The Golden Rule & The Prisoner's Dilemma – Reciprocity and Non Zero Sum Play May Teach an Adaptive AI to Be Kind.

A Conceptual Introduction

It is difficult to define "kindness" without recourse to some notion (implicit or explicit) of *reciprocity*, wherein kindness to another is linked to kindness to one's *self*. This reciprocal notion is perhaps more universally known as The Golden Rule (formally as The Law of Reciprocity) which states: 'Do unto others as you would have them do unto you.' A more refined version of this rule (some have asserted, a more "sadist proof" version) re-phrases it in its "negative" form: 'What you do not want done to yourself, do not do to others.' Let us refer to these as GR 1 and GR 2.

Now, either of these formulations requires a notion of 'self' or The Self, which is perhaps the crux, the 'holy grail' of Artificial Intelligence: *self-awareness* (the prerequisite for a concept of The Self). Programming or engineering a self-reflective capacity into a would-be AI would seem to be a gargantuan task (for this challenge, and, for the field of AI in general). But perhaps there is a roundabout, or *proxy strategy*, for achieving something like this 'self-awareness' such that a given AI – properly trained (more on this later) – may learn an ethical "understanding" and/or behavioral repertoire.

Specifically, if the AI assumes alternate "sides" in a competitive game (akin to "playing a role") in which said roles are routinely reversed (being both giver and recipient), thereby learning the value of, if not the meaning of, the Golden Rule (of reciprocity).

Further, if we also expand (or redefine, or illustrate) *the meaning of kindness* to include well-defined behaviors observed in a certain type of reciprocity game (behaviors that can be modeled or represented mathematically), I speculate that this strategy make provide the right footing for a computational (the scientific basis of AI) form of ethics.

The Prisoner's Dilemma (PD) – A game of cooperation and defection – a proxy for reciprocity

Technically, in what's known as Game Theory, the game referred to here is called Iterated Prisoner's Dilemma (PD) first developed by M. Flood and M. Dresher of the RAND Corporation. In IPD, repeated encounters (a long series of game "rounds') with an opposing player generally favors cooperative behavior (cooperation) in which both players gain something (typically, equal amounts of money) though this is not required by pre-ordained rule. A player may choose to "defect" (defection) to gain "unfair" monetary advantage.

There are several versions of the PD, but in its general form, the game works as follows: Two players. One player receives a specific amount of play money (one that is unknown to the other player). The player who goes first can split the money evenly by keeping half and giving half to his opponent. This is the cooperative strategy. But the trick is: the second player doesn't know for sure what amount the other player was secretly given. The second player must trust the first player. The two players engaged in many rounds (iterations) of play in which each takes turns going first. In some versions, after each round (each player has a turn giving and receiving), the results (the monetary totals) are revealed. In

others, only at the end of the game and each player's total is revealed (and this is not always required in some PD games) is it revealed how much cooperativity (and how much defection) occurred. In the Flood and Dresher IPD experiments, players achieved "mutual cooperativity" 60 % of the time. This result comports with much later, and much larger, experiments.

Not all PD Strategies Reflect the "Spirit' of the Golden Rule

Does cooperation, or "cooperativity" in this context, approach some notion of "kindness"? It certainly bears resemblance to how we define or visualize *reciprocity*. But this basic game behavior may be a ways off from providing a model (or its basis) for kindness. To see why it could be problematic, we need to take a better look at this dilemma game.

In the 1980's, Robert Axelrod organized a large IPD tournament for computer programmers. After many thousands of rounds, the winning-est strategy turned out to be *tit-for-tat* (also known as The Rapoport strategy, after Anatol Rapoport, who participated).

We can see that *tit-for-tat* is clearly a form of reciprocity. In this winning strategy, the first 'receiver' always cooperates; thereafter, that player copies what the other player did (i.e., if the other keeps more money and gives less to you, then next time, you do the same to her/him). A person may defect (i.e., *not cooperate for mutual, equal benefit*) if s/he thinks that a monetary advantage (greater wealth) can be gained by defection. But this behavior can be "punished" through reciprocity. Clearly then, a society in which members are frequently defecting from the expectations of that society (generally, cooperative behavior) and/or being punished for this defection, is not the ideal society nor the basis for an ethical model that might lead to *some notion* of kindness in an AI.

William Press and Freeman Dyson, in their researches on the IPD game, discovered what they termed a *memory-one* strategy in which a *long memory of previous moves* by your opponent (say, the last 100 moves), wherein you might surmise his next most probable move, is NOT an advantage in the IPD; the "short" memory player-strategy (wherein only the last previous round is recalled) has the same effect on the opponent. In deed these *memory-one* strategies can take complete control of the game (and thus your opponents score) and manipulate it as desired...for a while.

Zero-Determinant Strategies and Beyond

In short, this type of strategy can *force your opponent to cooperate* (or, perhaps even force him to defect when it's next her/his turn). Memory-one strategies are denoted by four numbers that represent the probabilities of each pairing of PD choices or "moves" [see: NOTE, below]. These strategies are also known as dictatorial and/or extortionist and/or coercive strategies. But even these more "powerful" player strategies can succumb to their initial success and evolve the game towards dominance by other strategies, such as the Pavlov strategy [see; NOTES, entry 2, below]

NOTES:

- There are four outcomes in IPD (Dyson and Press): *cc, dc, cd, dd* (where '*c*' = cooperate, '*d*' = defect); both cooperate, one defects/other cooperates, one

cooperates/other defects, both defect, respectively. The **tit-for-tat** strategy yields probability results of 1, 0, 1, 0, respectively. An 'extortionist' *zero-determinant strategy* that forces equality of scores. But this is also a *memory-one* strategy.

- As another example of a zero-determinant strategy, the **Pavlov strategy** yields probabilities of 1, 0, 0, 1 – cooperate when you and your opponent make the same choice (*cc, dd*), defect when you and your opponent make opposite choices (*dc, cd*). [see the cited article in *American Scientist* for more strategies and their probabilities].
- In a certain subset of zero-determinant strategies, only the probabilities for *cc* or *dd* are "free choices"; these determine the probabilities for the variant combinations (*dc, cd*). Press and Dyson observe that these former strategies allow for a one-sided encounter where in the "stationary state" (the state at any given moment in time, following a round of play) of the PD game is controlled *entirely by one player*, and note that this "leads to much mischief". Indeed, it undermines the ideal situation of mutual cooperativity (the presumed basis of our ethical model).

Darwin's Necessity

Researchers Adami and Hintze showed that, from an *evolutionary* perspective (note: keep this in mind as we are talking about human morality or ethics (underpinning our notion of kindness) which has evolved out of the ancient pro-social behaviors of primate societies) this initially "victorious" strategy can lead quickly to it dominating the evolutionary game, being copied over and over in the general population. When this happens, we encounter more people like ourselves (who display the same coercive/dictatorial behavior) with the result that the dictators and extortionist types can no longer thrive or achieve further gain (for everyone else is copying them). This leads to *evolutionary instability*, which in social/societal terms means that society is on the verge of "collapse". We may conclude from this research that "winning isn't everything." In fact, too much winning by one player, or a few of his/her followers, can be an evolutionary dead-end.

An 'eye for an eye' leaves the whole world blind

These *zero-determinant strategies* may generally be characterized as "eye for an eye" type behaviors. Extortionist and/or coercive strategies may indeed "work" (for a time) until others figure out what is going on, realize that the "social contract" has broken down, and dare to defect themselves, or just copy what the dictator/extortionist does. This situation quickly devolves into a "state of play" in which most players no longer derive equitable benefit (i.e., one player achieves unfair gain); continuation of the game (or, the social exchange) becomes pointless (this is akin to the "stationary state" noted by Press and Dyson, earlier).

> NOTE: One other advantage to the memory-one strategy is that the calculated frequencies of choices (the number of *cooperative* and *defecting* moves, in a defined run of IPD) all converge on a *stationary state* (i.e. the predicted result/outcome of the IPD run). Thus it is not even necessary to run an IPD-type program (i.e., run thousands iterations of the game, tracking each

move) to reach this state – it can be calculated directly [see: article by B. Hayes and NOTES, **page 4, entry 1**].

In some cases, this unfairness and defection from the social norm during a single (competitive) encounter takes the form of a "zero-sum game"* or "winner take all" outcome in which one "wins" only if everyone else, or here, the other player, loses. This cannot be the basis of a functional human society in which we engage in *repeated encounters* with others (who remember how we treated them in the past), regardless of our bravado talk of Social Darwinism ("survival or the fittest") or related applications of Hamilton's Law.** In its basic expression, I call this the "there can be only ONE" mentality, which typical of most ritualized competitions (e.g., sports) for mass entertainment.

* A recently popularized term derived from Game Theory; it is a paraphrase of one of famed mathematician John Nash's economic game equations). The 'Nash Equilibrium' refers to a stationary state in a *non-cooperative game* in which no player ('Alice' and 'Bob') can benefit by changing their strategy and while the other player's strategy does not change. Perhaps a variant of this equilibrium (in calculus form) could provide the basis of an "ethical algorithm" [see; **RATIONALE, page 6**, and **https://en.wikipedia.org/wiki/Nash_equilibrium**]

Hamilton's Law (also known as the Social Law) of "genetic fitness" can be expressed as an inequality, in which a member of a herd (a social grouping) may *cooperate altruistically* if the benefit (*B*) provided to that member (and weighted by its genetic relatedness, denoted as *r*) exceeds its reproductive cost (*C*) to that member (wherein the member may lose out on chances to reproduce due to the altruistic sacrifice, Thus, ***rB - C > 0*.). This law underpins most examples of emergent biological complexity in animals from ants to elephants. [see: Martin A. Nowak and Roger Highfield, 'SuperCooperators: Altruism, Evolution, and Why we need each other to Succeed', Free Press, 2011]

Do unto others – Mercy v. Justice.

It may seem like fair justice to repay an "eye for an eye". However, in groups or societies that depend on other member's for survival, *mercy* (possibly the equivalent of the memory-one strategy, or, what this author calls a "forget and forgive" strategy, a slightly altered *memory-one* strategy) is a more effective strategy – allowing for the maximum number of members of society to benefit, and thus continue the cooperative "play" that makes a *functioning* society possible.

An AI heuristic with an algorithm designed to perform this calculation can quickly determine the outcome(s) -- and the probability of each outcome – thus determining if the strategy is more or less "pro-social" (roughly defined here as "mutually cooperative") compared to another. But, if one wishes to play out a number of rounds of PD just to be sure, one can do so (i.e. run an IPD program). Remember, the results of these referenced IPD experiments come with large data sets.

NOTES:

- In two-player IPD players can be represented as player X, and player Y, and their respective scores denoted by S_x and S_y. As Brian Hayes has shown [see: citation, **page 6**], one "mischievous" strategy manipulates the ratio between the two scores, such that player Y may impose a linear relation on X's score ($S_y = 1 + M\{S_x - 1\}$, where M is an arbitrary constant greater than 1. Player X has the option of always defecting, but this limits both players to a minimal "pay-off" of just 1 point. Any attempt by X to improve his return will increase S_y 's return by *M* times Thus, this strategy is "extortionist" as defined by Press and Dyson.
- If our putative AI algorithm (which is yet to materialize) is guided by a pre-programmed "ideal", pro-social strategy (e.g., mutual cooperativity or reciprocity), and/or a proxy (or analogue definition) of kindness that comports with the most beneficial IPD strategy, then we may approach the designing or programming of an ethical-behavioral repertoire...and from there we build up a notion of "kindness". More on this later.

Enter the Super-Cooperators

Stewart and Plotkin conducted a series of evolutionary experiments –using a "generosity" subset of zero-determinant strategies (*generosity* being another possible proxy for 'Kindness"). With exception of very small populations (< 10 persons), the researchers found that the generosity strategy is the most "robust" strategy and quickly proliferates throughout the populations (comprised of diverse behaviors, not all generous or kind). Its robustness includes the benefit or ability to "repel" other, "invasive" behaviors (such as the dictator or extortionist strategies). In a quasi "turn the other cheek" type phenomenon, the team found that it "pays" (i.e., benefits all) to put up with a degree of selfishness (which may be viewed as defection from the mutual cooperative social strategy); those who practice the *always cooperate* ("Super Cooperator") strategy clear a path for others to follow and in doing so, create opportunities for mutually cooperative behaviors and beneficial opportunities. Thus, I speculate that the society based upon such super cooperation is a thriving society, and one more likely to survive long-term (which is the "point" of evolution, and thus, is positively selected for).

Back to Basics – *Do Unto Others*

Reciprocity underpins social cooperation (pro-sociality) which in turn underpins the *composite* behavior we call Kindness (i.e., kindness is composed of many possible behaviors). In describing these zero-determinant strategies, I have try to show that forms of reciprocity (one of which we seek as a basis for our theory of kindness) can be rendered algebraically and rendered as a logical/rational set of rules (algorithm?) and which may form the basis of an AI heuristic. Perhaps so, but perhaps it doesn't matter (like the short memory strategy). There's a reason I chose the IPD as a training "model" (and all PD research has large data sets with which to train an AI upon) for It should be noted that every move/choice in PD can be rendered in algebraic form (thus computational form, thus algorithmic). So, if we can find an appropriate proxy (or analogue) for human Kindness within the behaviors strategies of the Prisoner's Dilemma, then we may say that we can "program" kindness – or an ethical analogue of it – into an AI.

NOTE: The Nash Equilibrium, in which no additional move(s) can produce gain or loss for either player. This is roughly congruent with the positive variant of the *zero-determinant strategy* of "generosity" (which also may be used as a proxy for "equality").

An Adaptive AI – Substituting Monetary Gain for Social Benefit

If we substitute money or "monetary gain" with "social benefit" or "social opportunity" (broadening our definition) we may reframe the Prisoner's Dilemma in purely social terms (as opposed to economic terms, which is its original framing). An AI doesn't know what money is (it only computes its utility in terms of a PD game scoring [see **NOTES, page 5**]). So if an adaptive AI instead views every social interaction as an ethical cooperation or defection "game" in which the goal is to maximize social benefit and opportunity for both (or all) participants in the interaction (the encounter) then we have a viable conceptual framework (and viable IPD data sets) for constructing an AI ethics of human (socially beneficial) interaction. This construction should be mathematically tractable and computationally feasible.

In the IPD, two players engage in multiple rounds of play (in theory, any number of players can participate in IPD). We can imagine one being our "Self" and one being the "Other" (we can call these players A and B, or X and Y). We humans only see or *learn* how our *Self* interactions (in the contest of IPD) impact the *Other* when this opponent copies our strategy of interaction (assuming that our 'Self' goes first) and the game evolves from thereafter as described (assuming that, as children, we never watched a program that taught generosity, sharing, and kindness). The beauty of training an AI with IPD is that the AI can "assume" (take on) both A and B roles or strategies (*X* and *Y* may be more conventional designations for rendering PD strategies into algebraic form). Our AI need not have true "self-awareness" to reflect upon the games' outcomes, as it can easily calculate the probabilistic outcome(s) of each strategy taken up as it plays both roles [see: prior NOTES, **pages 2, 3, 4, 5**].

Towards and Ethical Algorithm

If the basic, programmed goal is *long-term, mutual benefit*, an AI trained on this ("redefined") IPD can "learn" what the most beneficial strategy is for both (or all) players in the long-term. This may serve as the platform for constructing an AI "notion" of ethics (from which a notion of "kindness", reciprocal benefit, may emerge).

An old maxim in human morality and ethics goes: *do not judge someone until you have walked a mile in their shoes* (or in this case, tested her PD strategy). An IPD-trained AI can be both 'Self' and 'Other' (even simultaneously, through parallel running of multiple iterations of PD). In the context of the idea of reciprocity (e.g., *tit-for-tat*), we can see that the Prisoner's Dilemma is a reciprocity game that can be rendered algebraically, and, that its analogous (cooperative) expression in the Law of Reciprocity ('*Do unto others…*') is a nearly perfect analogue formulation (let's call it 'higher level code') of this PD reciprocity. We therefore have the rudiments of an "ethical algorithm" (possibly an "algorithm of kindness"?), to be embedded in our AI ethics heuristic. This is more than some "kindness *ex machina*" or emergent evolutionary fantasy; it is the beginning of intelligent, ethical learning – what, in some

rudimentary form, human infants (and even human-domesticated dogs) must master to become *ethical beings in this world.*

REFERENCES / CITATIONS (for this proposal and for learning data sets):

The main reference for the IPD content of this proposal is the *American Scientist* article:

> **'New Dilemmas for the Prisoner'** by Brian Hayes. [*American Scientist*, November - December, 2013; pp, 422-425]; main cited references in the above articles' bibliography (with **usable/viable data sets**):
>
> > Adami, C and A. Hintze. 2012. 'Evolutionary instability of zero-determinant strategies demonstrates that winning is not everything'. *Nature Communications* 4:2193.
> >
> > Press, W. H. and F. J. Dyson. 2012. 'From extortion to generosity, the evolution of zero-determinant strategies in the Prisoner's Dilemma.' *Proceedings of the National Academy of Science of the USA*. 109:10409-10413.

Additional References (**with mathematical formulations and data sets**):

> William H. Press and Freeman J. Dyson, 2012. 'Iterated Prisoner's Dilemma contains strategies that dominate any evolutionary opponent'. *PNAS*. June 26, 2012. 109 (26) 10409-10413; Paper links: http://www.pnas.org/content/109/26/10409 or https://doi.org/10.1073/pnas.1206569109
>
> **Stewart, A. J. and J. B. Plotkin. 2013. 'From extortion to generosity, the evolution of zero-determinant strategies in the prisoner's dilemma.' *PNAS of the USA*. 110:15348-15353.**
>
> Links to paper: http://www.pnas.org/content/110/38/15348 or https://doi.org/10.1073/pnas.1306246110
>
> **Figures/Graphs and Supplemental Information (data sets): http://www.pnas.org/content/110/38/15348/tab-figures-data**