Dana 2007: Exploiting Moral Wiggle Room: Experiments demonstrating an illusory preference for fairness

* Does generosity in experiments really mean we care about desirable social outcomes?
  + Binary version of ‘dictator game’. Several treatments where subjects can ‘leave’ the relationship between their actions and the outcomes uncertain
    - Either to themselves or another subject influenced by these actions, thus providing the plausible deniability or ‘wiggle room’ to behave in self-interested ways.
* Finds significantly less generous behavior in these manipulations, relative to the baseline where how actions relate to outcomes is transparent.
* Conclusion is that ‘fair behavior’ is mostly because people intrinsically dislike appearing unfair, either to themselves or others

Fairness can be easily looked at in ‘dictator games’, where a dictator makes a 1-shot division of money between themselves and an anonymous recipient who has no recourse. A selfish person would keep the entire amount, but the majority of subjects give SOMETHING, and the average amount given is over 20%. Even when there is DOUBLE-blind anonymity!

“Social Preference” theories argue this is due to a preference for equitable outcomes or social welfare. People MAY share with each other because they have increasing utility in others’ payoffs, are averse to advantageous payout differences, or want to maximize TOTAL social payoffs or the lowest payoff to any one party. The common feature here is that the dictators’ preferences are characterized by only looking at final distribution of wealth.

THUS ‘Giving’ can be interpreted as being like any other consumption good, except that the dictator is ‘buying’ equity or social welfare!

It’s possible that you may want to give in some cases! Even when you are selfish, because you do not want to APPEAR selfish, either to themselves or others. The motivation driving this ‘fair’ behavior could be self-interest, and/or a desire to maintain the illusion of not being selfish. The same people who give in a context like the dictator game could prefer the unfair outcome, as long as there is an excuse to not have to give (or not know the consequences of their actions!), e.g.

* Person who would donate bone marrow if they had a match on the registry, removes themselves from the registry
* People who have STD who might stop having sex if they knew for sure, may avoid testing in order to keep having sex.
* Alternatively, other ‘dictators’, or those who COULD help, in order to not feel like they themselves have to help.
  + “I wont stop to help the stranded car b/c someone else will’
  + “This crime victim will be fine, someone else will help”
* Dictator can exploit (possibly asymmetric) uncertainty about what, precisely, causes unfair outcomes
  + Under plausible deniability, could act more in self-interest
  + E.g. a Manager would act in a manner beneficial to employees if it’s clear that the manager is solely responsible, but if that’s not the case… they can behave self-interestedly at their expense!

Rather than having a preference for a ‘fair outcome’, people may ‘conform’ to situational pressure to give in certain contexts. Conversely, some situations can be exploited for justification to behave selfishly.

* Test this w/ a nonstandard ‘dictator game’
  + Historically, ‘dictator game’ is transparent, there is common 1-to-1 mapping between the dictator’s actions and outcomes
* Series of manipulations, capitalizing on uncertainty to eliminate transparency.
* Most dictators ARE generous in a baseline transparent game
  + Selfishness increases SIGNIFICANTLY in the absence of transparency.
  + Are people aware of this happening, and is that part of why there is hesitation to trust/support something that is nontransparent (UHC service distribution?)
* This cannot be due to motivation to prefer some types of outcome, as transparency would be irrelevant there, the dictators can STILL force the outcomes they want.
* Additionally, reciprocity and perceptions of others’ intentions are important to determine utility in a social outcome.
  + This doesn’t matter here though, as the receiver is in a ‘passive’ position

Study

* Binary choice between equal and unequal (and welfare inefficient) wealth allocation
  + Baseline game was transparent
  + Three manipulations relax transparency assumption
* Subjects randomly assigned to one treatment
  + Undergraduates at University of Pittsburgh, volunteering in exchange for money
  + Drew cards to represent role (dictator or passive)
* Subjects were told that all members in the group would be paid according to the dictators (Player X in most treatments, X and Y in multiple dictator condition)
* Subjects were then given instructions and info on the actual payoffs.

Four main treatments

* Baseline (n = 38)
  + Dictators choose between A or B, with 60 seconds of time looking at the payoff matrix to force consideration of what they would do.
  + Relationship between actions and outcomes is transparent!
* Hidden Information (n = 64)
  + Dictator remained ignorant to the consequences for recipient, they got 6$ for A, and 5$ for B, but the amount the other person would get was unknown, told that they would either get 1 and 5, or 5 and 1 (where both parties are better for choosing A)
    - True payoffs were told that they could not be revealed publicly, but could be done by pressing a button. Informed that this choice would be private.
  + If giving money in the baseline = preference for equitable distribution, then % of dictators who give should be equal to proportion that reveals true payoffs, AND chooses the most equitable action.
    - Instead, if dictators are seeking an excuse to not feel compelled to give, then they might choose to remain uninformed and choose A under ignorance.
* Multiple Dictator (n = 30)
  + Second dictator added, thus no dictator is solely responsible, although the ‘fair’ outcome can be implemented unilaterally by either.
    - 2/3rd of subjects assigned to dictator X or Y, 1/3rd assigned to recipient.
    - Both dictators must choose A to get the inequitable outcome.
  + If the baseline reasoning holds, should have same proportion of B choices in the treatment as in the baseline.
  + Note that this treatment BREAKS transparency, the selfish choice A no longer guarantees the unfair outcome for the recipient.
* Plausible Deniability (n = 58)
  + Allow for the ‘unlikely’ possibility of the dictator losing agency, thus allowing outcomes to plausibly result from causes other than the dictator’s actions.
    - Adds a ‘cut-off’ feature, there is a 10s window to make choices, but if not made within a random cutoff time in the interval, it would choose between A and B with equal probability.
    - Only the dictator knows if a cutoff occurs.
  + Cutoff feature is largely irrelevant as most choices can be made within 2 seconds, and at most, 4 seconds (most cutoffs were at 5 and 6).
  + This relaxes the assumption of transparency, distinguishing between two types of mechanisms for moral ‘wiggling’
    - First, receivers can’t differentiate between dictators and nature, dictators choosing A more frequently would be an ‘other-deceptive’ motive. But the dictator still knows they were responsible for the inequity!
    - If ‘self-deception’ is responsible, dictators might ‘dither’ and let themselves be cutoff, with a 50% likelihood of getting the fair outcome they are compelled to choose, but with 50% likelihood of ‘winning’ with the selfish outcome.

Results

* Baseline game: Majority of dictators acted fairly, 74% chose the even split. All receivers chose B. This is consistent w/ previous evidence of sharing in dictator games, and is INTERPRETED as supporting the idea that people prefer the generous outcome.
  + Also consistent with the idea that dictators feel compelled to give in transparent situations.
* Hidden Payoff treatment: Of the 16 dictators looking at the original payoff, 63% chose the uneven split, even though revealing this information is no-cost.
  + Only 56% even chose to reveal the payoff!
  + Only 47% chose to reveal AND chose the ‘fair’ option, less than the proportion in the original game (even though revealing was ‘no-cost’ and happened!)
  + Hypothetical receiver choices mirror this, all receivers wanted the fair option in baseline, only 59% wanted it in the hidden payoff condition.
    - Suggests that several choices were NOT the result of dictators wanting to implement the ‘fair’ outcome
    - Many dictators appeared to exploit the payoff uncertainty as an excuse for self-interested behavior.
* Multiple Dictator treatment: only 35% of subjects chose fair choices in the two dictator solution. All receivers correctly predicted that ‘unfair’ would be the most popular choice.
  + Likely that a great deal of ‘generosity’ isn’t derived from socially desirable outcomes, but to be perceived as doing so
* Plausible deniability treatment: This treatment allows the researcher to examine whether or not decreased giving is due to self or other deception.
  + Amongst dictators not cut off, 55% chose the selfish action, A, higher than the proportion in the baseline
    - Receiver uncertainty on payoff types seems to be enough to promote increased self-interest
  + 24% of dictators were cutoff and did NOT make a choice.
    - Many subjects seemed willing to delay making a choice, with the hope of avoiding making a choice altogether.
  + Only 10 out of 29, 34% are consistent with the desire to be ‘fair’
  + Dictators engaged in moral wiggling are heterogeneous in how they obtain the selfish outcomes. Some directly choose A (exploiting receiver uncertainty), others allow themselves to be cutoff by the computer (exploiting their own lack of agency and uncertainty over outcomes)
    - 45% of receivers would implement the ‘fair’ outcome in this condition

Conclusions

* Historically, generosity in experiments is interpreted as a preference for a fair/efficient outcome.
* Giving is actually consistent with people feeling COMPELLED to give due to situational factors, while not really valuing the corresponding outcome.
  + Relaxation of the transparency standard allows for enough moral ‘wiggle room’ to behave selfishly while maintaining the illusion of fairness.
* In all three conditions relaxing transparency, support for fairness plunges from around 75%, to around half, 35%.
  + The differences in context matter, but this pattern is similar to other experiments measuring ‘moral wiggling’
* Some environments have STRONG prescriptions for fair behavior (e.g. fully transparent dictator games), but these norms or constraints are less binding/compete with other norms once transparency is eliminated!
  + E.g. the ‘mind your business’ norm to justify not acquiring information on the other party’s payoff.
* The lack of ‘certainty’ regarding the consequences of the other party allows much of the self-interested behavior to occur.
  + The proportion of dictators who choose with certainty to implement unfair outcomes is less than ½!
* Relating to other work on fair behavior, previous studies show that people capitalize on information asymmetry or uncertainty to behave more to self-interest
  + Mostly to avoid sanctions or keep others ignorant of whether or not an outcome is fair.

Dana 2012: Ethical Immunity: How People Violate Their Own Moral Standards Without Feeling They Are Doing So

* Ethics are important in modern economy/organization. There are tons of opportunities to leverage malfeasance into personal upside.
* When people begin acting unethically, do they recognize that they are doing so?
  + Literature shows that the desire to appeal ethical to oneself and others has strong influence on human behavior.
  + Even if incentives exist to behave selfishly, we see punishing unethical others with no benefit to self, honesty when lying is profitable and undetectable, and egalitarian distribution of wealth at the expense of oneself.
* Given this, how is it possible that bad things happen so often?

We wish to examine the study of Behavioral Ethics

* Similar to behavioral economics, explains when and why ethically minded individuals behave unethically.
  + E.g. Behavior Econ shows how people’s behavior falls short of fulfilling their own goals (dynamic inconsistencies, procrastination, etc.)
  + Similarly, evidence exists that different but consequentially equivalent frames in a decision lead to different choices, showing that there are some contradictions for people with only minor variations on the context.
* People can fail to ‘systematically’ live up to their own ethical standards.
  + People might not be as ethically as even they themselves wish!
  + Difference from normative or legal ethical study, we want to understand how EVEN WELL INTENTIONED people can sometimes behave unethically.
  + Going beyond the idea that bad people are just ‘a few bad apples’
* Current instruction shows how people can use ethical principles to discriminate between ethical and unethical behavior.
  + Assumes however, that that by highlighting the moral components of choices, the people will choose the moral path
  + Behavioral ethics can illuminate when and why people who know the rules, and want to follow them, end up breaking them
  + This allows for us to potentially avoid some ‘ethics traps’
    - Ideally, allows for us to design institutions to reduce those traps, and make it ‘easy and natural’ for people to behave in conformity with their own ethical standards.

Why do people violate their ethical standards? Biggest reason: People are able to persuade themselves that they are NOT doing so!

* People can rationalize unethical behavior or position themselves to achieve ‘ethical immunity’ (avoiding information, delegating unethical activities to others). Not facing up to consequences and obvious interpretations of their actions.
* Examines 3 main ‘tactics’ that people use to avoid holding themselves ethically accountable.
  + Each line demonstrates people behave more ethically when there is moral accountability (actions clearly/directly reflect their ethical standards)
  + If you can avoid this, people will do so, and behave less ethically.
  + ‘People avoid moral accountability, so they will not feel compelled to behave in accord with their own standard of ethics’

Brief Note: Are Lab experiments even good for looking at ethics?

* Lots of literature shows that some ‘fairness’ and ‘honesty’ are indeed gains.
  + Giving up some money so others can benefit, not lying for gain, etc.
* This is good to understand ethics in clear and simple positions/situations.
* While lab experiments show there is some propensity for ethical behavior, this does not provide evidence about how people are ethical or NOT in other related contexts.
  + No research can do so, as all research focuses by necessity on a limited set of circumstances, with a particular set of characteristics.

Diffusing Responsibility

* People can avoid accountability for difficult ethical choices when the responsibility for the choice itself is diffused. Other people existing allows for us to diffuse responsibility in a way that enables people to act self-interestedly.
* Two main types of diffusion, Vertical and Horizontal:
  + Vertical: Occurs when there is an intermediary between the decision maker and the stakeholders in the choice. The decision maker is less responsible for the outcome.
    - E.g. Company can use ‘firing consultants’ which don’t do much other than firing employees.
    - Firms can outsource to contractors that pay less than the standards of the firm itself
  + Horizontal: Decision makers rely on others to act in a stakeholder’s best interests rather than doing so themselves, essentially the ‘bystander effect’.
* Diffusion works even for external observers as well as those directly involved.
* The extent to which people hold *themselves* accountable for actions is mitigated when acting through intermediaries.
  + “Principles” decided which of several “agents” to hire, to decided how much money given to the principle would be shared by a third party.
  + Principles here tended to fire those who gave a lot, and hire those who gave a little. The “employed’ agents were those who gave almost none of the dictator’s money to third parties.
  + Motivated to be hired, the agent’s choices become more favorable to the dictator over time, leaving the other party almost nothing.
    - Relative to where the dictator makes the choice themselves, much less was shared. Even though the dictators hired the agents, they felt less responsible for low payoffs for third parties.
* Delegation makes selfish outcomes easier because you don’t feel as morally accountable when someone else (even someone you hired) is doing the dirty work.
* Another example: Production of two subjects, which has costs for a third party.
  + The subjects decide the amount of product, which leads to their own payoff and the payoff for the third party.
  + Alternative arrangements: decision is made vertically (one subj sets production, the other carries it out), or horizontally (both subjects jointly determine production, w/ consensus or reaching avg) and varying the amount of communication b/w subjects.
  + Less harm for a third party when both subjects under ‘avg’ responses, but even worse when it’s vertically diffused, or horizontally w/ forced consensus (b/c the consensus choice chosen is usually the less ethical)
    - When allowing lots of communication, the vertical structure had the best outcomes, giving subordinates voices and responsibility reduces the effects of vertical diffusion.
* Horizontal diffusion can be just as bad (e.g. bystander effect)
* However, the bystander effect can be seen as confounding two phenomena.
  + People might not care about what ethical choice is made if they don’t feel responsible
  + Consistent w/ a free rider problem, people may truly care about ethics, but would prefer not to incur the costs of taking the ethical action themselves if there is a good chance someone else will.
* Dana et al. finds the first effect in an economic game:
  + 2 choice makers choose between actions giving money for them and a 3rd party.
  + Each choice maker can ensure a fair outcome, and the unfair outcome only happens if both decision makers select it, thus, either choice maker can unilaterally guarantee a ‘fair’ outcome.
  + No free rider problem was present, there is no material advantage to letting someone else ensure equity.
  + This resulted in selfish choices more often, as this only happens if the other person is also selfish, which diffuses responsibility

Exploiting Uncertainty

* In binary dictator scenarios, when the other group didn’t realize that resources can be split fairly, was seen as maximizing the total amount of net payout is prioritized unless there is uncertainty/obscuration on the relation between payoff and action.
  + Hidden information treatments, the decision maker simply reveals the ‘true payoffs’ and selects the more generous option
    - Roughly half did not choose to reveal, allowing for their choices to be unfair. Not knowing how others are impacted, provides the decision maker w/ ethical immunity to pursue self-interest, regardless of whether or not it’s easy to find out how others are impacted.
* These experiments suggest people can sacrifice personal gain for others when information is known, but doesn’t hold under some uncertainty (mostly irrelevant uncertainty, even if it’s easy to resolve
  + People abuse this as an excuse to be unethical.
* Haisley et al shows another example: In their experiment, decision makers make binary choices, wit one being more equitable, but one more personally beneficial (selfish).
  + The payoff was contingent on a ‘lottery’, when the probability of winning could be anywhere between 0 and 1 (ambiguity) or 50% (known) condition.
  + When the payoff is known, the dictator cannot manipulate perception of likelihood
  + Under ambiguity, the subject could convince themselves that the lottery is favorable for the recipient, which allows the dictator to ask selfishly, while believing that they were not doing so!
    - This relies on ethical immunity provided by ‘ambiguity’ and what they’re doing is ‘not that bad’
* Generally, selfishness occurs more often under ambiguity as compared to ‘known risk’. Biasing the incentivized estimates of payoffs to third parties upward under ambiguity, showing that they convinced themselves that the lotteries would work out better for the third party.

Seeking Justifications

* People play ‘mind games’ with themselves, serving a function of not holding them to account, based on ignorance of information and situations.
* Van Avermaet et al: participants filled questionnaires until told to stop, and they left with money to pay themselves and another person who left, told one of four things.
  + The other person spent ½ time and ½ output
  + Other person put in ½ time and 2x output
  + Other person put in 2x time and 2x output
  + Other person put in 2x time and ½ output
* Almost no one kept all the money, but splitting was based on effort.
  + Those who worked twice as long and twice as output were given twice as much money.
  + The participants kept > ½ money in when the other worked twice and did less, worked less and did twice, which is based on ‘merit’ but systematically favors the subject.
  + Half as much and half as long resulted in half the money as well.
* ‘What is a fair allocation’ is malleable, towards judgements that favor the self
  + Most people want to be fair/ethical, but these systems of fairness can systematically be biased towards the self.
* Behavioral economists show that self-serving bias is caused by selective interpretation of information, in both labs and in field studies, to failures/impasses in negotiation.
* Lowenstien et al., Experiment on settling, defendants had money to finance settlement, division of the money was what they took home as pay, the longer it took to settle, the more both sides lost as the total money shrank.
  + Under no settling, the money was split by a neutral judge (and asked to predict how the judge would rule, and paid extra based on accuracy).
  + Estimates of ‘fair’ settlements were biased in self-serving manner. This was reflected in bad predictions of judge ruling, which were higher than those of the defendants (even tho the estimates were secret and didn’t effect the settlement)
  + Larger discrepancies between plaintiffs and defendants were associated with lower likelihood of settling.
  + Self-serving biases are unintentional, even when it’s best to not have them!
* Repeated but where the evidence is all read out first, and then sides are assigned.
  + This prevents motivation from interpreting evidence as favorable to one side or another, while researching. Perceptions and predictions of fairness, were more closely related to a ‘neutral’ third party, and this lead to drastically increased ‘settlement’ rates.
* One similar approach uses individual’s initial perception of ambiguity to constrain how much they can abuse it to act unethically.
  + Asked dictators their preference towards ambiguity versus risk before the choice. This mild intervention was sufficient to constrain later actions/judgements.
  + These dictators did not view ambiguity as favorable, resulting in less selfish behavior.
  + Obtaining the initial ‘unbiased’ judgements of what’s fair or ethical, can be very powerful at preventing deflection of responsibility.
* Rationalization for accepting gifts as a doctor:
  + Most doctors early in career were asked to see if getting gifts were good
  + Some were asked about sacrifices in med school first, others were asked the same question, but also the idea that hardship might make accepting gifts OK to do.
  + Control group were asked about getting gifts without priming questions.
* Reminding physicians of sacrifices made gifts seem more acceptable, even though most disagreed with the presented rationalization.
  + Thus, reminders of hardship lead physicians to conclude that they deserved gifts.
* “Die in Cup”: Subjects were asked to roll a dice under a cup, private see the result, and then get paid based on the # rolled.
  + The distribution of reports differed significantly from a fair dice roll, with only a modest amount of lying (37% claimed to roll a 6), which fit the distribution of reporting the best of three rolls, even though they only were supposed to report the first.
  + Done again with hiding the result from being seen, we saw that there was less lying, as the participants needed justification to be able to lie (having the opportunity to roll multiple times was the justification, they could roll until they saw a better #, then they felt that their report was ‘less of a lie’)

Conclusion

* Ethics is complex, with lots of context around right or wrong choices.
* There are a lot of opportunities to act unethically in business, as there is mass diffusion of responsibility.
  + Also lots of uncertainty
  + Also lots of room to make shitty justifications that are self-beneficial.
* We can attack ‘ethical immunity’ by creating greater self-accountability.
  + “choice architecture” can be used to help rational people make better choices.

Hansson 2020: Losing sense of fairness: How information about a level playing field reduces selfish behavioraH

* Inaccurate beliefs about procedural fairness can lead folks to act in selfish manners.
  + Does information about a ‘level playing field’ mitigate these behaviors?
  + Manipulate whether participants are informed about the fairness of a competition or not.
    - Following the competition, subjects (who won or lost) decided to distribute earnings for them and their opponent.
    - Informing participants about the fairness of the competition reduces selfish behavior for losers (not affecting winners), but losers who were not informed about the fairness incorrectly viewed it as being unfairly stacked against them.
  + Information about perceived ‘fairness’ reduces selfish behavior and is important for understanding when and why motivated reasoning about fairness helps people uphold positive self-image.

Winners don’t complain about fairness, but losers often do (for referees). Especially for academics, people at work, and in general folks want to see the world ‘as they want it to be’, which doesn’t work well in conditions of ambiguity and uncertainty.

* External attribution for failures
* Internal attribution for successes

Because this his hard asf, people generally search for info that bolsters their views/beliefs, which leads them to remembering good things and bad things that they were struggling through more than the equivalent for other people. Viewing lack of success as due to unfair processes can be problematic, and subjective beliefs on whether the outcomes were fair might have them act selfishly to ‘level the playing field’

Study

* Losers of a competition who get info that both were done fairly will be less selfish than losers with no info
  + When no info is given, losers will be more likely to believe (than winners) that the competition was unfair
  + Beliefs about fairness will be correlated to selfish behavior for losers in a competition
* N = 444, English speakers recruited online, done in Qualtrics
* Assigned to one of two conditions, information or no information.
  + Participants did competition against other people and given more earnings for doing well. 12 easy and 12 hard tasks were given
  + Afterwards, they were told that they won or lost, in the info condition, participants were told that the same tasks were equivalent, and in the no info condition, people had to guess for themselves to estimate how hard or easy it was for the opposition.
  + Then, people gave the ability to split rewards afterwards, losers were asked how many they wanted to ‘take’ from the winner, and winners were ask how many they wanted to ‘give’ to the losers.

Results

* Information reduces selfish behavior in losers, about 15% less selfish behavior. Winners have no main difference either way.
* Had scored performance objectively, for those told that they lost, and split losers into high and low performance.
  + For high performing losers, told that the competition was fair, took much less than those high performing losers in the no info condition.
  + No difference in selfish behavior between low performing losers, with or without information.
    - Losing a competition after doing well makes concerns about fairness more salient, and increases potential selfish behavior.
* Why are losers less selfish, but winners not more altruistic, when given information?
  + Belief was that the competition had less difficult tasks, or that they had more difficult tasks than the opponent. Losers believed that even more so than those who won.
  + No real difference between high or low performing losers however.
  + Estimation of number of difficult tasks were not robustly associated with selfish behavior in losers, or altruistic behavior for winners.

Conclusion

* People who form subjective beliefs on fairness act more self-serving and selfish, providing explicit information on fairness reduces this.
  + This has no effect on winners.
* This information about ‘fairness’ can reduce the “moral wiggle room” people use to justify selfish behavior.
* The absence of explicit information about the procedure of a competition may have similar effects on selfish behavior, especially to those who are ‘losing’
  + How can this be applied to distribution of healthcare goods? The ‘procedure’ by which individuals are selected to be able to get care?
  + Arranging fair procedures and practices is vital to justice/ethics, but INFORMING people of this procedural fairness is ALSO key!
* Many performance outcomes can be due to factors outside of control, which reasonably leads some to conclude that losing wasn’t their own fault, but by reinforcing how ‘fair’ the playing field is, we reduce selfish behavior.