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’