottery and standard incentives, which have also succeeded in increasing blood donations (24–26) and small gifts that have successfully reduced paper waste (27). Overall, the literature on motivational crowding-out is inconclusive.

This paper investigates the effect of an exogenous increase in extrinsic reward in the prosocial domain of recycling. We analyse a unique natural experiment resulting from the Swedish Central Bank's (Sveriges Riksbank) decision to withdraw from circulation the SEK 0.5 coin (approximately USD 0.07). From the 1st of October 2010, when instead the SEK 1 coin became the coin of lowest denomination, recyclable beverage cans carrying a deposit of SEK 0.5 were gradually replaced by cans carrying an SEK 1 deposit (approximately USD 0.14). As a result, per can expected reward from recycling gradually

increased from SEK 0.5 in September 2010 and finally converged on SEK 1 in January 2014. For the average recycling customer, returning 18 cans, this meant an increase in expected reward from SEK 9 to 18 (approximately USD 1.2 to 2.5) over the period. This gradual change allows us to estimate the supply of recycling over the whole range of values between SEK 0.5 and SEK 1 as a function of expected reward, and the exogenous decision by the Swedish Central Bank to remove the SEK 0.5 coin allows us to interpret this as a natural experiment (28).

Data on recycling and deposits are collected and stored by the recycling company Returpack. The company provided us with a balanced panel of monthly aggregated recycling data for a period of 42 months, spanning 485 stores of one of Sweden's largest chain stores. We complemented this with census data on median income from Statistics Sweden (Statistiska Centralbyrån) and voting data from the Swedish Election Authority (Valmyndigheten), measured at the postcode and municipal level, respectively. The panel spans the whole period, August 2010 to January 2014, during which the increase in expected reward occurred. The analysed data set contains 20,370 independent observations of aggregate monthly recycling quantities measured at the store level, corresponding to approximately 27 million individual recycling decisions.

In line with motivational crowding theories, we find support for an s-shaped aggregate supply curve of recycling. When expected reward increases, the quantity of recycled cans first increases up to an inflection point, after which it decreases, and the supply curve becomes negatively-sloped. This decrease in recycling continues until it reaches another inflection point, after which the supply curve once again becomes positively-sloped, giving rise to the s-shaped graph of the supply curve (14, 29). We also find that the intercept and the two inflection points on the estimated relationship depends on our proxy measure for intrinsic environmental motivation. Since the crowding-out effect is dependent on pre-existing levels of prosocial motivation, this provides additional support for motivational crowding theories. This dependency, combined with the s-shaped supply curve pattern, can help reconcile the mixed literature on motivational crowding-out.

We exploit data generated by a large-scale natural experiment (28) to examine motivational crowding-out by evaluating the effect of exogenously induced variation in reward on the supply of prosocial behaviour. Although past research has shown that introducing extrinsic rewards to promote prosocial behaviour can be counter-productive in a range of domains (14, 30), our study differs from past research in several important ways. First, past studies are only able to collect a few data points on the underlying relationship between reward and prosocial behaviour (12, 15, 31), which can lead to misinterpretation of empirical results, either as support for standard economic theory or for an alternative (but incorrect) crowding-out theory. This data enables us to overcome this limitation, potentially reconciling the mixed results emerging from the literature, and substantially constrain the set of plausible crowding-out theories. Second, previous studies have typically not complemented their behavioural data with any proxy measure of intrinsic motivation level, which precludes testing the dependency of the crowding-out effect on motivation levels (for an exception, see (2)). Here we use the share of votes for the Green Party as a measure of willingness to participate in pro-environmental activities. Third, participants in our study are unaware that they are participating in an experiment, in contrast to what is the case in the lab, which is a methodological advantage when studying social behaviour (28). Finally, many studies are significantly smaller and may not constitute

representative samples of the studied population, which is important for external validity and replicability of estimated results (28, 32, 33).

In addition to improving our understanding of prosocial behaviour and informing policies intended to promote social outcomes, our study has implications for economic models of individual decision making. In most models, when crowding-out occurs, it is because intrinsic motivation has vanished and is unlikely to be restored (5, 34). Alternatively, any positive amount of extrinsic reward will result in a drastic reduction in the supply of prosocial behaviour (22, 35, 36). Our study suggests that the process of crowding-out prosocial behaviour is significantly more gradual, and that positive but relatively modest rewards can improve aggregate social outcomes at a lower cost. Future efforts ought to be directed towards understanding how such opportunities could be harnessed.

Results

We calculate monthly expected reward per recycled can as the weighted mean of recycled SEK 0.5 and SEK 1 deposit cans. Since distributions of cans with different deposits differ across store location in a given month, so does expected reward. Figure 1 illustrates the gradual increase in expected reward $E[r_t]$ and average quantity recycled over time, across all stores in the sample, by month. The median of expected reward gradually increases from SEK 0.5 and converges on SEK 1, since restocking with cans carrying an SEK 1 deposit depends on individual store sales and producer stock. As expected reward differs across stores, both within and over time, it induces the necessary variation to examine the continuous relationship between reward and behaviour. Naturally, expected reward distributions are narrower at the beginning and end of the study period as most cans are carrying a deposit of either SEK 0.5 or SEK 1 respectively, while the transition period is characterised by wider reward distributions (see also Extended Data Figure 1). Store level quantity recycled ranges from 19,751 to 36,944 and tracks broad beverage consumption patterns, exhibiting highs in summer months and lows in winter months (see also Extended Data Figure 2). Overall, both measures follow expected patterns.

Fig. 1. about here

To assess the relationship between expected reward and prosocial behaviour, we combine recycled quantity and expected reward in the scatter plot presented in Figure 2, representing the aggregate supply curve. Visual inspection reveals a pattern inconsistent with standard economic theory, which predicts a monotonically increasing relationship between expected reward (price) and average quantity recycled (quantity supplied). Instead, this figure reveals a negatively-sloped segment on the supply curve which begins when expected reward approximately equals SEK 0.7. This segment remains negatively-sloped until expected reward approximately equals SEK 0.85, after which the supply curve appears positively-sloped once more.

To formally test for crowding-out effects, we turn to statistical analysis of the relationship (details in Methods). Using the Random Effects Generalised Least Squares technique, we begin by estimating quantity of recycled cans as a function of expected reward levels in t-1 (see Extended Data Figure 2

for box plots of Cans). Dummy variables for expected reward levels are generated by dividing the continuous reward variable into five equal intervals, which we denote L1-L5 in rising order; L1 corresponds to reward in the SEK 0.5-0.6 range, L2 to the SEK 0.6-0.7 range [0.6-0.7), and so on. Each dummy variable takes value 1 if expected reward is within the corresponding range, otherwise they equal 0. This approach enables a direct test of the law of supply without making assumptions about functional form otherwise necessary for parametric estimation. Since standard economic theory predicts a monotonically increasing relationship between expected reward and quantity recycled, this implies a pattern of strictly increasing coefficients.

Table 1 about here

Table 1 reveals a non-monotonic pattern of coefficients, inconsistent with the law of supply. Crucially, an increase in expected reward level from L2 [0.6-0.7) to L3 [0.7-0.8) leads to a reduction in quantity recycled as the coefficient on L3 [0.7-0.8) is significantly smaller than the coefficient on L2 [0.6-0.7) (L2 [0.6-0.7) vs. L3 [0.7-0.8): Chi2(1) = 30.43, $P < 0.001$, two-tailed, 95% Confidence Intervals = [2.0 3.2] vs. [0.2 1.5]). Further increases in expected reward levels lead to increases in quantity recycled relative to L3 [0.7-0.8). While all estimated coefficients are significantly different from baseline (see Table 1), the coefficient on L5 [0.9-1.0] is significantly larger than the coefficient on L3 [0.7-0.8) (L3 [0.7-0.8) vs. L5 [0.9-1.0]: Chi2(1) = 22.33, $P < 0.001$, two-tailed, 95% Confidence Intervals = [0.2 1.5] vs. [1.2 2.9]). This overall pattern suggests a non-linear, and non-monotonic relationship between expected reward and quantity recycled, consistent with the pattern in Figure 2. We summarise these results below.

Result 1: The relationship between quantity recycled and expected reward is non-monotonic, violating the law of supply.

We now estimate quantity recycled as a polynomial function of expected reward and present our results in Table 2. Specifically, we use a linear and quadratic term and in model 1 to allow for a single inflection point on the relationship, and add a cubic term in model 2 to allow for two inflection points. This approach obviates the need to specify intervals of the expected reward variable. Since a key prediction of motivational crowding-theory is the income independent interaction between price and intrinsic motivation, we add measures of income and intrinsic environmental motivation (Green votes (%)) and their interactions with expected reward to models 1 and 2. We present these as models 3 and 4, respectively.

Table 2 about here

We find in model 1 that the coefficient on the linear term is positive and statistically significant (β = 15.8, $P < 0.001$, 95% Confidence Intervals = [7.9 23.7]) and that the coefficient on the quadratic term is negative and statistically significant (β = -7.7, $P = 0.004$, 95% Confidence Intervals = [-12.9 -2.5]). This pattern is consistent with a single inflection point and a "backward bending" supply curve. We further find in model 2, that the linear and quadratic terms remain statistically different from zero (β's = 379.5

and -503.3, *P*'s < 0.001, 95% Confidence Intervals = [208.5 550.4] and [-734.9 -271.6]) and of the same signs as in model 1. In addition, the coefficient on the cubic term is positive and statistically different from zero (β = 218.1, *P* < 0.001, 95% Confidence Intervals = [115.7 320.5]). This pattern is consistent with two inflection points on the estimated relationship and an "s-shaped" supply curve (11, 13). In addition, interactions between expected reward and our proxy measure for intrinsic motivation to participate in pro-environmental behaviour, Green votes (%), are significant when the cubic expected reward term is included. Importantly, this effect is present when controlling for income and income-reward interactions (see model 4 in Table 2). Since monthly dummy variables are included, these estimations control for the seasonal effects apparent in Figure 1 (see also Supplementary Tables 1 and 2 for additional robustness tests). Overall, regression analysis results are consistent with predictions of crowding-out theories.

Result 2: Signs on estimated coefficients on expected reward and their interactions with the intrinsic motivation measure are consistent with predictions of motivational crowding theories.

We now turn to analysis of post-estimation result based on model 4 in Table 2. Figure 3 (a) represents linear predictions of quantity recycled as functions of expected reward, for three levels of our intrinsic motivation measure, Green votes (%), corresponding to the 10^{th} percentile (2.54), the median (5.76) and the 90^{th} percentile (13.87) of the Green votes (%) distribution. Overall, a higher level of Green votes (%) is associated with a higher level of recycling for each expected reward level, consistent with our interpretation of the measure as a proxy for willingness to participate in pro-environmental behaviour. The negative segment on each curve represents the crowding-out effect, which is more pronounced for higher levels of Green votes (%). This suggests that the crowding-out effect is heterogeneous with respect to intrinsic motivation, in line with motivational crowding theories (11, 13).

Figure 3 about here

A sufficient condition for motivational crowding-out is a negative marginal effect of any price on quantity supplied. Figure 3 (b-d) presents average marginal effects of expected reward on quantity recycled, evaluated over the whole range of expected reward from 0.5 to 1, by each level of Green votes (%). For all levels, point estimates of marginal effects are initially positive, followed by a negative segment, and finally positive once more as expected reward increases over the range. Negative segments below the zero-line identifies a net negative crowding-out effect. In (c) and (d), Confidence Intervals indicate statistically significant and negative marginal effects, but not in (b) (see Source Data 3). Importantly, the crowding-out effect is dependent on pre-existing levels of intrinsic motivation. Our analysis of marginal effects therefore rejects the hypothesis embedded in the law of supply – positive marginal effects of price on quantity supplied for all prices – and provides evidence in favour of motivational crowding theory. We summarize these results below.

Result 3: Negative marginal effects of expected reward on quantity recycled provide direct evidence for motivational crowding-out. The crowding-out effect is heterogeneous with respect to pro-environmental motivation.

To illustrate the significance of these effects for policy, Table 3 presents the predicted global and local maximum and minimum of quantity recycled, and the maximal crowding-out effect. The latter is calculated as the difference between the local maximum and local minimum quantities for levels of Green votes (%) on the basis of model 4 in Table 2. Calculations of the monthly per store effect reveals that the crowding-out effect is significant. In absolute terms, it ranges from 451 to 2,359 cans at the 10^{th} percentile (2.54) and the 90^{th} percentile (13.87) levels of Green votes (%), respectively. Evaluated at the average level of Green votes (%), the crowding-out effect is 1,093 cans. Aggregated across stores in the study sample, this implies a crowding-out effect of 5,616,000 cans annually. In material terms, this is approximately equivalent to 1,123m^3 waste, weighing at least 81 tonnes, or 1.12mill kWh (37). Extrapolation to the whole Swedish market implies a crowding-out effect of 89mill cans annually, corresponding to 16,183$m^3$of waste, weighing at least 1,168 tonnes, or 16,18mill kWh. To recover from the crowding-out effect, expected reward would need to increase from 0.88 to 0.98, equivalent to an 11.36% increase. Overall, these results clearly illustrate the existence of limitations when using the relative price mechanisms for design of incentives intended to increase prosocial behaviour, both in terms of the desired effect and cost-efficiency of interventions.

Table 3 about here

Discussion

This study shows that providing extrinsic monetary rewards to increase prosocial behaviours can have both intended and unintended consequences, and the observed outcome will depend on the net balance of two opposing forces. On the one hand, the price effect of increasing extrinsic monetary rewards is always positive. On the other hand, increasing extrinsic monetary rewards can have a negative effect on intrinsic motivation to participate in prosocial behaviour, in this case recycling. If an extrinsic reward causes a sufficiently large reduction in intrinsic motivation, the net effect of reward on quantity supplied is negative, since the effect on intrinsic motivation dominates the price effect. The theoretical literature offers two main candidate explanations consistent with an s-shaped aggregate supply curve as observed in this study.

Signalling theory emphasises the use of behaviour to signal intrinsic motivations to oneself, or to others, when motivated by image concerns (11, 13). It explains the s-shape as the result of how the net balance of the positive price effect and the negative effect on signal value evolves as reward increases (11, 13). While the price effect is always positive, the presence of a reward casts doubt about the underlying reason for engaging in prosocial behaviour since it becomes harder to infer whether behaviour is motivated intrinsically or by the reward. Behaviour therefore becomes an increasingly noisy signal of motivation and its value gradually diminishes as rewards continue to increase. When reward begins to increase from a relatively low level, the price effect first dominates the negative signal value effect, leading to a net positive effect on supply. Although signal quality falls as reward

increases, the net effect of reward on supply remains positive as long as the price effects dominates. This accounts for the initial positively sloped segment on the supply curve. When reward reaches a sufficiently high level, however, behaviour becomes too noisy a signal of intrinsic motivation. The negative effect begins to dominate the price effect, causing a net negative effect of reward on supply. This accounts for the negative segment on the supply curve. As reward continues to increase and becomes sufficiently high, its negative effect on signal value becomes relatively low since signal quality has already deteriorated. At this point, the positive price effect will again dominate, accounting for the second positively sloped segment on the curve (31), completing the s-shape at the individual level. Since the mechanism generates an s-shape at the individual level, this theoretical prediction is also consistent with an s-shaped aggregate supply curve. In the context of our study, self-signalling is always possible, but a desire to signal good behaviour to others cannot be ruled out as recycling areas are usually served by more than one machine, and others may be waiting in line to recycle. Exploring the relative significance of each signalling source might provide a fruitful direction for future research on this topic.

A second explanation is offered by "moral disengagement." This is predicted to occur when the introduction of an extrinsic reward initiates within the decision maker a process of cognitive re-evaluation of behaviour, placing less weight on its moral and ethical aspects of behaviour and more on extrinsic costs and benefits. This was the mechanism famously suggested by Richard Titmuss (16) when predicting that a small monetary incentive would crowd-out blood donations since it is unlikely to compensate for the time and effort needed to donate blood. Moral disengagement has been corroborated across studies using different subject pools, economic games and cultures (14). In addition, evidence from brain scans shows that introducing rewards induces neural activity switches between self-interested economic decision-making and social reward evaluation (22), consistent with moral disengagement. In this case, its explanatory validity hinges on an assumption about individual level responses to the presence and size of an extrinsic reward. Moral disengagement in the strong sense implies a homogenous and instantaneous reduction in prosocial behaviour immediately after an introduction of an extrinsic reward, followed by positive relationship between reward and prosocial behaviour, in aggregate producing a "forward bending" supply curve. This pattern is inconsistent with results presented here and is ante implausible given that a reward for recycling was available at the beginning for the study period. Moral disengagement in the weak sense, however, posits individual level heterogeneity in reward threshold required for moral disengagement to occur (39, 40). In aggregate, this can produce an s-shaped supply curve as it allows for an initially positively sloped segment. The latter mechanism can explain our result and moral disengagement therefore provides a valid theoretical account.

To distinguish between decision making mechanisms, it is necessary to elicit behavioural responses to changes in extrinsic rewards at the level of the individual. So far, research on motivational crowding theory has typically relied on comparisons of aggregate group level data, for example, experimental treatment groups (15, 19, 31), implicitly invoking the common assumption of homogeneous preferences or a representative agent (see (41) for an exception). While useful for establishing aggregate patterns and plausible theoretical accounts, such data cannot distinguish between, or infer, individual level decision-making mechanisms, except in rare cases. Future work may use within subject designs to evaluate the relative significance of competing theoretical accounts in this domain.

However, the context of this study does allow us to rule out several alternative accounts of the crowding-out effect. The conjecture that "bad news" (20) and "control aversion" (42) produce an s-shaped supply curve provides competing theoretical accounts, but cannot explain the results observed here. When a decision maker in a strategic setting uses extrinsic rewards to draw inferences about beliefs about her, increased rewards can be interpreted as the decision maker being expected to be selfish or unfair (12, 19). Reacting to such bad news, decision makers may choose to behave in line with these perceived beliefs. Depending on the balance of the negative reactance effect and the positive reward effect, an s-shaped supply curve can be obtained. Similarly, when extrinsic incentives are absent, behaviour is likely viewed as an expression of autonomy and self-determination; but when present, it can be viewed as an expression of compliance (30, 42, 43). If an incentive is perceived as a means of controlling behaviour, violating a decision maker's sense of autonomy and self-determination, a crowding-out effect can result from a decision maker's "control aversion." Since in this case, the change in expected reward resulted from the Swedish Central Bank removing a coin from circulation, these interpretations are implausible.

These theoretical distinctions have implications for policy attempting to encourage prosocial behaviour. For example, moral disengagement is based on the dismal prediction that intrinsic motivation permanently vanishes, while the signalling account holds that the ability to credibly signal motivation determines behaviour. A study in the domain of blood donations (17) finds that extrinsic reward accompanied by the opportunity to donate rewards to charity eliminated the crowding-out effect (see also (40)). This shows how results attributable to moral disengagement can obtain from an inability to signal motivations. To address instances of under-supplied prosocial behaviour, these mechanisms must be distinguished if extrinsic incentives of the right magnitude are to be provided. Such efforts have proven problematic for at least two reasons.

First, experimental treatments typically rely on a small number of point estimates on the relationship between reward and prosocial behaviour. Empirical results therefore crucially depend on the calibration of experimental treatments; arbitrarily selecting up to four expected reward values in the scatter plot presented in Figure 2 can result in support for any supply curve pattern: "upward sloping", "forward bending", "backward bending", or "s-shaped". Absent knowledge of pre-existing levels of intrinsic motivation, significant uncertainty regarding the appropriate calibration remains and can preclude negative sections on the supply-curve predicted by motivational crowding theory (41, 44). This can in turn lead to false negative results since only the price effect would be detected. Moreover, since theoretical interpretation hinges crucially on whether a negative segment is preceded by a positive segment or not, researchers may incorrectly favour one crowding-out theory over another. In this way, the s-shaped pattern found here can help explain the body of mixed results in the literature on motivational crowding-out.

Second, uncertainty over pre-existing levels of intrinsic motivation further complicates calibration due to differences in inflection points on the supply curve (45–47). This compromises the legitimacy of aggregate level comparisons between groups unless a reliable measure of intrinsic motivation is available. Such measures are, however, inherently difficult to obtain since intrinsic motivation is not directly observable. Instead, it has to be inferred from observational or stated preference data, but neither type of data can solve this problem on its own. Revealed preference data do not provide

information about underlying motivations, and stated preference data may be subject to measurement error when respondents are unaware of their true motivation, reluctant to respond or post hoc justify their behaviour (44). While successfully identifying the crowding-out effect, our results suggest that the crowding-out effect is most likely to occur for relatively low levels of reward and that caution must be exercised when interpreting empirical results in this domain.

The findings presented here need to be viewed in light of two main limitations. First, the study sample is drawn from the general Swedish population, which can be characterised as WEIRD (Western, Educated, Industrialised, Rich, and Democratic) (48). While it is not possible to determine *a priori* whether our results represent a general finding with respect to human decision making across cultures, the heterogonous within-study effect found here, suggests that variation in population characteristics may be particularly important for assessing the generality of the crowding-out phenomenon. Second, the range and level of reward analysed in our study is both narrow and low in a Swedish context. Although low rewards are thought to be particularly prone to crowding-out effects, we cannot rule out that aggregate supply of prosocial behaviour will not exhibit additional backward bending segments outside the range of reward observed in our study. Preference heterogeneity and aggregation allows for crowding-out effects both above and below this range. These limitations together imply that the presence and size of crowding-out effects remains an empirical question, necessitating careful testing of reward levels prior to any large-scale policy implementation.

In conclusion, our analysis of a large natural experiment finds that a higher level of prosocial behaviour can be incentivised at lower cost. Models of prosocial decision making can be improved by recognising the role of interactions between extrinsic and intrinsic rewards. Finally, empirical studies attempting to understand the crowding-out phenomenon must estimate supply over a denser and wider range of values, and collect data on proxy variables of intrinsic motivation for participation in prosocial behaviours in order to avoid misinterpretation of results and support effective policies.

Methods

Data and natural experiment. In March 2009, The Swedish Central Bank (Sveriges Riksbank) decided to withdraw the SEK 0.5 coin from circulation, effective from 1st of October 2010. From October onwards, recyclable beverage cans carrying a deposit of SEK 0.5 were gradually replaced by cans carrying a deposit of SEK 1, as SEK 0.5 cans were recycled and removed from circulation. As a result, expected reward from recycling a single can gradually increased from SEK 0.5 and converged on SEK 1 in January 2014 (approximately USD 0.07-0.14). The exogenous decision by the Swedish Central Bank allows us to interpret this as a natural experiment (28).

Swedish law stipulates that any commercially sold can has to be part of the national recycling system. A deposit, explicitly stated on the container, is paid for each can at the time of purchase. The deposit is subsequently refunded when the can is returned to a recycling machine. Each store owns its machine(s), but maintenance, collection of recycled material and repayment to the store is managed

by Returpack, a non-profit organisation in charge of the recycling system. Returpack collects information on what and how much is recycled at each machine, and aggregates this information monthly in order to return the correct deposit to each individual store.

Kooperativa Förbundet (KF), one of the largest grocery retail chains in Sweden, provided us with monthly data on all of their 683 stores covering most Swedish counties. The data span the relevant time period from August 2010, two months before the new SEK 1 deposit cans were introduced, to January 2014 (T=42), the last period where expected reward is statistically different from 1 (two-tailed t-test: $t(484) = -6.582$, $P < 0.001$, mean = 0.999, 95% Confidence Intervals = [0.999 0.999]). The data include each store location, municipality, county, postcode and recycled items. To achieve a balanced panel dataset, any store with incomplete data was excluded. This resulted in a panel consisting of detailed recycling information for a total of 485 stores distributed across Sweden, containing 20,370 independent observations of monthly aggregates of recycled quantity, corresponding to approximately 27,454 million individual decisions.

From Statistics Sweden (Statistiska Centralbyrån) we retrieved additional data on median income on the postcode level and voting data on the municipality level. Finding a measure for intrinsic motivation is inherently challenging as motivation can only be observed indirectly using proxy variables. Here we use the proportion of people who voted for the Green Party (Miljöpartiet) in the 2010 municipality election, denoted Green votes (%). The participation rate in the election was 84.63 percent, and the proportion of people who voted for the Green Party varied across municipalities from 0.3 to 16.6 percent (49). Data from the SOM Institute (SOM Institutet) reveal that the Green Party is generally regarded as a single issue party focusing on the environment and that the majority of the party's voters report the environment as their greatest concern (50). Using this measure relies on the assumption that Green Party voters are more likely to be intrinsically motivated to participate in pro-environmental activities compared to voters in the general population. Finally, to control for pure income effects we used store postcode to match data on the median residential income retrieved from national tax records. These two measures allow us to test if the crowding-out effect depends on the level of intrinsic motivation, above and beyond the pure income effect.

Statistical analysis. The outcome variable is store level monthly quantity of recycled cans in thousands, denoted Cans. The independent variable of main interest is expected reward from recycling calculated as the proportion of SEK 0.5 cans multiplied by SEK 0.5 plus the proportion of SEK 1 cans multiplied by SEK 1, for each time period t. The average customer recycling 18 cans, saw an increase in expected reward from SEK 9 to 18 (USD 1.25 to 2.5) over the study period.

Table 1 presents regression analysis of Cans as a function of expected reward levels $E[r_{t-1}]$. The continuous expected reward variable is divided into five equal intervals. These are used to create dummy variables equal to one if expected reward takes a value within the specified interval, else zero. This approach allows us to examine the relationship between recycled quantity and expected reward without making a priori assumptions about functional form.

We use the one period lag of expected reward due to the time lag between purchasing or collecting cans and recycling them in-store. This is a better measure of the expected reward at the time of purchase and was informed by a customer survey implemented by Returpack, designed with the expressed purpose of mapping the recycling habits of the general Swedish population. The survey found that the median customer recycles once a month (see Supplementary Figure 1).

Table 2 presents regression analysis of Cans as non-linear functions of expected reward. To allow for estimation of the non-standard supply curve shapes predicted by crowding-out theories, we use specifications including second and third degree polynomials of expected reward. These different specifications allow for all empirical patterns predicted by crowding-out theories: forward and backward bending supply with a quadratic term, $E[r_{t-1}]^2$, and an s-shaped supply curve with a cubic term, $E[r_{t-1}]^3$.

The panel structure of our data allows for Random Effects Generalised Least Squares model estimation. To correct for the presence of heteroskedasticity and autocorrelation, we estimate all models using robust standard errors clustered on store id (51). We opted for the Random Effects model rather than the Fixed Effects model on the basis of the conventional Hausman test which we did not reject ($Chi2(3) = 6.77$, $P = 0.080$). This result implies that the Random Effects model is efficient, unbiased and the correct estimation method (52). Importantly, the Random Effects model permits modelling of time invariant effects, allowing us to include income and voting data.

To control for effects of seasonality of recycling apparent in Figure 1, we include monthly dummy variables in all main analysis with further robustness tests reported in Supplementary Tables 1 and 2. First, our main results hold when controlling for individual years and quantity of beverages sold (Supplementary Table 1). Second, when estimating the supply curve within time periods, using individual cross-sections instead of the aggregate data, we find evidence of negative marginal effects, indicating the presence of a crowding-out effect, in the majority of cross-sections (Supplementary Table 2). This shows that the crowding-out effect observed in our study is not a result of seasonal variation arising from the pooled data combining all individual cross-sections as in our main analysis applying panel data techniques (see also Source Data 4).

The analysis in models 3 and 4 in Table 2 extends the analysis from models 1 and 2 by adding Income and Green votes (%), and their interactions with expected reward of all polynomial degrees. This allows us to test the hypothesis that the effects of changes in expected reward on prosocial behaviour are contingent on the level of intrinsic motivation, and control for the pure income effect beyond changes in expected reward.

A number of considerations indicate that the cubic model specification is preferred to the quadratic specification. First, rejection of a mixed model Likelihood-Ratio test (model 1 vs. model 2: LR $Chi2(1) = 65.52$, $P < 0.001$) as well as Akaike and Bayesian Information Criteria indicate that the cubic model specification provides a better fit of the data. We use the Mixed Model as it yields estimates near equivalent to the random effects model estimated in our main analysis, but permits the Likelihood-

Ratio tests which the Random Effects model estimation does not. Together with theoretical predictions, the scatter plot in Figure 2 and the results from the semi-parametric regression presented in Table 1, the cubic specification in model 4 is preferred to the quadratic specification in model 3. This was also supported by the associated Mixed Model Likelihood-Ratio test (model 3 vs model 4: LR Chi2(3) = 100.21, $P < 0.001$).

Figure 3 (a-d) presents post-estimation results based on model 4 in Table 2 using the Margins command in Stata 13. To illustrate the supply curve shape at varying levels of Green votes (%), (a) presents linear predictions of quantity recycled as functions of expected reward; (b-d) display average marginal effects at varying levels of Green votes (%). All point estimates of marginal effects are accompanied by their 95% confidence intervals (see also Source Data 3).

Finally, Table 3 presents predicted crowding-out effects on the basis of model 4 in Table 2. For each level of Green votes (%), the local maxima and minima of linear predictions for quantity recycled from Figure 3 (a) are provided. These values correspond to the inflection points on each estimated supply curve, and their difference consequently correspond to the size of the crowding-out effect provided in Table 3.

Data availability

The data that support the findings of this study are available for replication from the corresponding author upon request. The data cannot be made publicly available due to property rights held by COOP and Statistics Sweden.

Code availability

Custom code that supports the findings of this study is available from the corresponding author upon request.

Acknowledgements

Financial support from the Adlerbertska Foundation (Adlerbertska Forskningsstiftelsen) (C.W.), and the Swedish Research Council (Vetenskapsrådet, 2018-04793) (C.W., P.M.) is gratefully acknowledged. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript. Access to data was generously provided by Per Rosengreen (COOP), Katarina Lundell (Returpack), Annelie Jensen (Kooperation utan gränser) and Erik Ebbeson (Returpack). We thank participants at the Copenhagen Network for Experimental Economics (CNEE), the International Meeting on Experimental and Behavioural Economics (IMEBE), Cinla Akinci, Emil Persson, Hector Gutierrez Rufrancos and members of Stirling Behavioural Science Centre, especially Philip Ebert, Mirko Moro and Frans de Vries, for helpful comments. Views herein are exclusively those of the authors.

Author contributions

C.W., M.K., and P.M. designed research; C.W., M.K., and P.M. performed research; C.W., and M.K. analysed data; C.W. wrote the paper.

Competing interests

The authors declare no competing interests.

Tables

Table 1. Random effects GLS results: Estimated quantity of recycled cans as a function of expected reward levels.

Dep. Var.	Cans
L2 [0.6-0.7)	2.6 (0.3)
	[2.0 3.2]
	<0.001
L3 [0.7-0.8)	0.8 (0.3)
	[0.2 1.5]
	0.013
L4 [0.8-0.9)	1.5 (0.2)
	[1.0 1.9]
	<0.001
L5 [0.9-1.0]	2.0 (0.4)
	[1.2 2.9]
	<0.001
Month	Yes
Constant	22.2 (1.2)
	[19.8 24.5]
	<0.001
Observations	20,370
Groups	485
R-squared:	
Within	0.206
Between	0.000
Overall	0.013

Notes. L2 [0.6-0.7) - L5 [0.9-1.0] denote dummy variables that equal 1 if expected reward E[r_{t-1}] is in the specified interval (else 0); Baseline expected reward level is L1 [0.5-0.6); Robust standard errors clustered on id in parentheses and 95% Confidence Intervals in square brackets as [low high] with exact *P* values below (no brackets). H0: Estimated coefficient is equal to zero for all tests (Chi2-test, two-sided). No adjustment has been made for multiple comparisons.

Table 2. Random effects GLS results: Estimated quantity of recycled cans as a function of expected reward, including green votes, income and their interactions with expected reward.

Model	(1)	(2)	(3)	(4)
Dep. Var.	Cans	Cans	Cans	Cans
$E[r_{t-1}]$	15.8 (4.0)	379.5 (87.2)	85.2 (26.9)	1279.4 (346.4)
	[7.9 23.7]	[208.5 550.4]	[32.6 137.8]	[600.4 1958.4]
	<0.001	<0.001	0.002	<0.001
$E[r_{t-1}]^2$	-7.7 (2.7)	-503.2 (118.2)	-47.8 (17.3)	-1682.5 (462.5)
	[-12.9 -2.5]	[-734.9 -271.6]	[-81.7 -13.9]	[-2588.9 -776.0]
	0.004	<0.001	0.006	<0.001
$E[r_{t-1}]^3$		218.1 (52.3)		722.4 (200.9)
		[115.7 320.5]		[328.7 1116.2]
		<0.001		<0.001
GV			1.7 (0.4)	-8.7 (4.1)
			[1.0 2.3]	[-16.7 -0.7]
			<0.001	0.033
GV - $E[r_{t-1}]$			-0.4 (1.4)	43.2 (18.4)
			[-3.2 2.3]	[7.2 79.2]
			0.760	0.019
GV - $E[r_{t-1}]^2$			0.5 (0.9)	-58.7 (24.5)
			[-1.3 2.3]	[-106.7 -10.7]
			0.582	0.017
GV - $E[r_{t-1}]^3$				25.9 (10.6)
				[5.1 46.7]
				0.015
Income			-0.2 (0.4)	13.5 (4.4)
			[-0.9 0.6]	[4.8 22.1]
			0.691	0.002
Income - $E[r_{t-1}]$			-3.4 (1.6)	-61.3 (20.0)
			[-6.5 -0.4]	[-100.5 -22.2]
			0.027	0.002
Income - $E[r_{t-1}]^2$			1.9 (1.0)	81.1 (26.6)
			[-0.1 3.9]	[28.9 133.3]
			0.060	0.002
Income - $E[r_{t-1}]^3$				-34.9 (11.5)
				[-57.6 -12.3]
				0.002
Month	Yes	Yes	Yes	Yes
Constant	16.0 (1.5)	-69.7 (20.3)	8.3 (7.0)	-272.4 (77.3)
	[13.2 18.9]	[-109.6 -29.9]	[-5.5 22.1]	[-423.9 -120.9]
	<0.001	0.001	0.237	<0.001
Observations	19,885	19,885	19,885	19,885
Groups	485	485	485	485
R-squared:				
Within	0.206	0.208	0.209	0.213
Between	0.003	0.011	0.057	0.057
Overall	0.012	0.012	0.066	0.066

Notes. GV denotes Green votes (%); Robust standard errors clustered on id in parentheses and 95% Confidence Intervals in square brackets as [low high] with exact *P* values below (no brackets). H0:

Estimated coefficient is equal to zero for all tests (Chi2-test, two-sided). No adjustment has been made for multiple comparisons.

Table 3. Linear predictions of average monthly quantity recycled and crowding-out effect in thousands, by Green votes (%).

Green votes (%)	2.54	5.76	6.46	13.87
Local max. quantity	20.5	26.0	27.2	39.9
Local min. quantity	20.0	25.0	26.1	37.5
Crowding-out effect	0.5	1.0	1.1	2.4
Crowding-out effect (%)	2.2	3.8	4.0	5.9

Notes. Values of Green votes (%) correspond to the 10^{th} percentile, median, mean and the 90^{th} percentile (2.54, 5.76, 6.46, 13.87, respectively) (Source Data 3).

Figure Legends

Figure 1. Percentiles of expected reward (lines) and average quantity recycled (bars) over time. The solid line represents the median and the dashed lines represent the 2.5th (lower line) and 97.5th (upper line) percentiles, spanning 95 percent of the expected reward E[r_t] distribution. Median expected reward increases gradually from SEK0.5 to 1 and average quantity recycled exhibits a seasonal pattern (n=20,370).

Figure 2. Average quantity recycled and expected reward. The figure presents a scatter plot of average quantity recycled and expected reward for each 0.01-interval of E[r_t] between SEK 0.5 and 1. The resulting pattern appears to violate the fundamental "law of supply" by failing to exhibit a monotonically increasing relationship between expected reward (price) and average quantity recycled (quantity) (n=20,370).

Figure 3. Linear predictions and marginal effects. The figure combines linear predictions of quantity recycled as a function of expected reward by levels of Green votes (%) and average marginal effects of expected reward on linear prediction with 95% confidence intervals. (a) Lines represent linear predictions of quantity recycled as a function of expected reward by levels of Green votes (%) corresponding to the 10th percentile (blue line), the 50th percentile (red line), and the 90th percentile (green line). (b-d) Solid lines represent point estimates of average marginal effects on linear predictions and fields represent 95% confidence intervals (see Source Data 3 for full statistical results). (a) Indicates non-monotonic linear predictions as a function of expected reward consistent with a crowding-out effect. (b-d) Reveal ranges of negative marginal effects, statistically significant at less than the 5% level for Green votes (%) levels 5.76 (c) and 13.87 (d). Overall, results suggest that the crowding-out effect is heterogeneous with respect to level of Green votes (%) (n=19,885).

References

1. Dwenger, N., Kleven, H., Rasul, I. & Rincke, J. Extrinsic and intrinsic motivations for tax compliance: Evidence from a field experiment in Germany. *American Economic Journal: Economic Policy* **8**, 203–32 (2016).
2. Fehr, E. & Gächter, S. Do incentive contracts undermine voluntary cooperation? Preprint at SSRN http://dx.doi.org/10.2139/ssrn.313028 (2002).
3. Kolm, S.C. Introduction to the economics of giving, altruism and reciprocity. in *Handbook of the Economics of Giving, Altruism and Reciprocity* (eds. Kolm, S.C. & Mercier-Ythier, J.) 1–122 (North-Holland, 2006).
4. Meier, S. A survey of economic theories and field evidence on pro-social behavior. in *Economics and Psychology: A Promising New Cross-Disciplinary Field* (eds. Frey, B.S. & Stutzer, A.) 51-88 (MIT Press, 2007).
5. Gneezy, U., Meier, S. & Rey-Biel, P. When and why incentives (don't) work to modify behavior. *Journal of Economic Perspectives* **25**, 191–210 (2011).
6. United Nations Development Programme. *Human Development Report 2006* (Palgrave Macmillan UK, 2006).
7. World Health Organization. *WHO Global Strategy for Containment of Antimicrobial Resistance.* (World Health Organization, 2001)
8. Ostrom, E. *Governing the commons.* (Cambridge Univ. Press, 2015).
9. Hardin, G. The tragedy of the commons. *Science* **162**, 1243–1248 (1968).
10. Nyborg, K. et al. Social norms as solutions. *Science* **354**, 42–43 (2016).
11. Frey, B.S. *Not Just For The Money: An Economic Theory of Motivation* (Edward Elgar, 1997).
12. Falk, A. & Kosfeld, M. The hidden costs of control. *American Economic Review* **96**, 1611–1630 (2006).
13. Bénabou, R. & Tirole, J. Intrinsic and Extrinsic Motivation. *The Review of Economic Studies* **70**, 489–520 (2003).
14. Bowles, S. & Polania-Reyes, S. Economic incentives and social preferences: substitutes or complements? *Journal of Economic Literature* **50**, 368–425 (2012).
15. Gneezy, U. & Rustichini, A. A fine is a price. *The Journal of Legal Studies* **29**, 1–17 (2000).
16. Titmuss, R. *The Gift Relationship (reissue): From Human Blood to Social Policy* (Policy Press, 2018).
17. Mellström, C. & Johannesson, M. Crowding out in blood donation: was Tituss right? *Journal of the European Economic Association* **6**, 845–863 (2008).
18. Holmås, T.H., Kjerstad, E., Lurås, H. & Straume, O.R. Does monetary punishment crowd out pro-social motivation? a natural experiment on hospital length of stay. *Journal of Economic Behavior & Organization* **75**, 261–267 (2010).
19. Fehr, E. & Rockenbach, B. Detrimental effects of sanctions on human altruism. *Nature* **422**, 137-140 (2003).
20. Dickinson, D. & Villeval, M.C. Does monitoring decrease work effort?: The complementarity between agency and crowding-out theories. *Games and Economic Behavior* **63**, 56–76 (2008).
21. Bewley, T.F. *Why wages don't fall during a recession* (Harvard Univ. Press, 2009).
22. Li, J. Xiao, E., Houser, D. & Montague, P.R. Neural responses to sanction threats in two-party exchanges. *Proceedings of the National Academy of Sciences of the United States of America* **106,** 16835-16840 (2009).
23. Murayama, K. Matsumoto, M., Izuma, K. & Matsumoto, K. Neural basis of the undermining effect of monetary reward on intrinsic motivation. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 20911–20916 (2010).

24. Goette, L. & Stutzer, A. Blood donations and incentives: Evidence from a field experiment. Preprint at SSRN: https://ssrn.com/abstract=3376608 (2019).
25. Lacetera, N. Macis, M. & Slonim, R. Will there be blood? Incentives and displacement effects in pro-social behavior. *American Economic Journal: Economic Policy* **4**, 186–223 (2012).
26. Iajya, V., Lacetera, N., Macis, M. & Slonim, R. The effects of information, social and financial incentives on voluntary undirected blood donations: Evidence from a field experiment in Argentina. *Social Science & Medicine* **98**, 214–223 (2013).
27. Rommel, J., Buttmann, V. Liebig, G., Schönwetter, S. & Svart-Gröger, V. Motivation crowding theory and pro-environmental behavior: Experimental evidence. *Economics Letters* **129**, 42–44 (2015).
28. Harrison, G.W., List, J.A. Field experiments. *Journal of Economic Literature* **42**, 1009–1055 (2004).
29. Frey, B.S. A constitution for knaves crowds out civic virtues. *The Economic Journal* **107**, 1043–1053 (1997).
30. Deci, E.L., Koestner, R. & Ryan, R.M. A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin* **125**, 627-668 (1999).
31. Gneezy, U. & Rustichini, A. Pay enough or don't pay at all. *The Quarterly Journal of Economics* **115**, 791–810 (2000).
32. Deaton, A. & Cartwright, N. Understanding and misunderstanding randomized controlled trials. *Social Science & Medicine* **210**, 2–21 (2018).
33. Open Science Collaboration. Estimating the reproducibility of psychological science. *Science* **349**, 1-8 (2015).
34. Meier, S. Do subsidies increase charitable giving in the long run? Matching donations in a field experiment. *Journal of the European Economic Association* **5**, 1203–1222 (2007).
35. Bandura, A. Social cognitive theory of moral thought and action. in *Handbook of Moral Behavior and Development*: *Volume 1* (eds. Kurtines, W.M., Gewirtz, J. & Lamb, J.L.) 69–128 (Psychology Press, 2014).
36. Cárdenas, J.C., Stranlund, J. & Willis, C. Local environmental control and institutional crowding-out. *World Development* **28**, 1719–1733 (2000).
37. Pettersson, G. *Miljöstatistik* (Skill, 2012).
38. Mas-Colell, A. Whinston, M.D. & Green, J.R. *Microeconomic Theory*. (Oxford Univ. Press, 1995)
39. Bénabou, R. & Tirole, J. Incentives and prosocial behavior. *American Economic Review* **96**, 1652–1678 (2006).
40. Beretti, A., Figuieres, C. & Grolleau, G. How to turn crowding-out into crowding-in? an innovative instrument and some law-related examples. *European Journal of Law and Economics* **48**, 417–438 (2019).
41. Müller, S. & Rau, H.A. Motivational crowding out effects in charitable giving: Experimental evidence. *Journal of Economic Psychology* **76**, 102210 (2020).
42. Frey, B.S. & Jegen, R. Motivation crowding theory. *Journal of Economic Surveys* **15**, 589–611 (2001).
43. Richard, M. & Deci, E. *Intrinsic Motivation and Self-Determination in Human Behavior* (Springer US, 1985).
44. Rode, J., Gómez-Baggethun, E. & Krause, T. Motivation crowding by economic incentives in conservation policy: A review of the empirical evidence. *Ecological Economics* **117**, 270–282 (2015).
45. Bohnet, I. & Baytelman, Y. Institutions and trust: Implications for preferences, beliefs and behavior. *Rationality and Society* **19**, 99–135 (2007).

46. Kessler, E. Behavioural Economics of Performance Incentives. (University of Nottingham, 2008).
47. Carpenter, J. & Myers, C.K. Why volunteer? Evidence on the role of altruism, image, and incentives. *Journal of Public Economics* **94**, 911–920 (2010).
48. Henrich, J., Heine, S., & Norenzayan, A. The weirdest people in the world? *Behavioral and Brain Sciences* **33**, 61-83 (2010).
49. Valmyndigheten. Valresultat 2010. https://data.val.se/val/val2010/673slutresultat/R/rike/index.html (2010).
50. University of Gothenburg, SOM Institute. The National SOM Survey 2010. Swedish National Data Service. Version 1.0. https://doi.org/10.5878/002316 (2012).
51. Cameron, A.C. & Miller, D.L. A practitioner's guide to cluster-robust inference. *Journal of Human Resources* **50**, 317–372 (2015).
52. Wooldridge, J.M. *Econometric Analysis of Cross Section and Panel Data.* (MIT Press, 2002).