Thank You for Playing: A Rational Choice Theory of Friendship

Elena Barham

Advised by Fabrizio Cariani

Spring/Fall 2015

Word count: 11,788

I. Introduction: Can Friendship Arise from Self-interested Rational Behavior?

Friendships play a central role in our lives. We attend social functions we might prefer not to, perform inconvenient favors, curb our more critical opinions, all on behalf of the happiness of our friends. This all seems necessary if we assume that friendship is, in some indirect way, worth it due to the unconscious reactions of our biological reward system. However, if we look past a biological imperative and treat humans as rational agents, where does that leave us in terms of understanding friendship? In this section, I will explore existing theories relating to the conditions and purpose of friendship, and its relationship with rationality.

To evaluate the rationality of friendship, we must assess the conditions that justify sustained altruism to a friend; in a sense, understanding how individual preferences function and interact within a friendship. Existing theories of friendship suggest that friendship grants both pleasure and usefulness to the friends. Telfer defines the phenomenon of 'liking' a friend as a "quasi-aesthetic attitude, roughly specifiable as 'finding a person to one's taste.'"[1] However, Telfer acknowledges that this conception of engaging in friendship because one 'likes' one's friend is separate from the bond of friendship. One can imagine individuals that one 'likes,' but are not friends, and similarly have friends that (at times) one does not 'like'. Moreover, the phenomenon of 'liking' seems to not necessarily be rationally justifiable, as two very similar people could be 'liked' or 'disliked' for subtle reasons unclear even to the individual who holds the preference.

Another existing theory highlights the ability of a friend to teach us more about ourselves, that perhaps a close friend makes certain understandings epistemologically accessible to us when it would not be otherwise. Grounded in an Aristotelian discussion of a friend as 'a second self', Cooper posits that one can view a friend objectively enough to discern faults they would not find in themselves. More concretely, Cooper states that "one recognizes the quality of one's own character and one's own life by seeing reflected, as in a mirror, in one's friend."[2] In this argument, the ways in which an individual fails to objectively assess their own character can be made accessible to them through similar behavior they analyze objectively in their friend. In this

---

[1] Telfer, E., 1970-71, "Friendship". *Proceedings of the Aristotelian Society,* 71. 226.
[2] Cooper, J.M., 1977, "Friendship and the Good in Aristotle", *Philosophical Review,* 86: p.299.

way, friendships are valuable, perhaps necessary, in that they improve our self-knowledge in ways that could arguably be rationality enhancing. Although this seems plausible, such an intimate understanding of the shortcomings of a friend's character would be territory of a mature friendship. Seeking to know oneself better fails to explain initiating a friendship, and moreover doesn't clearly support the more frivolous interactions that form the gateway to a deeper friendship.

A final account of the rationality of friendship could be that friendship is rational through a 'character-relative' or 'ethocentric' model as proposed by J.E. Whiting[3]. Friendship, in this view, is based on valuing another individual's character because they are a certain type of person, such that one considers this 'type' good and one seeks to emulate this 'type' as well.[4] This line of reasoning holds that friendship's rational justification relies on an individual clearly valuing their friend for the reasons the individual loves herself, but avoids the egoist critique by arguing that they don't value merely that their friend is like them. Rather, both they and their friend are like a certain ideal type of person (in terms of character and values). However, true character and deeply held values are difficult to assess accurately prior to beginning a friendship. It seems unlikely that individuals at the outset of a friendship could really understand these essential traits about their friends. Therefore, this theory provides an explanation for continuing a friendship that is consistent with a rational choice point of view, but likewise fails to justify beginning a friendship; examining friendship in rational choice terms should illuminate the cooperative mechanisms throughout the scope of a friendship.

The reasons behind friendship as a concept likewise complicate this investigation. C.S. Lewis posits that friendship relationships are the *least necessary* of all of our relationships. In *The Four Loves*, Lewis argues that friendship is free from biological necessity, unlike romantic, kinship, and charitable love.[5] Following this line of thought, rational choice theory would have to answer why we ought to pursue friendship at all? What about friendship justifies its existence, especially if this comes at the expense of time individuals could be putting into other

---

[3] Whiting, J.E. 1991, "Impersonal Friends", *Monist*, 74. p.7.
[4] Ibid., p.7.
[5] Lewis, C.S. *The Four Loves*.

relationships? This paper will offer a framework to interpret these questions, assessing how our self-interest reflects in our friendships, and in turn, how our friendships manifest our values.

If we posit that friendship demands altruism, that in friendships we must be willing to subjugate our own immediate desires and prioritize those of a friend, it becomes clear that there must be a clear individual benefit to friendship in order to rationally justify these choices. One possible theory of friendship compatible with a decision-theory framework would argue that sustained mutual altruism over time evens out or increases the overall utility of both individuals. It could be possible that there are things one friend appreciates more than they cost to the other friend to perform. Perhaps my friend giving me advice costs them a small amount of time and patience, but their advice significantly enhances my ability to navigate a situation. Through exchanging these 'social goods', each friend has a better experience than they would achieve on their own. Thus, although individual altruistic acts may penalize the individual performing them, creating a disposition that encourages the friend's altruism becomes to one's overall advantage.

This could happen in a few ways. The first would be immediate gratification—that friendship more or less amounts to a kindness exchange. In this version of friendship, each act of altruism is repaid within a reasonable time horizon; sustained relationships are cases in which two individuals are willing to meet each other's payback expectations. The second possibility for this could be more long-term, that although friendship demands preliminary sacrifice, there could be a long-run reward to friendship, which the friendly individual thinks merits pursuit. Examples of this could be eventually having an individual to care for one when one falls sick, or to whom one leaves their children's custody in the case of a tragedy, who in some sense 'owes it' to them to fill this role.

Another explanation for the rationality of friendship falls in line with the valuation of a friend as its own end. In this version, moral, ethical, or other personally held normative judgments compel an individual to form friendships because they feel that they *should*, that it is the *right* action to take. A concrete case of this would be the situation in which an individual maintains social contact with a family friend, due not to organic interest in this person but rather due to a personal belief that she should honor and maintain relationships that are important to her parents out of love for her family. Acting in this way comes with the pleasure of living aligned

with one's principles. In this way, partaking in a friendship supplies utility that makes altruistic acts rational, all things considered. Given this explanation, friendships form between individuals who feel that acting altruistically towards one another is compliant with their moral values, and not overpowered by other preferences.

A final defense of the rationality of friendship would be that each individual act of altruism could be explained as constructive of the characteristic dynamics of a friendship, which over time will grow to be its own source of utility for the partaking individuals. In this view, a friend performs a favor for the friend instrumentally, as a signal to demonstrate trustworthiness, kindness, competence, or some other trait that they feel is important to their friend. Altruistic acts are rational insofar as they construct expectations and mutual understanding so that friendship can develop, with the friendship being enjoyable to the friends in its own right. Over time, these interactions enable both individuals to experience greater happiness than they would absent the existence of the friendship. This long-term enhanced enjoyment makes it rational to partake in the friendship throughout.

As shown thus far, assuming a rational, self-interested framework for decision-making complicates any explanation for the phenomenon of friendship. This paper will take seriously the requirements of rational choice theory as a tool for analyzing human rationality in both descriptive and action-guiding ways. To be clear, rational choice theory describes the broader field of study that comprises decision theory, game theory, and social choice theory. Applying these concepts to friendships provides an avenue to explore the rational motivations behind social behaviors, and what sort of values or beliefs this reveals about individuals who partake in friendship. Moreover, decision theory applied to these questions can serve an action-guiding purpose, allowing understanding about ways in which friendships ought to be pursued.

Looking ahead, this paper uses decision theory to analyze individual decisions within a friendship, and propose a game theory model through which to understand friendship as a series of sustained, cooperative, mutually altruistic interactions. Game theory provides a necessarily complement to decision theory, as friendship involve multiple agents, and game theory provides better mechanisms to deal with choices that depend on pair dynamics. This paper contributes an understanding of how agents sustain friendship given the self-interest and rationality

requirements of rational choice theory, and explores what this can tell this about our relationships and ourselves. I explore a rational choice model to evaluate the rationality of friendship, conceived as a stochastic, extended-form game, and create a stylized toy model to explore generalizable strategy equilibria. Finally, this paper argues for deep, robust friendships (based on Kantian and Aristotelian notions of such) as the most rational relationship an individual can form.

II. Background: Basic Concepts of Decision Theory and Game Theory

Rational choice theory, which includes decision theory and game theory, is a technical fields that demand clarification before delving into its relationship with theory of friendship. This section should provide for a working understanding of both to facilitate understanding throughout the paper. Combining the two subfields allows for an analysis that deals better with both the interpersonal and intrapersonal aspects of friendship than either theory could alone.

Game theory and decision theory both employ a framework in which certain choices are likely to lead to certain *outcomes*. A quintessential example of this can be as simple as the set of choices {wear raincoat, leave raincoat at home} having potential outcomes {don't get wet if it rains, get wet if the rains}. These outcomes are linked to subjective valuations from the decision-making agent, who has their own unique set of values and preferences. The subjective value an agent places on a certain outcome is referred to as *utility*. Utility measurements express the agent's valuation of the outcomes, where a higher utility corresponds to a greater enjoyment of the outcome. Utilities are treated as an interval scale, in which a single unit of utility (a *utilon*) denotes a consistent amount of utility throughout. This would mean that, in terms of raw utility, an agent prefers a 6 utilon outcome to a 5 utilon outcome by the same amount as they prefer a 13 utilon outcome to a 12 utilon outcome. It is worth noting that interval scales are not divisible, so there would be no sense in saying that one prefer a 12 utilon outcome twice as much as a 6 utilon outcome.

Using these concepts, both game theory and decision theory define rationality as pursuing a course of action that *maximizes expected utility*. This means that rational agents should act in ways that they believe are likely lead to the outcomes they prefer. Expectation is key, as individuals have imperfect information about future occurrences, and can only evaluate choices in the contexts of *potential* states of the world, and how likely they believe it is that each state will occur.[6]

Given this focus on maximizing expected utilities, we describe the quality of outcomes largely based on how they compare to other outcomes. Imagine an individual, agent A, choosing

---

[6] Briggs, Rachael, "Normative Theories of Rational Choice: Expected Utility", *The Stanford Encyclopedia of Philosophy* (Winter 2015 Edition), Edward N. Zalta (ed.).

between two bike routes to get to the grocery store. One road is poorly maintained (route B), the other well maintained (route C). Agent A also worries about traffic levels. Given that agent A feels strongly about not falling off their bike, she must assess what would be the route most likely to help her arrive at her destination safely. Table 1 describes this decision:

	Traffic	No Traffic
Bike on route B	5	8
Bike on route C	6	10
Stay home	4	4

Table 1. Biking and Road Conditions Decision Matrix

Demonstrated by comparing these utilities, the utility-maximizing strategy is to bike on the better road. To ascertain this, we calculate the expected utility of each possible action. These expected utility calculations (for an N state decision matrix) follow the general format $EU(x) = (P_1)(U_1) + (P_2)(U_2) + ... (P_N)(U_N)$ where P_i refers to the probability of the outcome and U_i refers to the utility of that outcome. In this example, suppose a 50% chance of traffic. We then calculate the expected utility of biking on route B as: $EU(B)=(.5)(5) + (.5)(8) = 6.5$; similarly $EU(C)= (.5)(6) + (.5)(10)=8$. Such expected utility calculations form the general basis for evidential decision theory.[7]

A final concern with this scenario would be the argument that traffic is more likely on the better road. Probabilistic relationships between actions and states of the world require us to use *conditional probabilities* in the expected utility calculation.[8] Here, we use probabilities of traffic conditioned on the road. Suppose there is a 90% chance of traffic on the better road, and a 20% chance of traffic on the worse road. The expected utility of biking on road B would then become $EU(B) = (.2)(5) + (.8)(8) = 7.4$; correspondingly, in this case $EU(C)=(.9)(6) + (.1)(10) = 6.4$.[9] Taken together, the concepts of utility and maximization assessed through agent actions and

---

[7] Jeffrey, Richard. *The Logic of Decision*, second edition, Chicago: University of Chicago Press. 1965.

[8] Weirich, Paul, "Causal Decision Theory", *The Stanford Encyclopedia of Philosophy* (Winter 2012 Edition), Edward N. Zalta (ed.).

[9] Any discussion of conditional probabilities opens itself to deep philosophical questions regarding how to interpret this conditionality. For a fuller discussion of these questions, see *The Foundations of Causal Decision Theory* by James Joyce (1999).

states of the world provide a basis for understanding rationality as conceptualized by decision theory.

Game theory, which also falls under the umbrella of rational choice theory, expands from decision-theory's single agent point of view to dynamic systems involving multiple agents. The most basic game theory unit, a *normal form game*, is defined by a tuple (N, A_i, u_i) where N is the set of players, A_i denotes a finite set of actions available to player i, and u_i denotes the set of utilities available to player i based on the outcome arrived at through their and others' choices. Normal form games assume simultaneous action by each player, based on ignorance of the other player's actual choice but knowledge of the incentives faced by the other players. The most fundamental games are played once, achieve an outcome, and end. However, *extensive form games* allow for games to play out over time, with sequential outcomes in each subgame. Extensive form games can include either perfect or imperfect information. In *perfect information games*, players know exactly the choice they face, as well as the choices facing the other player. Games can likewise have *imperfect information*, a structure I will employ but modify in analyzing friendship.

Repeated and stochastic games also provide a way to organize games with a time horizon. Within *repeated games*, a game is played over and over, repeated finitely, infinitely, or indefinitely. *Stochastic games* are a class of repeated games in which the outcome of the previous game corresponds to a probabilistic distribution which determines the next game played. Concretely, an example of this would be a cooperative outcome in Game A corresponds with a 50% chance that Game A is repeated, and a 50% chance that Game B is played. In a stochastic game of this format, the game played after Game A would reflect the corresponding probabilities. Stochastic games utilize a set of normal form games, determining the game played at any time probabilistically based on the previous game and the actions taken by the players.

An agent's strategy profile, then, consists of all the strategies available to an agent in a given game. These strategies fall into two broader categories. *Pure strategies* indicate that a player chooses one action with certainty. *Mixed strategies* involve an agent randomizing over the available actions. Concretely, a mixed strategy could be as simple as a student flipping a coin to decide whether or not they will study for a test. Mixed strategies tie the agent's choice to a

random occurrence with a certain probability distribution. These strategies are particularly useful in games in which one agent desires to coordinate and the other player wishes to avoid coordination. As mixed strategies exist on a distribution, there are infinitely many mixed strategies available to any player facing at least two choices, however there is generally only one utility-maximizing mixed strategy.

When dealing with extensive-form games, strategies take on greater complexity, as these games grant each player the opportunity to teach the other player about herself. Therefore, a player's strategy refers to their plan for the entire form of the game, rather than just the current iteration. These strategies involve desires to influence each other's strategies; each player attempts to shape the other player's strategy through behaving in ways that encourage certain choices from the other. For example, in an infinitely repeated game one player might find it advantageous to punish their opponent harshly early on (even at a loss to themselves) in order to convince their opponent to choose in ways that player finds favorable later in the game.

A final important concept in understanding the operation of game theory is the idea of a stable outcome, referred to as a *Nash equilibrium*. *Nash equilibria* contain the choice that each player makes if they knew with certainty the strategies of the other agents. Formally, Nash equilibria follow this definition:

(1) Let (N, A_i, u_i) be a game with n players, A_i possible actions, and u_i associated utilities.

(2) Let S_i be the strategy set for player i, and $S = S_1 \times S_2 \times \ldots \times S_n$ be the set of all strategy profiles.

(3) Let $f = (f_1(x), \ldots, f_n(x))$ be the payoff function for all $x \in S$.

(4) Let x_i be the strategy profile of player i and $x_{\sim i}$ be the strategy profile of all players except for player i.

(5) Nash equilibria choices (x_i^*) occur under the condition: $\forall i, x_i \in S_i : f_i(x_i^*, x_{\sim i}^*) \geq S_i : f_i(x_i, x_{\sim i}^*)$. [10]

Less formally, Nash equilibria are situations of no-regret, in which each player has done the best they can, given the strategies of other players.[11] No player can unilaterally improve his

---

[10] Giocoli, Nicola. "Nash Equilibrium." *History of Political Economy*. 36:4 (Winter 2004). p. 639-666; p.639.

[11] Binmore, Kenneth. *Game Theory: A Very Short Introduction.* Oxford: Oxford UP. 2007. p. 14.

or her outcome. In extensive form games, players achieve ‘strategy equilibria’ as well as outcome equilibria. In these cases, equilibria may be a set of recurring actions by the agents, rather than the outcome interpretation of equilibrium present in one-time games.

III. Preliminary Philosophical Concerns with Game Theory and Friendship

Game theory applies well to topics such as chess or international relations in which players have well defined interests and clear beliefs about situations. But can it be applied to something as subjective and seemingly irrational as friendship? In interpersonal relationships, passions, offense, and momentary frustrations alter our decision-making and add a degree of impulsiveness that game theory seems not to capture. It seems likely that in our interpersonal relationships we choose based on intuition or impulse, and then later invent a rationale to defend our choices to ourselves. Fortunately, a clear explanation of what is meant by utility, as well as setting an 'interaction-significance' threshold can help avoid these concerns and improve the framework for this exploration.

The first key understanding for this model of friendship is to understand the inherently subjective nature of utility. Unsophisticated game theory might assign utilities based on a supposed objective evaluation of a payoff. Such utilities would resemble payoffs U_1=-10 for ten years in prison and U_1=-5 for 5 years in prison in a standard Prisoner's Dilemma. Decision theory, rather, demands subjective valuation of payoffs; the value of an outcome depends on the preferences of the agent. Treating utilities in a more decision theoretic, subjective way, allows for the preferences of the agent to be captured in the values assigned. For the purposes of this paper, I assume that the utilities assigned reflect individual preferences correctly, but do not specify the underlying preferences unless it is useful to the discussion, as values and preferences vary between individuals and likewise influence friendship and compatibility. Although these are certainly relevant questions, incorporating individual values and their justifications would make this investigation overly complex while reducing the scope of my model.

Second, it seems plausible that friendship involves a series of instantaneous decisions, such as how to reply in a conversation and whether or not one laughs at a friend's joke. These smaller interactions could account for perceptions of 'conversational chemistry' or 'compatibility' that contribute to divergences between genuine connections (friendships) and minor acquaintanceships without the potential for growth. For the purpose of this investigation, I will ignore these smaller interactions, for the following reasons. Primarily, these instantaneous perceptions of friendship compatibility may somehow account for a subset of attempted

friendships, making them a potentially necessary condition for friendship, but are not independently sufficient to ensure that the players are compatible as friends. Secondly, although a large amount of interpersonal communication takes place in these brief and subtle interactions, this line of thinking supposes a deterministic model of friendship. In doing so, individual values and long-term preferences become secondary to impulsive reactions and unrelated to behavior and outcomes. Empirically, psychologists have documented these reactions as deteriorating in rational quality with 'decision fatigue', a necessarily irrational phenomenon.[12] For these reasons, this impulse-bound conceptualization of friendship reduces down to an argument in which individuals are insufficiently conscious to apply values and long-term objectives to smaller interactions, which is not an argument I wish to engage; it seems overly pessimistic about the decision-making capacity of individuals to assume that friendship cannot be consciously chosen. With these two points in mind, I will now propose my model for analyzing friendship through the rational choice framework.

[12] Tierney, John, 2011. "Do you suffer from decision fatigue?". NYT Magazine Online.

IV. Definitions and Preconditions for Friendship

What sort of friendship is rational? What sort of outcomes would we consciously commit ourselves to if we had perfect information regarding the scope of a friendship? Who can partake in a friendship? Existing philosophical work can ground an understanding of what friendship means, how it manifests, who can partake in friendship, and thoughts about the role of rationality.

This paper conceptualizes friendship based mainly on Aristotelian and Kantian perspectives to orient the model. Aristotle defends a theory of friendship based on a division between friends for utility and unconditional friendships. For Aristotle, friendships of utility rely on each individual having their needs met by the friend, and thus the friendship's continuance relies on the other individual staying instrumentally valuable.[13] Unconditional friendships, rather, arise between two virtuous individuals who wish to do good for the other's sake, and bring out the best in and for each other.[14] It is worth noting that both of these friendships are rational, as both friends derive a positive utility from partaking, but that the second type of relationship is a more robust sort of friendship. Aristotle also notes friendships of pleasure, in which the other person's company is pleasant, and this provides the basis of the friendship. I argue that this is not particularly different from a friendship of utility in terms of choice, as you would likewise terminate the friendship if you cease to enjoy their company.

Kant posits a similar division, labeling these categories as friendships of need, of taste, and disposition.[15] Kant's 'friendships of need' are similar to Aristotle's 'friendships of utility', and likewise his friendships of taste are similar to Aristotle's of pleasure. This is not to say that the two do not differ at all. However, understanding these friendships through a rational choice framework makes it clear that these genres of friendship are clearly concerned with the immediate outcomes more so than with deeper traits or value correspondence between friends. Kant's friendship of disposition, however, involves true friends that know each other deeply and

---

[13] Aristotle. *Nicomachean Ethics: Books VII and IX. Other Selves: Philosophers on Friendship.* Edited by Michael Pakaluk. Indianapolis: Hackett Pub., 1991. 28-69. Print., p. 36-37.

[14] Aristotle, Op. Cit., p.36.

[15] Kant, Immanuel. "Lecture on Friendship." *Other Selves: Philosophers on Friendship.* Edited by Michael Pakaluk. Indianapolis: Hackett Pub., 1991. p. 208-217. Print. p.212.

genuinely, and both have complete trust that the other will help them to the best of their ability whenever they truly need it.[16]

Thinking about these notions of friendship, it seems clear that one can't rationally commit to either Kant's or Aristotle's more profound variety of friendship in the early stages of a friendship. These varieties of friendship demand deeper knowledge and understanding of the friend than is accessible at the preliminary stages of a friendship. Rather, the individuals can begin a friendship based on what they believe to be the case about the other person, holding these beliefs subject to revision over the course of the relationship. Thus, necessary conditions for friendship involve two individuals who mutually perceive that the other seems sufficiently likeable to begin interacting. From here, rationality relies on the interaction between the individuals and how that informs their beliefs about the other and shapes the dynamic of their relationship. With this in mind, I posit a rational choice model for friendship.

[16] Ibid, p. 215.

V. Proposition: A Basic Rational Choice Model of Friendship

To ground the model in a concrete example, let's imagine a simple case of a friendship beginning. Amy and Ben are college freshmen who have (along with the rest of their class) just arrived on campus for the beginning of the school year. New to college, nobody knows anybody else particularly well, and most everyone would like to make friends within this group. On the first day of class, Ben decides to sit next to Amy. Recognizing him from their residence hall, Amy greets Ben and asks him how his morning was. This interaction demonstrates cooperative socialization: Ben extended a social gesture by choosing to sit next to Amy over sitting by himself or with someone else. Amy cooperated with this by engaging Ben once he sat down. It is foreseeable that because of this interaction, Amy will talk to Ben after class or sit with him if she runs into him in the dining hall. If this interaction had gone differently—whether Ben had not chosen to sit next to Amy, or Amy had looked annoyed with Ben once he sat down, it would probably make it less likely that they would later interact in a friendship context.

Looking at this interaction through game theory, we would consider this a successful attempt at response coordination, which resulted in both players having a more friendship-positive beginning to class than they would have otherwise. Both Ben and Amy could discern a genuine friendliness and mutual friendship interest in each other, which gave them a friendship payoff from the interaction and increased the likelihood of continued friendship games. They both, likewise, exposed themselves to a small amount of risk (Ben risked Amy's annoyance that he sat next to her, and Amy likewise risked being shut down when she asked Ben about his day). Their mutually positive handling of this interaction increases the possibility of later interaction, as Ben and Amy begin to form an opinion of the other that includes their preliminary friendliness and cooperative behavior.

Now imagine a different example. Camille and Denise are trapeze artists in a traveling circus. They were both recently hired by the circus, have never met before, and are part of a large group of trapeze artists who all joined the circus on the same day. The choreographer has required that Camille and Denise execute a trapeze act together. This routine involves both of them performing acrobatics, relying on each other to ensure that they do not fall. Assuming competence, any error in the routine would be due to negligence or ill will between them.

However, their executing the routine successfully does not necessarily enhance the probability of a friendship outcome for Camille and Denise. Although they both exposed themselves to risk at the hands of the other person, their cooperation was given as there would have been significant repercussions for Camille and Denise had they failed to ensure the other's safety. Failure for either would be extremely costly for her own career; their actions are determined by the incentives of the situation, not due to care for the other specifically.

This example allows us to discern a few qualitative differences in defining friendship through game theory. Both of these cases exhibit similar characteristics: individuals expose themselves to risk and rely on the other individual to cooperate in order to ensure a mutually beneficial outcome. The probability of further interactions (games) increases if this game has a cooperative outcome. However, in the case of Amy and Ben, they co-operated specifically with that person, and began to voluntarily form a relationship based on genuine choice to take socially 'risky' behavior in pursuit of connecting with the other. Camille and Denise co-operated because it was designed in the payoffs of the current game, rather than because of an interest in the other person or a strong personal belief. My game theory model of friendship will center on these ideas of intentional altruism specific to another person where the outcomes of each situation help determine future interactions.

To capture person-specificity, I posit a model of friendship in which what matters is not the precise action, but how the action undertaken on the friend's behalf seems *to that friend*. In the context of the example, had Ben sitting down offended Amy, his action would not have been friendship-enhancing. Whether an act is considered friendship-enhancing or friendship-harming will depend a good deal on the subjective preferences of the individuals involved in the friendship. Although there may be more common preferences, such as loyalty or honesty, defining what I believe all individuals to seek in the context of friendship would necessarily reduce the applicability of the model. In setting up the decision matrix for friendship interactions, therefore, I will divide the universe of all possible actions into two zones, each defined by what the other player would prefer the individual do in the situation in question. The first zone is defined by failing to meet the friend's expectations for the other player's actions (*disappointment zone*). This includes the other player making the choice that the player would have preferred they

not make with respect to the friendship. The second zone includes option that pleases the individual beyond what they expected out of their friend in that situation (*growth zone*). This, similarly, involves the action that the friend preferred their friend take in that situation.

Two potential objections to this idea follow: firstly, one could object that there are more than two relevant actions in each decision. I allow that this is a possibility, however I view my formulation as a necessary simplification, and that otherwise the model quickly becomes too complicated and unwieldy for inference. Secondly, one might object that there are likely a number of possible actions that would neither impress nor disappoint the friend. However, I hold that games with this outcome maintain the status quo in the friendship, and thus are in a sense irrelevant to friendship development (and therefore I will not consider them).

Table 2. The Basic Friendship Game

	Growth_1	Disappointment_1
Growth_2		
Disappointment_2		

Note: The subscripts on row or column identifiers refer to the player for whom these subjective expectations are defined.

Within this game theory model, major strategic interactions in a friendship follow essentially this structure. Both friends face a universe of possible actions, each of which fall into one of these categories. An iteration of this game could be collaborating on a homework assignment, choosing to attend an event together, or any other shared activity in which friends partake. The iterative structure of the game relies on common-sense understanding of the way in which friendships progress. Relationships develop as a series of progressive interactions, which lends itself naturally to a long-form iterated game.

The actual payoffs that face friends within the structure rely heavily on the Kantian notion of displaying a friendship disposition through demonstrating willingness to prioritize the friend's well-being. Kant identifies several types of friendship to help orient these payoffs. The first Kantian friendship is a 'friendship of need', in which friendship "presupposes a benevolent disposition and a helping hand in need, and on the other abstention from abusing it by making

calls upon it".[17] The seconds relevant Kantian friendship is the 'friendship of disposition or sentiment', a "friendship in the absolute sense" in which an individual feels comfortable with complete personal disclosure to the other friend.[18] Based on these ideas, I argue that to display a disposition of friendship, a player must undertake voluntary altruistic acts on behalf of their friend. Likewise, acts that are convenient to a player but demonstrate less kindness to their friend would be a disappointment strategy. The following table formally demonstrates these ideas:

Table 3. The Basic Friendship Game, with payoffs:

	$Growth_1$	$Dissappointment_1$
$Growth_2$	M_1-C_1, M_2-C_2	M_1, N_2-C_2
$Disappointment_2$	N_1-C_1, M_2	N_1, N_2

This structure includes intuitive payoffs (M, X, and N), which denote the optimal outcome, the status quo outcome, and the sub-optimal outcome respectively. These payoffs must satisfy the inequality "M>N". This merely signifies that growth outcomes grant the friend a higher degree of utility than disappointment outcomes. The coefficient C denotes the cost of altruism (any inconvenience or diminished utility a player undergoes to act in the other player's interest, traditionally part of the definition of altruism). I will restrict C to values greater than 0 (C>0), as it necessarily denotes a cost to the agent that performs it. This 'cost of altruism' draws on Kant's argument that self-love and love of the other are necessarily in conflict,[19] as well as traditional definitions of altruism as necessitating a cost on the individual performing the altruistic act.[20] Finally, note that these payoffs include the most objective view of the situation. There are no other emotional reactions included (such as a betrayal coefficient for when you are altruistic and your friend is not). This is to avoid specifying the beliefs, values, and desires of the agents within the construction of my model, as this would necessarily reduce how broadly this model can be applied.

---

[17] Ibid, 213.
[18] Ibid, 214.
[19] Ibid, 210.
[20] Okasha, Samir. "Biological Altruism", *The Stanford Encyclopedia of Philosophy* (Fall 2013), Edward N. Zalta (ed.).

A thick understanding of this model necessitates an understanding of each iteration from the player's subjective perspective. In order to choose rationally, the player considers (a) the present payoffs and (b) the effect of their choice on the friendship dynamic and associated future payoffs. Finally, the friend assesses their own level of confidence that their action will have the intended effect (be perceived by their friend in the way the player means it). Through considering present and future payoffs (utilities) and assigning probabilities to these outcomes occurring as a result of their action, the friends can perform an expected utility calculation of the choice at hand. All of these things depend on the beliefs a player has regarding the friendship, which I will expand on below.

Within friendship, we learn about our friends from our interactions with them. Rational individuals accumulate experience that allows them to form expectations about their friend's behavior, beliefs, values, and character. For the purposes of this model, I will treat each friendship separately, assuming that interactions within a friendship affect that particular friendship, and that other (prior) interactions shape the expectations with which an individual enters the friendship. Formally, a set of actors (A, B) comprise the friendship, and outcomes of games between these two players affect their relationship (the iterated game between A and B) but not other relationships. 'Learning' refers to agents updating their beliefs about the other agent, and the friendship, based on these sub-game outcomes. Both agents enter the friendship with expectations based on prior relationships, however their expectations update to include interactions with their friend. I will not deal with the expectations from previous relationships as a dynamic within friendship, but rather consider them part of a necessary set of traits for being a 'player' in potential friendship games. This requirement is necessary because actual social life involves actors within a web of relationships, and would be absurd to simultaneously model all of these different potential pairings in describing individual behavior.

To understand the longer-term structure of the game, imagine the simple example in which one friend (agent A) invites a friend (agent B) to go to on a vacation together. If agent B says 'yes' to the invitation, and enthusiastically helps plan the trip and they have a great time, this would constitute a growth zone outcome for agent A. If agent B enjoys agent A's company, the vacation, and was pleased that this plan came to be, this constitutes a growth outcome for

agent B. This positive experience for both players should make each more likely to invite the other to activities in the future. Moreover, they took a risk in pursuing this collective action, and this risk should shape their expectations towards greater trust in their friend. Within this example, a positive outcome in the first game (attending the play) makes a second game more likely to occur. In this way, friendship best imitates a stochastic game, a game in which each game is probabilistically dependent on the outcome of the previous game. The stochastic format implies a probabilistic distribution of possible future games but not a guaranteed next iteration. This structure makes sense due to the simple truth that friendships do not develop in a vacuum. The other events and obligations in people's lives create an environment that can crowd out even a promising friendship, and likewise a lack of distraction can cause certain friendships to go on longer than they perhaps should. As it would be hopelessly complicated to attempt to measure the comprehensive set of pressures facing any one individual, expressing a friendship continuing as probabilistic with this element of randomness accounts for these other pressures in a reasonable way.

This model seems to overlook a type of friendship interaction in which only one friend acts. An example of this would be when a friend is upset about something unrelated to the friendship, and the other friend steps in to cheer them up. In this instance, the first friend perhaps has no strategy, whereas the second decides what they would like to do in that instance. However, even within these interactions there are a number of ways the friends both choose their courses of action. The apparently action-less friend can reject their friend's support, or can make clear their vulnerability and appreciation. In this sense, every interaction between friends is necessarily two-sided, and can be modeled as such. Moreover, the former line of argument about one-sided interactions includes the possibility that all coordination games could also be expressed as a series of two one-sided games. This misses the notion that coordination games entail imperfect knowledge of the other player's strategy. Likewise, in coordination games, players face ambiguity that must be filled by prior knowledge of the other player's strategy, trustworthiness, and other traits. This uncertainty and reliance on prior knowledge is often very relevant in social situations, especially as relationships become important to the individuals

involved. Friendships can legitimately be expressed as coordination games given this condition of imperfect knowledge.

The stochastic distributions I propose should seem largely intuitive. Mutually optimal outcomes increase the likelihood of both players to engage in coordination games with both higher potential payoffs and greater potential for disappointment. This dynamic expresses what we would consider ‘trust’ in an informally conceptualized friendship. Rationally, a player risks disappointment if they believe that they will achieve the correspondingly higher outcome. Correspondingly, mutually suboptimal outcomes decrease the likelihood of future high-stakes games, and sufficiently negative outcomes increase the probability that the game will end completely.

VI. Preliminary Exploration: Mathematical Inferences from a Toy Model

In this section, I posit a stylized model to demonstrate several points regarding the rationality of friendship. This model utilizes a standard game theory format, and formalizes my earlier points regarding friendship as a long-form game in which both potential payoffs and the likelihood of cooperation increase as the players build "trust" through cooperative outcomes. This model provides insight into the dominance of cooperative strategies, at least until the late stages of a friendship game.

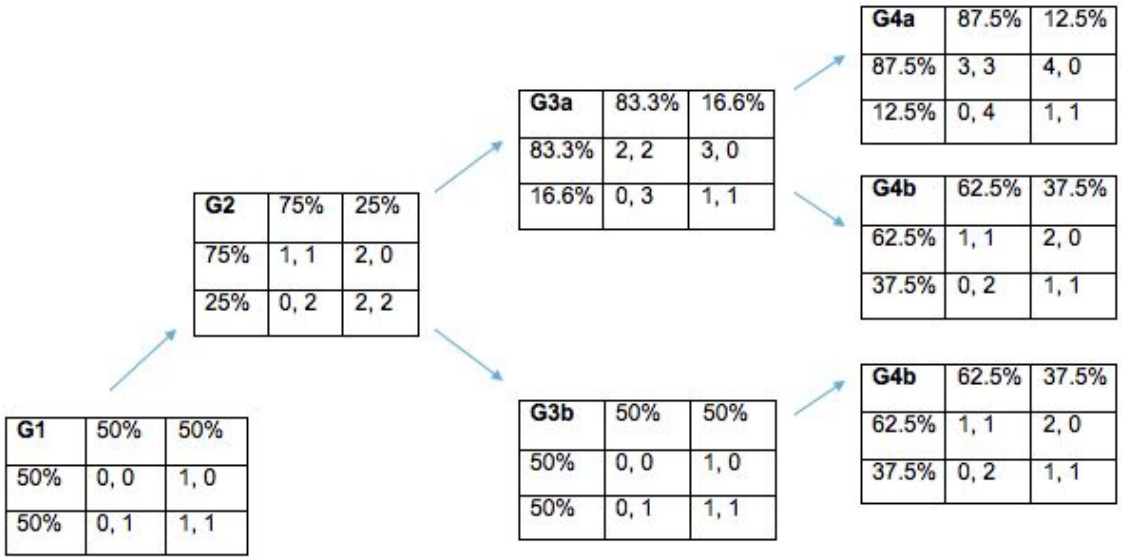
The diagram depicts a four stage friendship game (a truncated model of the hypothesized game of indefinite length). The payoffs (a, b) denoted in the games refer as (row player, column player). The boxes on the top and left side of the table denote the expectations each player should have for that game, given an outset belief (updated to reflect prior outcomes). Behavioral expectations are calculated in the following way: $P(\text{cooperation})=(P_1 + O_1 + \ldots + O_{n-1})/n$, where 'O' refers to the cooperative outcome of the indexed game, and 'n' is the iteration of the game currently in play. Games with a cooperative outcome are assigned a value of '1' for cooperation, while games with non-cooperative outcomes are assigned a value of '0'. I have chosen this rule as it strikes a balance between initial perceptions (factored in as a parameter) but also makes individuals responsive to their interaction in a way that allows them to adjust quickly to the player that they are playing. Although there are certainly underlying philosophical concerns about the strength of an individual's initial character judgement, I find it more appropriate to this framework to maintain players as rational and adaptive, and thus engaged in learning about the other player's cooperative tendencies through interaction. Predictably, using a calculation rule that favors the agent's initial judgement over situational learning biases the expected utilities in the direction of their initial perception; likewise, using a calculation rule that places more weight on the interactive outcomes gives prior outcomes a greater influence in expectations.[21]

Explaining the overall structure, games in which both players play 'growth' strategies lead to a game of higher payoffs; games in which either player disappoints lead to a game of lower payoffs (ie. if one player defects in G2, the players then play G3b rather than G3a). Within

[21] Disappointments from a friend so severe that they end the friendship definitively (such as, for example, discovering that the friend perpetrated a hate crime) are not handled well within this theory. The qualities and dynamics of such interactions deserve further exploration.

each subgame, this long-form game uses the variable identified in previous sections. The benefit from cooperation, M_i, is equal to 1 in G1 and increases or decreases by 1 for cooperation and defection, respectively. I will refer to the variation in payoffs between games as V_i. The payoff from defecting, N_i, begins at 1 and remains constant. The cost of altruism is set as C_i=1 throughout. Each player has an unbiased initial belief about the other player's likelihood to cooperate (P_1=.5). These are arbitrary values chosen to aid in mathematical inferences that can be drawn from the model; setting the outcome parameters as equal allows for simple relationships between variables to be analyzed.

Table 5. "Game A": a potential game following my proposed format



Within each subgame, this long-form game uses the variable identified in previous sections. The benefit from cooperation, M_i, is equal to 1 in G1 and increases or decreases by 1 for cooperation and defection, respectively. I will refer to the variation in payoffs between games as V_i. The payoff from defecting, N_i, begins at 1 and remains constant. The cost of altruism is set as C_i=1 throughout. Each player has an unbiased initial belief about the other player's likelihood to cooperate (P_1=.5). These are arbitrary values chosen to aid in mathematical inferences that can be drawn from the model; setting the outcome parameters as equal allows for simple relationships between variables to be analyzed.

The expected utilities for each available strategy in "Game A" are as follows:

Strategy Profile	Expected Utility
C, C, C, C	5.3125
C, C, C, D	6.3125
C, C, D, C	4.375
C, C, D, D	5.375
C, D, C, C	2.375
C, D, C, D	3.375

As is obvious from this table, the rational player should prefer the strategies in this order: *(C, C, C, D) > (C, C, D, D) > (C, C, C, C) > (C, C, D, C) > (C, D, C, D) > (C, D, C, C)*. This preliminary result illustrates that, for the game at hand, the strategies that involve cooperation in the first two games yield higher outcomes than a player who disappoints in this early stage. The highest expected utility occurs when the players cooperate throughout, and then can disappoint in the final game with a high degree of certainty that the other player will not expect that based only on the player's prior actions. Although this game is very specific, it begins to illuminate a more general principle: for a friendship game conceived with these parameters, a rational strategy delays defection until the last round *no matter how long the game*.

Playing with the values of different parameters of the game yields some further interesting points. Maintaining the initial condition (where $M_i=C_i=|V_i|$), the strategy profile (C, C, C, D) has the highest expected utility for all possible subjective assessments of the likelihood the partner will cooperate ($0\leq P_1\leq 1$). Counterintuitively, this means that even if the player expects with certainty that their partner will not cooperate at G1, the utility-maximizing strategy over the four-game strategy space should still be to play a cooperative strategy in the first round.

In reference to the other parameters, increasing $|V_i|$ preserves (C, C, C, D) as the most valuable strategy, with EU(C, C, C, C) > EU(C, C, D, D) for values $|V_i| > \frac{24}{13}$ for all ($0\leq P_1\leq 1$). Moreover, setting $|V_i|=1$, EU(C, C, C, D) remains the highest until for all cases such that $\frac{Ci}{Mi}\leq\frac{13}{12}$, in the case in which $P_1=0$. Increasing to $P_1=1$, $\frac{Ci}{Mi}\geq 2$ preserves (C, C, C, D) as the strategy with the highest expected utility.

Although extremely specific, these examples illustrate a few insights regarding strategy equilibria. Firstly, strategies that feature initial cooperation retain higher expected utilities even

under adverse conditions such as high costs of altruism (relative to benefits), low expected probability of cooperation, and low game-to-game returns to cooperation.

Even though this game presents a stylized example, it sets groundwork for extrapolation. As I have explored, the more cooperative strategies (C, C, C, C) and (C, C, C, D) have a higher expected payoff under a large range of conditions. More interestingly, this truncated game provides for generalization past this four subgame structure. In games with more iterations, the result will still hold that the highest expected utility outcome under certain constraints will always be strategy of cooperating until the last round. This is due both to the increased subjective probability of cooperation and the higher payoffs that accompany late-stage games. Moreover, in indefinite or infinitely recurring versions of this game, the highest expected utility will be a strategy that is *purely cooperative*. In the indefinite version, it is unclear whether a round would be the final round, and thus cooperation is a sustaining strategy. In the infinite version, there is no 'final' round (unless mandated by sufficient uncooperative behavior, which I have demonstrated would be irrational). With these long-term strategy equilibria of cooperation in mind, I will now argue for the rationality of a thicker, deeper notion of friendship complimented by this understanding of the game theory logic of cooperative behavior.

VII. A Rational Choice Theory of Friendship

In this section, I utilize the rational choice model of friendship to argue for the superiority of deep, intimate friendships as the most rational friendship to pursue. To accomplish this, I first provide a definition for such friendships in relation to the game theory model, noting the role of strategy equilibria in developing a friendship to its potential. I appeal to these friendships as being the maximally rationality-enhancing form of friendship, influencing overall decision processes and thus improving rationality of the individuals who partake in such friendships.

In the simplest game theory sense, friendship manifests as sustained mutual altruism in which both players achieve greater outcomes than they could alone. I base this on the Kantian conception of deep, genuine friendship as "the maximum reciprocity of love".[22] In technical terms, the reciprocated nature of altruism interactions demand that both players' strategies consistently play actions that are growth outcomes for the other player. However, I make another requirement of these friendships, likewise based on Kant. Kant's "absolute sense" of friendship describes individuals being genuine with each other, thus improving each other through this honest friend-relation.[23] For this reason, I require that agents notice how their friend perceived their action, the way in which their friend played, and update their knowledge of the game accordingly even if this does not immediately influence their strategy. Through this mutual awareness of the dynamics of the game, individuals learn and revise their strategies accordingly. I will touch on this idea throughout, noting that a friendship (over time) begins to approximate a perfect information game, although agents start out somewhat unaware of both their own and their friend's values, preferences, and outcomes in the situation.

A successful friendship, in game theory terms, would progress as follows. Both players choose strategies that correspond to growth outcomes for the other player. Each game, players update their expectations for the other player's strategy and, gaining an understanding of the other player's specific beliefs, values, and preferences. Through sustained mutual 'growth' strategies, both players come to believe that the other player will act in ways that protect each other's interests, and that the friendship should continue as long as they continue to act more or

---

[22] Kant, Op. Cit., p.211.
[23] Kant, Op. Cit., p.214-15.

less reciprocally. Through observing the other player's reaction to their choice, each player learns to judge with greater accuracy what strategies will please the other player, and why. Given enough iterations of the game, they will both become very competent at playing in mutually beneficial ways. Moreover, the trust dynamic (reflected in each friend's assessment of the likelihood of cooperation) will grow very strong between the players. As a result, they will feel comfortable communicating their beliefs with increasing clarity and placing themselves at more risk at the hands of their friend, as it becomes clear that their friend can be counted on to protect their interests.

I argue that the rational strategy for friendship is to be maximally altruistic, in a precise sense that I will define, establishing trust that allows friendship to be fruitful in mutual altruism payoffs and, more interestingly, rationality-enhancing for each player. I now demonstrate how these deep, genuine friendships function to rationally benefit both friends.

Consider the first iteration of the game. At this point, any expectation the players have about friendship resides purely in previous experience and personal belief. Were this a stand-alone game, the Nash equilibrium would be the outcome (Disappointment_1, Disappoinment_2), in which player exerts the minimum effort on behalf of the other and receives minimum effort in return. Thinking back to Game 1 in the toy model, this Nash equilibrium should be clear. Regardless of what the other player does, each player is always better off in that specific game by minimizing his or her own effort. However, once the time horizon expands to include indefinite interactions, this changes.

Over iterated prisoner's dilemmas, 'tit-for-tat' strategies, are the most successful strategy equilibrium. Generally, 'tit-for-tat' strategies are strategies in which one player cooperates as long as their friend cooperates, and defects immediately following their partner's defection. In a 'tit-for-tat' strategy, the player responds directly to their partner's previous move: if the partner defected, the player now defects; if the partner cooperated, the player now cooperates. A substantial amount of literature has been devoted to the superior success of this strategy in iterated prisoner's dilemmas over all other theorized strategies.[24]

---

[24] Leyton-Brown, Kevin and Yoav Shoham. *Essentials of Game Theory*. San Rafael: Morgan & Claypool, 2007. Print. p.51.

This is directly relevant for a rational theory of friendship as the friendship game closely approximates an iterated prisoner's dilemma in several important ways. First, the immediate payoff structure of a friendship mirrors that of a prisoner's dilemma as both players might be taken advantage of if they act in a way that is sensitive to the interests of the other. Although philosophers argue about the way in which friendships are valuable, they generally agree that friendship constitutes an important part of a high quality life[25], and therefore there is a real incentive to act in the interest of the other *if you believe the other will act in your interest as well.* Perhaps more importantly, mutually altruistic interactions provide the foundation for a potential friendship's progression, and thus have a real attraction for the agent. Secondly, friendships emulate prisoner's dilemmas in the symmetrical structure and joint preference for cooperation. Within both friendship and the prisoner's dilemma, it is unreasonable to expect the other individual to act on your behalf if you seldom act on theirs. Rational individuals would seldom seek to befriend an individual who antagonizes them. In both the friendship game and an iterated prisoner's dilemma, you play repeatedly against the same other player, establishing a dynamic of play specific to your relationship. In seeking to enforce the behavior of the other player, it is therefore important to make your strategy clear and predictable in order for them to understand it, thus making cooperative play rational for them.

However, friendship also differs from a prisoner's dilemma, creating a necessity to amend the use of a tit-for-tat strategy to include the rationality of avoiding retaliation when possible. Making oneself vulnerable communicates a disposition towards achieving robust personal friendships.[26] Similarly, the nature of the friendship game includes an 'existential imperative', in which part of the point of the game is *to continue the game*. Whereas in a standard iterated prisoner's dilemma the players do not know how long they play, the probability that a friendship will end depends largely on the behavior of the friends towards each other. In a friendship, players decide which games they feel comfortable engaging in. Disappointment strategies increase the likelihood that the game stops there. Friendship revises a tit-for-tat

---

[25] Helm, Bennett. "Friendship." *The Stanford Encyclopedia of Philosophy* (Fall 2013), Edward N. Zalta (ed).

[26] Not only does vulnerability communicate willingness to cooperate, it perhaps also plays to the idea of 'letting one's guard down' as a trait we might desire in our friends.

strategy as within a friendship there are two ways to punish the opponent: one is through defecting, the other not consenting to play another round. Given this additional choice, it becomes strategically optimal that players establish which actions are worth ending a friendship over, and punish those accordingly. One must be aware that, between rational players, defection signals at least the potential willingness to end a friendship, as in a context where both playing a retaliatory strategy one defection will lead to permanent retaliation. Thus, two key revisions of the tit-for-tat strategy occurs within friendship. Firstly, due to potential retaliation, agents should not be the first to play a disappointment strategy, unless they are willing to risk ending the friendship. Secondly, agents should only punish in situations where they would be indifferent between the friendship ending and that behavior continuing. The threat of ending the friendship makes a case for the optimal friendship strategy being less punishing than tit-for-tat.

Along these lines, the existential imperative of friendship game makes altruism rationally necessary because the altruistic action is a necessary condition for future payoffs and more profitable future games. An agent who is rational enough to account for long-term time horizons sees that a growth strategy includes the potential for future payoffs. From this analysis, it is clear that friendships should be conducted altruistically, as is necessary to achieve a stable, cooperative equilibrium.

One could object that overly altruistic acts towards a friend are likely to seem suspicious, as individuals may be averse to feeling as though their friendship has been ‘bought off’. However, this overlooks the essential idea that actions in a friendship are necessarily defined by their mutual perception. Thus, the optimally altruistic action must be in compliance with the friend’s preferences for genuinely necessary and appropriate altruism. This objection highlights the crucial notion within a friendship that both players must cultivate a genuine understanding of the other’s preferences and beliefs. This idea of genuine mutual understanding allows us to explore the way in which robust friendships are valuable in terms of the strategic accuracy and enhanced self-awareness achieved through genuine interactions.

Thus far I have argued that robust friendships are the best in terms of outcomes and strategy equilibria. Now, I argue that these friendships are also the most rational as they improve the players *as players*, increasing rational capacity through a deep understanding of the other,

and eventually improving self-understanding as well. Definitive traits of robust friendships act to improve access to personal information.

As a precondition for improving rational capacity, I argue that robust, authentic friendship demands *honesty*. Complete honesty could perhaps be unhelpful for friendship, such as harshly honest opinion, however in general authentic friendship demands that individuals present themselves (and their beliefs, etc) as they are to the friend. Within this model we see honesty in an individual's reactions to the actions a friend takes towards them or on their behalf. After each interaction, an agent understands whether what they did was a growth outcome or a disappointment outcome for their friend. As a heuristic, this sort of honesty is a necessary condition for authentic friendship as it improves both agents' access to rationally necessary information.

Inauthentic friendship would entail not the opposite but merely the *absence* of honesty. Agents might actively deceive or ignore their friends, but the most important is that in this situation, agents do not make their preferences and values perceivable. Intuitively, an inauthentic friendship means that the agents cannot develop a true 'knowing' of each other, as they lack access to genuine information about the other person and thus fail to accumulate accurate knowledge of the other. This impairs their ability to apply knowledge from past interactions in distinguishing between genuine growth strategies and genuine disappointment strategies. Inauthentic friends form an idea of the friend's persona's utility function, but not the friend's true preferences. As a result, the accuracy with which they assess the strategic implications of their actions is likely not to improve or only improve minimally over time with respect to their friend.

This notion of hidden preferences, therefore, gives rise to a clear split between the rationality of authentic versus inauthentic friendships. To clarify this connection, reconsider the basic friendship game:

Table 4. Revisiting the Friendship Game

	$Growth_1$	$Disappointment_1$
$Growth_2$	M_1-C_1, M_2-C_2	M_1, N_2-C_2
$Disappointment_2$	N_1-C_1, M_2	N_1, N_2

Within this game, both players face an incentive to disappoint, as irrespective of the other person's strategy this is the best payoff. However, in this situation, the divide between the 'Growth' and 'Disappointment' regions is defined in terms of both players' preferences. The action that defines each category is therefore defined not by the player who takes the action, but the player who receives it. Imagine an agent who plays a practical joke on the friend to help the friend alleviate stress during a tough week at work. Some percentage of people might laugh and appreciate the break from the tension of their work environment while others might become angry with the friend for wasting their time. In a long-term friendship, learning the friend's value structure should enhance the accuracy with which a player plays in different zones. Knowing which category the action is likely to fall into relies on knowledge about the friend's preferences. This improved accuracy justifies the honesty requirement of Kant's complete notion of friendship with respect to interpersonal rationality. However, this only accounts for interpersonal accuracy; I will now explore the intrapersonal rationality enhancement of genuine friendship.

To argue for friendship's rationality-enhancing value, I introduce the notion of a 'deeper' preference structure, of which individuals have imperfect information and limited access.[27] Robust, genuine friendships enhance individual rational capacity by granting individuals epistemic access to their deeper beliefs and preferences, and thus are valuable for the greater self-knowledge available to the players involved.

Rational choice theory relies on the notion that individuals assign subjective utility values to possible outcomes. Decision theory agents then maximize their potential outcomes according to these projected utilities. However, evaluating rationality in line with this framework entails resolving the question of how accurately individuals understand their own payoffs *before the actual experience happens*, as well as questioning how well individuals understand the relative strengths of any potentially conflicting preferences they may hold. It seems plausible that in many situations, especially personal or social experiences, we might not know exactly how an experience will make us feel until we have had it. In this sense, certain preferences and values are epistemologically inaccessible to us prior to experiencing them in practice. Similarly, societal norms may give us preconceived notions of how we ought to react or feel in these situations,

[27] For a fuller discussion of this idea, see *Transformative Experience* by L.A. Paul (2014).

obscuring the accurate anticipation of the 'payoff'. Imagine having a fight with a close friend, and being surprisingly upset as a result. Given an understanding of how upset one would become, one might not have instigated. In these situations, our subjective experiences once the outcome has occurred do not correspond perfectly to the utilities projected prior to the event.

Improving our rational capacity, as individuals, demands that we become more acutely aware of our own preferences and values, with respect to how they will make us experience outcomes in unfamiliar situations. Formally, improving rationality demands that individuals understand their own complex utility functions more completely and accurately. To be clear about the focus of this argument, improving rational capacity *is itself rational*, as it allows the individual to choose behaviors more likely to guarantee outcomes that satisfy their *true* preference structure.

In this light, friendships are rationality-enhancing as they force individuals to confront and understand their own preferences as they impact the friendship. Similarly, the pressures of friendship give these decisions more urgency and importance than the individual might place on them independently. To ground this in the model, consider the accumulated altruistic acts and positive interactions necessary to establish trust within a relationship. Once a friendship has established such trust, the player stands to sacrifice the pattern of sustained altruism if they break the pattern and play a disappointment strategy. Moreover, it is likely that at this point in the relationship trust has developed sufficiently that the conflict between self-interest and altruism calls upon some important beliefs or values. The friend facing this choice has a clear incentive to sustain the mutual altruism and preserve the friendship in order to protect their future payoffs. These pressures lend gravity to the decision, forcing the rational player to confront their own beliefs in contrast with those of their friend, ultimately arriving at the deep and conscious understanding of the situation and choice of action that the player can arrive at. Through these interactions which demand self-evaluation, a long-term friendship forces individuals to understand their preferences regarding loyalty, altruism, intimacy, and a host of other social values. Friendship may also help to ground pre-existing moral values in personal experience, or cause individuals to question these moral values in a way that leads to eventual clarity. Through

these repeated evaluative decision processes, accurate individual knowledge of one's own preferences emerge within sustained friendship interaction.

Perhaps even more important, however, is the notion that an independent player can only review their values and beliefs insofar as they themselves spot problems or inconsistencies. Robust friendships present an individual with a true understanding of the friend's values and beliefs. Insofar as these beliefs differ between friends, each friend makes these new values and beliefs accessible and relevant to the friend. Similarly, situations that place these values or beliefs in conflict create space for reflection. To demonstrate this, imagine a situation in which one friend is considering a personal choice that places their values in tension, a real life example might be a woman who asks a friend to drive her to an appointment to terminate a pregnancy. The friend in question might have a well formulated opinion about the ethics of this situation, or might not, but either way when this issue become part of the friendship they are forced to reconsider a number of extremely personal values and judgements in order to ensure that they make the correct choice. This decision process forces the agent to nail down areas in which their desires conflict, and understand at a deeper level what actually informs their decision-making. Notably, this line of argument can also apply to value-questions that fall outside the scope of direct friendship relevance. Although this example was ethical, deep and genuine friendships place every element of an individual's preference and value structure up for review through such experiences. Moreover, this review function is specific to close relationships, as discovering that one's friend holds a dramatically different ethical opinion or political belief makes one question one's own beliefs more than finding that a stranger holds these beliefs, precisely due to the trust and perceived predictability of a close friendship. These instances of belief conflict allow players to develop complete, revised sets of values and preferences that ultimately makes them better able to discern what they desire out of situations, and therefore achieving deeper self-realization in their strategies.

Finally, I argue that robust friendships are in fact the best relationship in which to evaluate these deeper beliefs and values due to their being the least biologically necessary relationship. In other close relationships, such as romantic or kinship relations, physical attraction or biological imperatives create pressures on the situation that are not present in

friendship. An individual may act against his or her own values to remain on good terms with a lover because of the erotic benefit this relationship grants them; likewise, a parent may act against their own values to further their child's welfare because of a more instinctive desire to see one's progeny succeed. Friendship relationships come without as much biological baggage, allowing the potential for a space of more values-based comparison. This is particularly salient in robust friendships where the individuals know each other intimately enough to credibly compare their value structures. The non-necessity of robust friendships provides individuals with an irreplaceable mechanism for understanding themselves without other factors diluting the decision process.

Therefore, as explored throughout, the rational choice model helps us understand how deep, robust, intimate friendships are the most valuable in terms of payoffs as well as in their capacity to enhance individual rationality. Through strategy equilibria, cooperative and other-regarding friends can improve their own outcomes. More fundamentally, maintaining and rationally evaluating these strategic equilibria facilitates greater knowledge of the other as well as oneself, in ways that make both friends more rationally capable individuals.

VIII. Further Exploration: Fungibility and Traits-Based Friendship

Having completed my major argument, I will finally turn to an debate in the literature on friendship to which my theory can add some clarity. This debate regarding friendship relates to how we should justify a friendship with a particular individual. One theory presents a traits-based model of friendship, in which we value certain qualities in our friends and these qualities justify the relationship. A second theory presents friendship as based in personal history and experience with that individual, where friendship is built on a shared past. However, these arguments both entail troubling results. If friendships are traits-based, then friends should be interchangeable to the extent that they share the relevant traits[28]. Justifying the friendship based on one friend's virtue or honesty (or other amiable trait) forces the friend to agree that they should therefore equally be friends with another person who shares the trait. This 'fungibility problem' presents a problem for individuals who would otherwise be considered rational but would not accept a replacement friend for a current friend, even if told their new friend would have the exact same traits.

Likewise, the personal history explanation of friendship arguably provides a rationale for the friendship, but fails truly justify it over other friendships in a rigorous way. Although the friendship may have emerged from knowing each other over this period of time, this fails to explain why this friendship should be more important than any other in an objective or externally verifiable way[29].

Given an understanding of friendship based on the model outlined throughout this paper, this debate arguably presents a false dichotomy by omitting key links between traits and history. As explored in section VI, resolving the game in a manner that looks like friendship can be expedited by shared traits and values which lower the cost of altruism and make friends better natural judges of each other's preferences and utilities. However, as discussed previously, different individuals may perceive different friendship traits in the same actions, or (similarly) perceive different actions as conveying certain traits. In this way, time provides a necessary means to observe and understand how certain desirable traits operate in any given friend. An

---

[28] Helm, Bennett. Op. Cit.
[29] Whiting, J.E., Op. Cit., 46.

individual seeking honesty in their friends relies on situations in which the friend could demonstrate the level of their honesty, in order to discern the genuine value of the friendship. In this way, a friend's traits become epistemologically accessible over the course of a friendship, but are not at the beginning. Understanding between friends develops in ways that can't be adequately summarized through mere trait description of the friend.

Secondly, if we take behavior as communicating underlying traits, individuals choosing a course of action based on the preferences of their friend may convey a trait that they wouldn't under other circumstances. To be clear, I do not mean to argue that all traits of an individual are circumstantial. However, friendships can shape individual behavior as each person begins to act in ways based not only on their preferences, but also on their desired friendship outcomes and the preferences of their friend. Because of this, an individual might express certain traits in the context of one friendship, and other traits in the context of another. In order for an individual to genuinely enact a trait, that trait must exist on some level in the individual exhibiting it. However, preference pressures from a friend should alter the balance of traits displayed by an individual insofar as that friendship shapes the individual's behavior. The possibility that a friendship alters the perceivable traits in an individual demonstrates that friendship history and individual traits are not oppositional theories for justifying friendship, but are actually inextricably linked in the formation of a friendship. Moreover, the past relationship is not an arbitrary or irrational justification. The intimate familiarity with what behaviors convey a trait for that friend enable accurate judgments (and thus, individual rationality,) and friends must achieve this complex understanding through sustained interaction over a period of time. Therefore, shared personal history enables an understanding of traits that can justify the friendship on both practical and ethical levels better than either theory does independent of the other.

Works Cited

Aristotle. *Nicomachean Ethics: Books VII and IX*. *Other Selves: Philosophers on Friendship.* Edited by Michael Pakaluk. Indianapolis: Hackett Pub., 1991. 28-69. Print.

Binmore, Kenneth. *Game Theory: A Very Short Introduction.* Oxford: Oxford UP. 2007

Briggs, Rachael, "Normative Theories of Rational Choice: Expected Utility", *The Stanford Encyclopedia of Philosophy* (Winter 2015 Edition), Edward N. Zalta (ed.),

Cooper, J.M., 1997, "Friendship and the Good in Aristotle," *Philosophical Review,* 86: 290-315.

Giocoli, Nicola. "Nash Equilibrium." *History of Political Economy*. 36:4 (Winter 2004). p.639-666.

Helm, Bennett. "Friendship." *The Stanford Encyclopedia of Philosophy* (Fall 2013), Edward N. Zalta (ed).

Kant, Immanuel. "Lecture on Friendship." *Other Selves: Philosophers on Friendship.* Edited by Michael Pakaluk. Indianapolis: Hackett Pub., 1991. p. 208-217.

Jeffrey, Richard. *The Logic of Decision*, second edition, Chicago: University of Chicago Press. 1965.

Joyce, James. 1999. *The Foundations of Causal Decision Theory*, Cambridge: Cambridge University Press

Lewis, C.S. *The Four Loves*. New York: Harcourt, Brace, 1960. Print.

Leyton-Brown, Kevin and Yoav Shoham. *Essentials of Game Theory*. San Rafael: Morgan & Claypool, 2007. Print.

Okasha, Samir. "Biological Altruism", *The Stanford Encyclopedia of Philosophy* (Fall 2013), Edward N. Zalta (ed.).

Paul, L.A. *Transformative Experience.* Oxford: Oxford UP, 2014. Print.

Telfer, E. "Friendship." *Proceedings of the Aristotelian Society* 71 (1970-71): 223-241.

Tierney, John. 2011. "Do you suffer from decision fatigue?". *New York Times Magazine Online.* Accessed 10/2/2015.

Weirich, Paul, "Causal Decision Theory", *The Stanford Encyclopedia of Philosophy* (Winter 2012 Edition), Edward N. Zalta (ed.).

Whiting, J.E. “Impersonal Friends.” *Monist* 74.1 (1991): 3-29. Web.