Team Reasoning, Framing and Self-Control: An Aristotelian Account

Natalie Gold, Kings College London

Author Note and Acknowledgments

I thank John Broome, Dan Hausman, Steve Pearce and Hanna Pickard for helpful comments on early presentations of this material, and Neil Levy and an anonymous referee, whose identity was later revealed as Jean Kennett, for helpful input on the paper.

This research received funding from the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013) / ERC Grant Agreement n. 283849.

Correspondence regarding this chapter should be addressed to Natalie Gold, Philosophy Department, King’s College London, Strand, London, WC2R 2LS, United Kingdom. E-mail: Natalie.gold@rocketmail.com

In the *Nichomachean Ethics,* Aristotle discusses the problem of how someone can intentionally act against their best judgment. The Ancient Greeks called this *akrasia*, or lack of control, and its opposite is self-control. Aristotle's example of akrasia involves someone who is tempted, against their better judgment, to eat something sweet. When the person eats, he is still "under the influence (in a sense) of reason" (1147b1-2).[1] Hence *self-control* involves the exercise of control over temptations. This usage of the term, which is fairly standard in philosophical discourse, contrasts with a modern colloquial usage, whereby the opposite of having self-control involves losing the ability to intentionally control one's actions.[2] In this paper I will adopt the philosophical usage of "self-control" to refer to over-coming temptations, whilst remembering that, as J. L. Austin said, we can help ourselves to two portions of dessert without ravening and that, "We often succumb to temptation with calm and even with finesse" (Austin, 1956/7, p. 198).

Aristotle gives a causal account of akrasia, as located in the reasoning process of the akrates. For Aristotle, reasoning is characterized as a syllogism, involving a major and a minor premise. The major premise is a universal principle and the minor premise is a particular one about the situation at hand. In book VI, Aristotle has declared that the second, minor premise is the "starting point" of reasoning (1143b4-6) and that the universals are reached from these particulars, of which we have perception.

Aristotle illustrates the reasoning of the akrates as follows:

Particular premise: This thing is sweet

Universal premise: Everything sweet ought to be tasted

Conclusion: This thing ought to be tasted[3]

Opposed to this is a second syllogism, whose conclusion is that the akrates should not taste the thing. But, according to Aristotle, the akrates uses only the universal and not the particular premise of the second syllogism. Although the akrates "has" knowledge of the particular premise s/he does not "use" it (1146b31ff), where "using" should be understood as thinking of the premise or having it before one's mind (Bostock, 2000 p.125-6).[4] Hence Aristotle says that the akrates is ignorant of particular facts that are "within the sphere of perception" (1147a26).

[1] All quotations from the Nicomachean Ethics refer to the translation by Ross in Barnes (1984) and the standard Bekker page reference is given.

[2] The modern usage is at some points implied in Aristotle, for instance when he says that the akrates "acts with passion and not choice" (1111b13). In determining Aristotle's position, much turns on the interpretation of "choice". But my aim is not to provide an accurate reconstruction of Aristotle, rather it is to explore an interesting idea that can be divined in his work. (For an interesting reconstruction that is broadly sympathetic to the ideas in this paper, see Moss (2009).) Interpretive issues, including interpretations that are at odds with those that I use, will be confined to the footnotes.

[3] It is generally thought that Aristotle intended that the conclusion is also an action. For a minority dissenting view, that the conclusion of the syllogism is not identical with the action, see Charles (1984), and for an argument that the conclusion of the syllogism cannot also be an action if akrasia is to exist, see Wiggins (1980).

[4] Some commentators favour an interpretation where the proposition that is known but not used is the conclusion of the second syllogism (e.g. Charles, 1984; Urmson, 1988).

However, Aristotle does not specify the content of the second syllogism, the particular facts that the akrates does not perceive, or the premise that s/he does not use. The Ancients were more concerned with the puzzle of how akratic action could exist than with how people exercise self-control.

This contrasts with decision theory, whose framework suggests that people will always give into temptations and where the puzzle is to explain how people exert self-control. In decision theory, problems of overcoming temptation are naturally modelled within the framework of *dynamic choice,* where one person makes a sequence of decisions over time. It is conventional to analyse problems of dynamic choice as if, at each time *t* at which the person has to make a decision, that decision is made by a distinct *transient agent*, 'the person at time *t*'. Each transient agent is treated as an independent rational decision-maker. In this framework, self-control is a problem of *diachronic consistency,* where the profile of choices that seems best from the point of view of an early transient agent relies on a choice by a later agent that will not seem best from the later agent's point of view. The early transient agent would like to implement a plan that she cannot rely on her later self to carry through. Standard examples of the problem are going on diets, giving up smoking, and studying for an exam.

One objection to the decision theoretic account is that it provides a neat model of temptation at the expense of an impoverished notion of agency. Agency is entirely vested in the transient agents; there is no notion of a self that extends over time. It is as if every transient agent asks "What should *I*-now do?". The instruments for achieving self-control are limited to *pre-commitments,* taking actions that constrain the choices or alter the incentives that will be available to future selves; making a resolution in the hope that it will affect future behaviour is "naive" (Strotz, 1955-56). There is no sense of an extended agency over time, whereby earlier selves make plans for the person over time which influence their later selves because of their status of plans (as opposed to because the later self would have taken that action anyway).

In this paper, I place the decision theoretic account of temptation within a framework that allows multiple levels of agency and use this to show how the rational agents of decision-theory can achieve self-control. The idea of multiple levels of agency has been articulated at the inter-personal level, where theories of team reasoning allow agents to ask "What should *we* do?" and identify distinctive modes of reasoning used by people in teams (Bacharach, 2006; Sugden, 1993). I apply team reasoning at the intra-personal level, modelling the self as a team of transient agents over time. The resulting account of self-control is Aristotelian in flavour, in that it involves reasoning schemata and perception, and it is compatible with some of the psychological findings about self-control.

In order to motivate the application of team reasoning to the problem of self-control, I begin by showing how the problem of self-control can be seen as a prisoner's dilemma. The prisoner's dilemma set-up is not essential to the idea of the person as a team over time, but it is an important aspect of the decision theoretic puzzle of self-control. The prisoner's dilemma is also one of the puzzles of game theory that motivated the development of team reasoning, so analysing the problem of self-control as a prisoner's dilemma makes a very clear the analogy between team reasoning in the inter-personal and the intra-personal cases.

1. Self-control as an Intra-Personal Prisoner's Dilemma

In the prisoner's dilemma, individual agents must choose between two strategies, commonly called *defect* and *cooperate*. It is in each individual's interest to defect, but if everyone defects that leads to a worse outcome than if all had cooperated. For instance, pollution is an example of a prisoner's dilemma, with *defect* being to pollute and *cooperate* refraining from polluting.

The reason the problem of pollution has the form of a dilemma is that there is an *externality,* where each person's action has effects on other players that are not captured in the player's own payoff. To illustrate with an example: the prisoner's dilemma is actualized experimentally in "public goods games" with the cooperative action being investing in a group account (Ledyard, 1995). Money in the group account is multiplied up and shared out amongst all the players, with the contributor receiving less back than she put in, but the group as a whole receiving more. If there are n players, and f is the factor by which the money is multiplied, $1 < f < n$. So contributing to the group kitty has benefits for the other players, as they receive more money as a result of the agent's investing. But the agent who invests a pound bears a cost of £1 and only gets back £f/n, which is less than £1.

We can see that the problem of pollution has the same basic structure. Each individual might prefer the outcome where the air is clean (all *cooperate*) to the outcome where the air is polluted (all *defect*). However, the cost of not polluting (or benefits that the non-polluter forgoes) is born entirely by the individual whereas the gains of cleaner air are shared between the whole community.[5]

Problems of self-control can have an analogous structure. Take an individual who is a smoker.[6] Each transient agent might think that it is better for herself, over her lifetime, to be a non-smoker rather than a smoker, as that reduces the risk of cancer. In order to implement this plan, the transient agent must bear a cost, namely forgoing the enjoyment of smoking a cigarette. However, the benefit of being a non-smoker is not all captured by the transient agent. We might think that it is shared across transient agents or even that it accrues entirely to the transient agents at the end of the individual's life. The benefit of not smoking in any period is an *externality*, it is not completely captured by the agent that period.

The plan that the agent would most prefer to implement is to smoke a cigarette today and give up from tomorrow onwards (which is analogous to the case of pollution, where the most preferred outcome is to pollute whilst everyone else refrains) but that plan is not

[5] With a very large number of individuals and small incremental benefits, which are hence spread very thin, this can result in the benefits from any one person's action being imperceptible. For discussion of imperceptible benefits see McCarthy and Arntzenius (1997). But it is the dilemma structure that is key to the incentive structure of pollution, regardless of whether or not the benefits are imperceptible.

[6] Nicotine is an addictive substance. Arguably, addiction is simply a species of self-control problem (Heyman, 2009), but any reader who thinks that addiction adds extra complications to problems of self-control should either assume our smoker is not an addict, or transpose the example to a case that clearly does not involve addiction, such as studying for an exam.

realizable because the transient agent of tomorrow will face the same preference structure and, hence, will not play her part. So we have a prisoner's dilemma with smoking equivalent to *defect* and refraining to *cooperate.*

The analogy between the inter-personal and the intra-personal cases is not exact. In the intra-personal case the players do not move simultaneously, it is an *asynchronous prisoner's dilemma*. For this to change the analysis, it is not sufficient that the agents move in sequence. It is also necessary that the second player knows what the first player did. If the transient agents have perfect recall of past moves then, in some respects, a better analogy for the intra-personal case is found in Hume's two farmers, who play an asynchronous prisoner's dilemma.[7]

> Your corn is ripe to-day; mine will be so to-morrow. 'Tis profitable for us both, that I shou'd labour with you to-day, and that you shou'd aid me to-morrow. I have no kindness for you, and know you have as little for me. I will not, therefore, take any pains upon your account; and should I labour with you upon my own account, in expectation of a return, I know I shou'd be disappointed, and that I shou'd in vain depend upon your gratitude. Here then I leave you to labour alone: You treat me in the same manner. The seasons change; and both of us lose our harvests for want of mutual confidence and security. (Hume 1739-40/1978, pp. 520-521)

In an asynchronous game, the second player's strategy can be conditional on what the first player does, so she has four strategies instead of two: *cooperate regardless, defect regardless, cooperate if player one cooperates and defect if she defects, defect if player one cooperates and cooperate if she defects*. In the inter-personal prisoner's dilemma *defect* is a *dominant strategy,* whatever the other player does each player does best by defecting. In the asynchronous dilemma, (*defect, defect regardless*) is the sole Nash equilibrium, however it is not a dominant strategy equilibrium since player two does equally well if she plays the strategy *cooperate if player one cooperates and defect if she defects.* But it is usual to expect Nash equilibrium strategies to be played and, as Hume shows in the farmer example, backwards induction also leads to both agents defecting.

Smoking makes for a nice analogy with pollution, but the framework of costs that are born by the current transient agent for benefits that are, at least partly and maybe wholly, accrued by later agents, can also accommodate other classic examples of self-control, such as dieting and studying for an exam.

It might be complained that, so far, I have assumed that each transient agent cares only about the here and now whereas, plausibly, they would also care about the past or future, i.e., they would exhibit some intra-personal altruism. However, it is equally plausible that people exhibit some "present bias", with each transient agent giving herself relatively more weight than the other transient agents. Indeed, there is evidence of present bias (Ainslie 1992, Thaler, 1981) and also of "hyperbolic discounting" (Frederick, Loewenstein & O'Donoghue, 2002), which is closely related to present bias (Ainslie 1991, 1992; Laibson, 1997). Even if there is some degree of intra-personal altruism, if each transient agent gives herself more weight than the other agents, then the agents fail to take into account fully the positive externality. Hence there is still likely to be an intra-personal prisoner's dilemma -

[7] For more on the asynchronous prisoner's dilemma in Hume see Vanderschraaf (1998) and Skyrms (1998).

as can be seen in the model of decision-making over time with some intra-personal altruism provided by myself and Robert Sugden in our conclusion to Michael Bacharach's *Beyond Individual Choice* (Bacharach, 2006).

Further, displaying intra-personal altruism is different from agency. Decision theory can accommodate altruism, or a concern for another agent's outcomes, which is usually modelled by a *payoff transformation* where the payoffs of the other agent appear in the altruistic agent's utility function (e.g. Collard, 1978). But agency involves planning (Bratman, 1987) and identity (Parfit, 1984, p.319), neither of which appear in standard decision theory. Being altruistic does not necessarily solve problems of planning and agency, nor lead to co-operation in a prisoner's dilemma.[8] The payoff transformation involved in altruism is not enough, introducing "agency transformation" is also needed. In general, payoff transformations and agency transformation lead to different classes of results (Bacharach,1999). Hence the development of theories of team agency.

2. Reasoning: Team Agency and Self-Control

The prisoner's dilemma is one of the puzzles of game theory that motivated the development of theories of team agency. Since any individual player does better by choosing *defect* than by choosing *cooperate*, regardless of what other players do, game theory both predicts and prescribes *defect*. However, a substantial number of people *cooperate* in real life.[9] Further, there is a tension between individual and collective rationality because the players each do better by all choosing *cooperate* than by all choosing *defect.* Whilst any individual player can reason to the conclusion that "The action that gives the best result *for me* is *defect*", it is also true that "The set of actions that gives the best result *for us* is not all *defect".* But reasoning about "our" outcomes has no status in standard game theory.

An analogous point can be made about the intra-personal prisoner's dilemma. In the syntax of the theory of dynamic choice, each transient agent asks separately "What should *I-now* do?" and, in the prisoner's dilemma, the answer is to *defect.* Intuitively, it seems reasonable for the players to ask a different question, "What should *I the person over time* do?', in which case the answer is surely not to *defect* in every time period. Indeed, this latter question seems more than reasonable; if anything, it is the standard model of dynamic choice that seems implausible, with its absence of intentions, plans or any sense of agency that extends over time.

[8] Two "golden rule altruists" playing a prisoner's dilemma, who each give equal weight to self and other's payoffs in their utility functions, may transform the dilemma into a Hi-Lo game, which has two equilibria and, hence, an element of co-ordination (Gold and Sugden, 2007a). Standard game theory cannot prescribe a unique course of play where there is more than one equilibrium, even if one of the two equilibria gives a strictly higher payoff to both players. Hi-Lo is another of the "puzzles" that motivated the development of theories of team agency.

[9] In experiments in which people play the prisoner's dilemma for money, anonymously and without repetition, the proportion of participants choosing *cooperate* is typically between 40 and 50 per cent (Sally, 1995).

Theories of team agency try to reformulate game theory in such a way that “What should *we* do?” is a meaningful question. The basic idea of team reasoning is that, when an individual reasons as a member of a team, she considers which *combination* of actions by members of the team would best promote the team’s objective, and then performs her part of that combination. Although the theory was originally developed with reference to individuals, it could equally be applied the transient agents of dynamic choice theory, with the person being a team of transient agents over time.

We can follow Gold and Sugden (2007a, 2007b) in representing reasoning using schemata of practical reasoning, where agents infer conclusions about what they should do from premises that include propositions about what they are seeking to achieve and about the decision environment (which might respectively be thought of as analogues of Aristotle’s general and particular premises). This is another way of representing the instrumental reasoning of game theory: the standard of success is taken as given and the conclusions tell the agent what s/he should do in order to be as successful as possible according to that standard.

There are four possible outcomes, O_i, corresponding to the four combinations of actions: O_1 from (*cooperate, defect), O_2* from (*cooperate, cooperate*), O_3 from (*defect, defect),* and O_4, from (*defect, cooperate).* An “outcome” includes everything that the players want to achieve.

In the case of a two-period prisoner’s dilemma, the backwards induction reasoning of the second player could have the following form (the propositions above the line are premises, while the proposition below the line is the conclusion):[10]

Schema 1: player 2’s reasoning (individual agency)

(1) I must choose either to *cooperate* or *defect*
(2) If the other player has chosen to *cooperate*, then the outcome will either be O_1 or O_2
(3) If the other player has chosen to *defect*, then the outcome will either be O_3 or O_4,
(4) I prefer O_1 to O_2 and O_3 to O_4, i.e. whatever player 1 has done, I prefer the outcome that results from my playing *defect*

I should choose to *defect*

The schema could be used by an individual playing an asynchronous prisoner’s dilemma or by a transient agent playing a prisoner’s dilemma over time with other transient agents. The reasoning is instrumental practical reasoning, and the “should” is the normativity of instrumental rationality.[11]

[10] I follow philosophical tradition in showing reasoning as the manipulation of propositions, which is also naturally interpreted as conscious manipulation. However, neither of these are necessary for my account. Decision theory is non-committal about the mental processes underlying the choice. In cognitive science a broader definition of reasoning operates, where reasoning can refer to sub-conscious processes and algorithms. The schema presents a “rationalization” of the choice, i.e. it shows how it could be the outcome of a rational process.

[11] One might think that instrumental rationality provides only prima facie reasons and that what an agent all-things-considered ought to do is a question about ethics or morality. In some theories of team agency, team reasoning is a required by morality (see the

In an asynchronous dilemma, the first player can reason by backwards induction, as follows:

Schema 2: player 1's reasoning (individual agency)
(1) I must choose either to *cooperate* or *defect*
(2) If I choose to *cooperate* the outcome will be O_1
(3) If I choose *defect* the outcome will be O_3
(4) I prefer O_3 to O_1

I should choose to *defect*

These schemata show how the two players in an asynchronous prisoner's dilemma can reason that they should defect. (In a synchronic prisoner's dilemma, both players will use a version of schema 1.) They are equivalent to individual rationality in standard game theory or to the reasoning of the transient agents in dynamic choice theory, where *I* may be understood as *I-now*.

If we allow, instead, that the players can ask "What should *we* do?" and consider all possible plans, we get a schema with the following pattern:

Schema 3: collective rationality (team agency)
(1) We must choose one of (*defect, cooperate),* (*defect, defect),* (*cooperate, defect),* (*cooperate, cooperate)*
(2) If we choose (*cooperate, defect)* the outcome will be O_1
(3) If we choose (*cooperate, cooperate)* the outcome will be O_2
(4) If we choose (*defect, defect)* the outcome will be O_3
(5) If we choose (*defect, cooperate)* the outcome will be O_4
(6) We want to achieve O_a more than we want to achieve O_b, O_c, O_d
----
We should choose *(x, y)*

In the inter-personal case the *we* is a team of individuals, in the intra-personal case the *we* is a team of transient agents that make up the agent over time. The actions that the team should take depends on how the team ranks the outcomes, i.e. on the content of premise (6). We might think of O_a as the group goal. The question of how team goals should relate to the rankings of its members is complex (see Gold, 2012), but it seems clear that the team would rank the outcome of (*cooperate, cooperate)* above that of (*defect, defect)* as the former is ranked higher by every player.

In the inter-personal case, it seems reasonable to assume that the players will be treated symmetrically, and Gold and Sugden (2007a) suggest that it is natural to think that the group will rank (*cooperate, cooperate)* above (*defect, cooperate)* and (*cooperate, defect).*

discussion in Gold & Sugden, 2007b). Under that interpretation of team reasoning, my analysis shows how reasoning as a transient agent (or as an individual, in the inter-personal case) can lead to deviations from ethically correct actions. This sort of lack of self-control is what Kennet and Smith (1996) label a failure of *orthonomy*, our capacity to act in accordance with our normative reasons.

In the intra-personal case, symmetry is a less obvious assumption.[12] However, whichever of the three remaining outcomes, (*cooperate, cooperate),* (*defect, cooperate),* and (*cooperate, defect),* the person over time ranks highest, it involves at least one transient agent taking an action that conflicts with her ranking as a person over time. If she follows team reasoning to its end and concludes that she should do her part in the best team plan, then that decision-maker has an Aristotelian problem of self-control, with two conflicting reasoning schemas, depending on whether she reasons as a transient agent or as a team over time.

3. Perception: Framing and self-control

Given that the agent has two conflicting reasoning schemas, what makes her use one over the other? In the Aristotelian account of self-control, the akrates does not use the second syllogism because of a failure of perception. Perception, in the form of "framing", also has an important role in team reasoning. As in the case of reasoning, above, we can apply insights from the inter-personal to the intra-personal case.

A *frame* is the set of concepts a person uses when thinking about the world. Frames are notorious because of Kahneman and Tversky's work on *framing effects*, where changing the description of a choice problem affects the choices that people make (Tversky and Kahneman, 1981). In their classic example, subjects were told that the US was threatened by a deadly disease, which is expected to kill 600 people, and asked to choose between two vaccination programs. Different groups of subjects received different presentations of the decision problem. One group received all the information in terms of how many of the 600 lives would be saved by each program, the other in terms of how many of the 600 would die. The modal choice of program differed between groups; the implication was that the presentation in terms of "saving" and "dying" influenced people's decisions.

Framing starts from the idea, familiar to philosophers, that seeing involves "seeing as". When you see the following marks, O Δ χ, you might see them as a circle, a triangle and a cross. Or, if you know Greek, you might see them as an omicron, a delta and a xi. If you do not know Greek, then the latter option is not a possibility. A larger set of descriptions is available to the linguist. However, the availability of a larger set of descriptions does not imply that they are all used. Someone reading Greek will tend to see the marks as letters, even though they could equally be described as shapes. She frames what she sees as letters, not as shapes. This is like the Aristotelian idea of having knowledge but not using it.

The standard agents of decision-theory use all the knowledge that they have, they always see their world under all of the infinite number descriptions available to them. However, real people are finite, so this is never a possibility for us. We have "bounded cognition".[13]

[12] With more than two players, is it possible to formulate more complicated *production functions*, which specify how combinations of cooperative contributions translate into benefits, where the optimal outcome for the team has some, but not all, team members cooperating. That might be a better model for examples like healthy eating and studying. In that case, it is the transient agents who are assigned to make a cooperative contribution that face a problem of self-control.

[13] The allusion to Herbert Simon's "bounded rationality" is intentional (Simon, 1955). *Bounded rationality* is the idea that, unlike the ideal agents of decision theory, real agents

At any time, we will be using only a small subset of the concepts that could describe our situation.

Framing is an important pre-cursor to decision-making. In order for something to feature in an agent's reasoning, she must have concepts related to it in her frame. Hence, in order to team reason, a player must have the concept "we" in her frame.

Many accounts of team agency emphasize the role of commitment in group identification, be it rational commitment, moral commitment or simply the endorsement of a particular mode of reasoning by the agent (Gold and Sugden, 2007a, 2007b). But even accounts of group agency that do not give perception a prominent role have an implicit framing step, as seeing that a decision can be described as a problem for "us" is a necessary pre-condition for team reasoning.[14]

In contrast, Bacharach (2006) gives an account where team reasoning is purely the result of framing. For Bacharach, certain features of choice-problems may, when salient, promote group identity which, in turn, primes team reasoning. In his model, these transitions are all the results of psychological processes, not of decisions. However, there is an implicit commitment to or pre-eminence of the *we*-frame. Team reasoners must have the *I*-concept in their frame, as they reason to conclusions about what individual actions they should take, and because team reasoning can be *circumspect*, taking into account the possibility that others do not group identify but act on individual reasoning instead. Nevertheless, there is an assumption that priming *we*-concepts tends to promote team reasoning.[15]

Bacharach elides the distinction between the noticing of *we*-concepts and their having what we might call "motivational grip" (Gold, 2012). Motivational grip has two components above and beyond merely noticing a feature or concept: noticing that it is choice-relevant and, given that it has been noticed as choice-relevant, deciding to act on the reason it provides. Hence Bacharach's team reasoners must either not find *I*-reasons to be choice relevant or else they have decided not to act on them. In order to stay within a purely cognitive framework, Bacharach might say that the salience of the *we*-concepts outweighs that of the *I*- concepts. But, in order for framing the decision as a problem for "us" to affect behavior, *we*-reasons have motivating power, which seems to presume an implicit commitment to the team agent.

are subject to cognitive limitations. Simon's research programme emphasized limitations on information processing. Bounded cognition is (at least partly) a cognitive limitation. So, strictly speaking, bounded cognition is a species of bounded rationality, albeit not one that has come under much scrutiny. One advantage of the moniker "bounded cognition" is that it does not mention rationality, hopefully avoiding the increasing tendency to confuse bounded rationality with irrationality (Gigerenzer, 1997), when questions of rationality are really still up for debate.

[14] See Gold (2012) for a more detailed examination of this point with respect to Sugden's account.

[15] In an earlier presentation of his theory, Bacharach (1999) allows that there is an *I*-frame, a *we*-frame, and a *superordinate* frame, oscillating between the *I* and *we*-perspectives. Smerilli (2012) explores this further, providing a simple model of vacillation between *I* and *we*-modes of reasoning.

We can draw parallels with team reasoning at the intra-personal level. In the same way that inter-personal team reasoning requires the decision-maker to frame the decision as a problem for us, intra-personal team reasoning requires the decision-maker to frame it as a problem for herself over time. Of course, there is a sense in which everyone knows they are an extended self over time but, as we saw above, it is possible to have that knowledge without using it to frame a decision problem. Arguably, it is more natural for people to think of themselves as selves over time than as transient agents, so the team frame might be the default in the intra-personal case. It certainly seems natural to think that people have an implicit commitment to their extended self over time. The lack of such a commitment is one reason why the pure transient agent model is impoverished.

When there is temptation, the divergence of interests between the transient agent and the person over time is salient. At the inter-personal level, social identity theorists say that awareness of common interest and a common fate promotes group identification, by raising awareness of a relevant basis for categorization into groups (Brewer, 1979). Conversely, awareness of divergent individual interests may inhibit group identification, and obscure awareness of any basis for group categorization, and this may be true in both the inter-personal and the intra-personal cases. Hence akrasia may be associated with a lack of identification with the self over time.

The idea that akrasia involves a failure of perception is central to the Aristotelian account. For Aristotle, the akrates is unaware of a particular premise. However, in the team reasoning account, the concept of “we” is in every premise. Even the group goal is “our” ordering, from the point of view of the team, so perceiving it involves having the concept of the team in one’s frame. Indeed, seeing the group goal may be more important in triggering team reasoning than any of the other premises if, as social identity theorists claim, recognizing a common interest can trigger group identity. Hence it must be that the whole team reasoning schema is not used by the akrates.

4. Psychological Evidence and Discrimination Between Theories

The intra-personal team reasoning account implies that we can improve self-control by increasing the salience of the self over time, and by increasing the salience of long-term goals relative to transient ones. The relative salience of long-term goals can be increased either by increasing the salience of the long-term or by decreasing the salience of the short-term, which can be done by focusing attention on the long-term goal or distracting oneself from the immediate temptation.

Here I present evidence that is compatible with the intra-personal team reasoning account of self-control, explore why current evidence does not discriminate between the account and other explanations of self-control, and suggest a way in which the account might have increased explanatory power.

There is evidence that distraction from temptation, by engaging in other activities and even just by thinking about other activities, can increase self-control (Mischel & Baker, 1975; Mischel, Ebbesen & Zeiss, 1972).[18] And, if directing attention is an effortful activity, then

[18] Mischel’s experimental paradigm involves a child who can obtain a less preferred food reward immediately or wait for a more preferred food reward. Since the temptation and the

an account where self-control involves focusing attention is consistent with some of the psychological evidence on *ego depletion*, the idea that willpower is a limited resource. Being asked to exert self-control in a first task adversely affects subjects' performance in a second task that also requires self-control (Baumeister *et al* 1998; Muraven, Tice Baumeister, 1998). Conversely, high glucose levels and consumption of calories can increase self-control (Gailliot *et al* 2007). Muraven and Baumeister (2000) make an analogy between willpower and a muscle, whose strength gets depleted as you use it, but whose strength can be built up with exercise. If the analogy is correct, it implies that there is some unspecified effortful mechanism underlying self-control. Focusing attention plausibly requires effort, especially when one is trying to focus attention away from a salient stimulus. If we think that controlling attention can be learned, then the account can also explain why people can improve their self-control with practice (Muraven, Baumeister & Tice, 1999).

However, whilst the efficacy of directing attention between outcomes is consistent with the intra-personal team reasoning account, we can talk about framing the options without introducing levels of agency. The idea that we can solve problems of self-control by re-describing the options is already present in the philosophical literature (Kennette & Smith, 1996; Mele, 2012), and there are models that explain akrasia as a conflict of reasons (Kavka, 1991; Pettit, 2003) and as conflicting perceptions of reasons (Gold, 2005; Schick, 1991).

The model of intra-personal team reasoning suggests that, in addition to re-framing the options, we should be able to improve self-control by re-framing the agent. This hypothesis needs further research. Here are two reasons for thinking that it might be fruitful.

First, there is some evidence on the self over time and delayed gratification, from the work of Dan Bartels and colleagues (Bartels & Rips, 2010; Bartels & Urminsky, 2011). This starts from Derek Parfit's (1984) argument that personal identity depends on *psychological connectedness,* having psychological connections with our future selves such as sharing memories, intentions, beliefs, and desires. Bartels wanted to see if there is a correlation between connectedness and the "discount rate", which leads us to choose sooner smaller rewards over larger later ones. In fact, he tested the relation between *perceived* connectedness and the discount rate, finding that subjects who rated themselves as more connected to later selves were more patient (Bartels & Rips, 2010) and that connectedness can be manipulated, resulting in increased patience (Bartels & Urminsky, 2011).

Bartels' work involves the perception of psychological connectedness. This is not the same as the perception of shared agency, but it is plausible that they are related. At the inter-personal level, identification with other members of a group enhances the accessibility of shared characteristics (Smith & Henry, 1996). Perceived similarities may also increase the likelihood of group identification. It is likely that the causality runs in both directions. At the intra-personal level, we might hope that future research will attempt to discriminate between the effects of psychological connectedness and those of team agency or, alternatively (should discrimination not be possible), to explore the connections between the two.

reward are in the same currency, distractions are distractions both from the temptation and the more preferred reward, which is the long term goal.

A further confound is that the perception of psychological connectedness might increase intra-personal altruism as well as or instead of increasing the perception of the self-as an agent over time. Again, intra-personal altruism and sense of agency may be related: framing a decision as a problem for us may also encourage a concern for the welfare of the other transient agents that belong to the self over time. The same issue occurs at the inter-personal level, where the question becomes how to discriminate between interventions that increase inter-personal altruism and those that make salient team agency. Bacharach (2006) claims that a test that discriminates team reasoning can be constructed by using the fact that team reasoners take actions that lead to the group utility maximizing outcome, whereas individual reasoners will sometimes end up at an outcome with lower payoffs, if it is salient and solves a co-ordination problem. More work remains to be done, at both the inter-personal and intra-personal level.

A second reason favouring the intra-personal team reasoning account is that it has increased explanatory power. There is a datum that cannot be explained by a simple framing of the object account: the relation between Borderline Personality Disorder (BPD) and self-control. A personality disorder occurs when the way that a person is inclined to think, feel, and act causes that person severe psychological distress, impairs them in important contexts and does them harm (Pickhard, 2011). BDP is defined by instability of interpersonal relationships, self-image, and affects, and a marked impulsivity. A person is diagnosed with BDP when she displays at least five of a list of diagnostic traits, which include "identity disturbance: markedly and persistently unstable self-image or sense of self", and "impulsivity in at least two areas that are potentially self-damaging" (American Psychiatric Association [DSM-IV] 2000, p.706).

BPD patients have a fractured sense of self. They do not identify with their later selves and they do not think through the consequences of their actions on either themselves or others. The amount of impulsivity displayed by BPD patients can be extreme and their inability to carry through plans causes severe detriment to their lives. They cannot hold down jobs and have impoverished relationships. However, although impulsivity is a diagnostic criterion for BDP, there is currently no theoretical explanation of the co-occurrence of impulsivity and identity disturbance.

BDP is also characterized by "affective instability", with intense emotions and mood swings (American Psychiatric Association [DSM-IV] 2000). Current treatments for BDP, such as therapies involving "mindfulness" (Breslin, Zack & McMain, 2002) and "mentalizing" (Fonagy & Bateman 2007) focus on patients' emotional shifts, helping them to take a more detached perspective on intense emotions, especially negative ones, and teaching them how to focus their attention. There is evidence that BDP patients have difficulty in controlling their attention (Hoermann *et al*, 2005; Lenzenweger *et al*, 2004; Posner et al., 2002) and that negative affective cues activate alcohol cognitions in problem drinkers with psychiatric disorders (Zack, Toneatto, & MacLeod, 1999; Zack, Toneatto, & Colin M. MacLeod, 2002). As with other evidence cited above, this could be explained by an account of self-control as involving re-framing options. However, in this case, that account leaves a feature of the condition unexplained, namely the connection between unstable self identity and impulsivity - exactly the extra component that the intra-personal team reasoning theory can provide.

The theory of intra-personal team reasoning is not the only account of self-control that invokes a division of agency and a diachronic perspective. There are models involving a long and a short-sighted self (Schelling, 1984), a planner and a doer (Thaler & Shefrin, 1981), global and local choice (Heyman, 2009), and short-range and long-range interests (Ainslie, 1992). Intra-personal team reasoning is compatible with some of these accounts, providing more detail about what the long-sighted or global perspective involves, and explicit modeling of the interaction of successive transient agents. However, it is not obviously compatible with accounts where the long-range interest is constrained to always take the same action in every period (e.g. Ainslie, 1992). In the team reasoning account, it is possible that the optimal team plan involves occasional lapses, if that produces the best outcome for the self over time.

Conclusion

Aristotle examined the reasoning of the akratic agent but did not specify the reasoning of the agent who exhibits self-control. Similarly, the framework of decision theory explains why people would give in to temptations but not how people can use willpower to exert self-control. I introduced the idea of levels of agency, with the self as a team over time that makes and follows plans, and showed how intra-personal team reasoning can lead to self-control. The account is Aristotelian in that it involves reasoning schema and a lack of perception on the part of the akratic agent, who does not see her decision-problem as a problem for her self over time. It suggests a role not just for for framing of the options, but also for the framing of the agent. This would merit further investigation in future research.

References

Ainslie, G. (1991). Derivation of "rational" economic behavior from hyperbolic discount curves. *American Economic Review* 81: 134–140.

Ainslie, G. (1992). *Picoeconomics.* Cambridge University Press, Cambridge.

American Psychiatric Association. (2000). *Diagnostic and Statistical Manual of Mental Disorders Revised 4th ed*. Washington, DC: Author.

Austin, J. L., (1956/7). A Plea for Excuses. In J. L. Austin 1979, *Philosophical Papers*, 3rd ed., J. O. Urmson and G. J. Warnock (eds.), Oxford: Oxford University Press.

Aristotle (1984). *Nicomachean Ethics*. In J. Barnes ed. *The Complete Works of* Aristotle, Volume II, Princeton: Princeton University Press.

Bacharach, M. (2006). *Beyond Individual Choice.* N. Gold and R. Sugden (eds.) Princeton: Princeton University Press.

Bacharach, M. (1999). Interactive team reasoning: A contribution to the theory of cooperation. *Research in Economics* 53, 117-47.

Bacharach, M. (1997). 'We' equilibria: a variable frame theory of cooperation. Working paper, Institute of Economics and Statistics, University of Oxford.

Barnes, J. (1984). *The Complete Works of Aristotle Volume II*. Princeton: Princeton University Press.

Baumeister, R. F., Bratslavsky, E., Muraven, M. & Tice, D. M. (1998). Ego Depletion: Is the Active Self a Limited Resource?. *Journal of Personality and Social Psychology* 74, 1252–1265.

Bartels, D. & Rips, L. (2010). Psychological connectedness and intertemporal choice. *Journal of Experimental Psychology: General* 139: 49-69.

Bartels, D. & Urminsky, O. (2011). On intertemporal selfishness: How the perceived instability of identity underlies impatient consumption. *Journal of Consumer Research* 38: 182-198.

Bostock, D. (2000). *Aristotle's Ethics.* Oxford: Oxford University Press.

Bratman, Michael (1987). *Intention, Plans and Practical Reason.* Cambridge Ma.: Harvard University Press.

Breslin, F. C., Zack, M, & McMain, S. (2002).An Information-Processing Analysis of Mindfulness: Implications for Relapse Prevention in the Treatment of Substance Abuse. *Clinical Psychology: Science and Practice* 9(3), 275-99.

Brewer, M. B. (1979). In-Group Bias in the Minimal Intergroup Situation: A Cognitive-Motivational Analysis. *Psychological Bulletin*, 86, 307-324.

Charles, D. (1984). *Aristotle's Philosophy of Action* London: Duckworth.

Collard, D. (1978). *Altruism and Economy: a Study in non-Selfish Economics* Oxford: Martin Robertson.

Fonagy, P., & Bateman, A., W. (2007). Mentalizing and borderline personality disorder. *Journal of Mental Health*, 16(1), 83 – 101.

Frederick, S., Loewenstein, G. & O'Donoghue, T. (2002) "Time discounting and time preference: A critical review" *Journal of Economic Literature* 40(2), 351-401.

Gailliot, M.T., Baumeister, R.F., DeWall, C.N., Maner, J.K., Plant, E.A., Tice, D.M., Brewer, L.E., & Schmeichel, B.J. (2007). Self-Control relies on glucose as a limited energy source: Willpower is more than a metaphor. Journal of Personality and Social Psychology, 92, 325-336.

Gigerenzer, G. (1997). Bounded Rationality: Models of Fast and Frugal Inference. *Swiss Journal of Economics and Statistics, 133* (2/2), 201–218.

Gold, N. (2012). Team Reasoning and Cooperation. In S. Okasha and K. Binmore (eds) *Evolution and Rationality: Decisions, Cooperation and Strategic Behaviour* Cambridge: Cambridge University Press.

Gold, N. (2005) Framing and Decision Making: A Reason- Based Approach. Unpublished D.Phil thesis, University of Oxford.

Gold, N., & Sugden, R. (2007a). Theories of Team Agency. In F. Peter & S. Schmidt (Eds.), *Rationality and Commitment* Oxford Oxford University Press.

Gold, N., & Sugden, R. (2007b). Collective Intentions and Team Agency. *Journal of Philosophy 104* (3), 109-137.

Heyman, G. M. (2009). *Addiction: A Disorder of Choice*. Cambridge, Mass.: Harvard University Press.

Hoermann, S., Clarkin, J. F., Hull, J. W., & Levy, K. N. (2005). The construct of effortful control: An approach to borderline personality disorder heterogeneity. *Psychopathology*, 38, 82 – 86.

Hume, D. (1739-40/1978). *A Treatise of Human Nature.* Oxford: Clarendon Press.

Kahneman, D.& Tversky, A. (1984). Choices, values and frames. *American Psychologist* **39** (4): 341–350.

Kavka, G. (1991). Is Individual Choice Less Problematic than Collective Choice?. *Economics and Philosophy*, 7 : 291-310.

Kennett, J. & Smith. M. (1996). Frog and toad lose control. *Analysis* 56: 63-72.

Laibson, David (1997). Golden Eggs and Hyperbolic Discounting. *Quarterly Journal of* Economics, 112:443-77.

Ledyard, J. O. (1995). Public goods: A survey of experimental research. In J. H. Kagel & A. E. Roth (Eds.), *Handbook of experimental economics* (pp. 111–194). Princeton, NJ: Princeton University Press.

Lenzenweger, M. F., Clarkin, J. F., Fertuck, E. A., & Kernberg, O. F. (2004). Executive neurocognitive functioning and neurobehavioral systems indicators in borderline personality disorder: A preliminary study. J*ournal of Personality Disorders*, 18, 421 – 438.

McCarthy, D. & Arntzenius, F. (1997). Self Torture and Group Beneficence. *Erkenntnis* 47: 129–144.

Mele, A., R. (2012). *Backsliding: Understanding Weakness of Will.* Oxford University Press

Mischel, W., & Baker, N. (1975). Cognitive Appraisals and Transformations in Delay Behavior. *Journal of Personality and Social Psychology* 31(2), 254-61.

Mischel, W., Ebbesen, E. B. & Zeiss, A. R. (1972). Cognitive and Attentional Mechanisms in Delay of Gratification. *Journal of Personality and Social Psychology* 21(2), 204-18.

Moss, J. (2009). *Akrasia* and Perceptual Illusion. *Archiv fur Geschichte der Philosophie* 91, 119–156.

Muraven, M., & Baumeister, R. F. (2000). Self-regulation and depletion of limited resources: Does self-control resemble a muscle? *Psychological Bulletin, 126*, 247-259.

Muraven M., Baumeister, R. F. & Tice, D. M. (1999). Longitudinal Improvement of Self-Regulation Through Practice: Building Self-Control Strength Through Repeated Exercise. *Journal of Social Psychology*, Vol. 139, 446-57

Muraven, M., Tice, D. M., & Baumeister, R. F. (1998). Self-control as a limited resource: Regulatory depletion patterns. *Journal of Personality and Social Psychology, 74*, 774-789.

Parfit, D. (1984). *Reasons and persons.* Oxford: Clarendon.

Pettit, P. (2003). Akrasia, Collective and Individual. In Sarah Stroud and Christine Tappolet, eds., *Weakness of Will and Practical Irrationality*, Oxford, Oxford University Press, pp. 68-96.

Pickard, H. (2011). What is Personality Disorder? *Philosophy, Psychiatry, and Psychology* 18(3), 181-4.

Posner, M. I., Rothbart, M. K., Vizueta, N., Levy, K. N., Evans, D. E., Thomas, K. M., & Clarkin, J. (2002). Attentional mechanisms of borderline personality disorder. *Proceedings of the National Academy of Sciences of the USA*, 99,16366 – 16370.

Sally, D. (1995). Conversation and cooperation in social dilemmas: a meta-analysis of experiments from 1958 to 1992. *Rationality and Society* 7, 58-92.

Schelling, T. (1984) Self-command in practice, in policy, and in a theory of rational choice. *American Economic Review* 74, 1-11.

Schick, F. (1991). *Understanding Action* Cambridge: Cambridge University Press.

Simon, H. A. (1955). A behavioral model of rational choice. *Quarterly Journal of Economics, 69,* 99–118.

Skyrms, B. (1998). The Shadow of the Future. In Coleman and Morris (eds.) Rational Commitment and Social Justice: Essays for Gregory Kavka, New York, Cambridge University Press.

Smerilli, A. (2012). We-thinking and vacillation between frames: filling a gap in Bacharach's theory. Theory and Decision. On-line first publication.

Smith, E. R. & Henry, S. (1996). An In-group Becomes Part of the Self: Response Time Evidence. *Personality and Social Psychology Bulletin* 22, 635-42.

Strotz, R. H. (1955-6). Myopia and Inconsistency in Dynamic Utility Maximization. Review of Economic Studies 23, 165-80.

Sugden, R. (2000). Team preferences. *Economics and Philosophy* 16, 175–204.

Sugden, R. (1993). Thinking as a team: toward an explanation of nonselfish behavior. *Social Philosophy and Policy,* 10, 69-89.

Thaler, R. (1981). Some Empirical Evidence on Dynamic Inconsistency. *Economics Letters* 8(3), 201–07.

Thaler, R. H. & Shefrin, H. M. (1981) Temporal construal. *Psychological Review* 110, 403-21.

Tversky, A, & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science* 211, 453–458.

Tversky, A. & Kahneman, D. (1986). Rational Choice and the Framing of Decisions. *The Journal of Business*, 59(4), pages S251-78,

Urmson, J. O. (1988). *Aristotle's Ethics.* Oxford: Basil Blackwell.

Vanderschraaf, P. (1998). The Informal Game Theory in Hume's Account of Convention. *Economics and Philosophy, 14*, 215-247.

Wiggins, D. (1980). Weakness of the Will, Commensurability, and the Objects of Deliberation and Desire. In A. Oksenberg Rorty (ed.) *Essays on Aristotle's Ethics* Berkeley: University of California Press.

Zack, M. Toneatto, T., & MacLeod, C. M. (2002). Anxiety and explicit alcohol-related memory in problem drinkers. *Addictive Behaviors* 27, 331–343.

Zack, M., Toneatto, T., & MacLeod, C. M. (1999). Implicit activation of alcohol concepts by negative affective cues distinguishes between problem drinkers with high and low psychiatric distress. *Journal of Abnormal Psychology*, 108, 518–531.