



MASTERARBEIT | MASTER'S THESIS

Titel | Title

The Rationality of Team Reasoning

verfasst von | submitted by
Rafaela Schinner B.A.

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of
Master of Arts (MA)

Wien | Vienna, 2024

Studienkennzahl lt. Studienblatt | Degree
programme code as it appears on the
student record sheet:

UA 066 642

Studienrichtung lt. Studienblatt | Degree
programme as it appears on the student
record sheet:

Masterstudium Philosophy and Economics

Betreut von | Supervisor:

Univ.-Prof. Dr. Hans Bernhard Schmid

Table of Contents

1. INTRODUCTION	2
2. THE PROBLEM OF COOPERATION IN PRISONER'S DILEMMAS	4
2.1. THE STANDARD PRISONER'S DILEMMA MODEL.....	4
2.2. REAL-WORLD APPLICATIONS	4
2.3. THE NORMATIVE PROBLEM	5
2.3.1. <i>Argument from the possibility of committed action</i>	6
2.3.2. <i>Argument from inefficiency</i>	6
2.3.3. <i>Conclusion</i>	7
3. TEAM REASONING AND THE JUSTIFICATION PROBLEM	8
3.1. TEAM REASONING AND THE PRISONER'S DILEMMA.....	8
3.2. DEMARCATION FROM SOCIAL PREFERENCE THEORIES.....	8
3.3. THE JUSTIFICATION PROBLEM	9
4. STATE OF THE ART AND RESEARCH QUESTION	10
4.1. STATE OF THE ART	10
4.2. RESEARCH QUESTION	12
4.3. ANALYSIS OF THE JUSTIFICATION PROBLEM	12
5. ELIZABETH ANDERSON'S PROPOSAL.....	15
5.1. A ROUGH RECONSTRUCTION OF THE KORSGAARDIAN CONCEPTION OF A REASON	15
5.2. SOCIAL IDENTITY AND THE UNIVERSALISABILITY TEST	15
5.3. SOCIAL IDENTITY AND PRISONER'S DILEMMAS	18
5.4. BACK TO THE RATIONALITY OF TEAM IDENTIFICATION.....	19
5.4.1. <i>The priority principle</i>	19
5.4.2. <i>The standpoint of the individual self</i>	21
5.4.3. <i>The choice principle</i>	22
6. CHRISTINE KORSGAARD'S PRACTICAL IDENTITY.....	26
6.1. WHY THE IDENTITY AS A HUMAN BEING IS NECESSARY	26
6.2. KORSGAARD AND TEAM REASONING	27
7. SUMMARY AND FURTHER THINKING.....	29
8. LIMITATIONS.....	31
9. CONCLUSION.....	33
BIBLIOGRAPHY	34
LIST OF FIGURES.....	36
LIST OF TABLES.....	36
APPENDIX: ABSTRACTS.....	37
ABSTRACT (ENGLISH).....	37
ABSTRACT (DEUTSCH).....	38

1. Introduction

Collective action problems are often characterised as n-person Prisoner's Dilemma situations. Prisoner's Dilemma games are characterised by the fact that it is rational for all agents to defect, but if all defect, we end up with a non-Pareto efficient outcome. Many proponents and critics of rational choice theory¹ have expressed dissatisfaction with this implication of rational action.

Team reasoning constitutes a prominent solution to explain cooperation in Prisoner's Dilemma situations: agents cooperate insofar as they reason as members of a team. From a normative perspective, this raises the question of why the agent should engage in team reasoning rather than individual utility maximisation (= justification problem). The most prominent team reasoning accounts have accounted for the justification problem in psychological, descriptive terms only (Bacharach 2006) or offered a normatively inadequate solution (Sugden 2015). Therefore, the justification problem is still unsolved (Gold and Colman 2020).

Elizabeth Anderson (2001) proposed an interpretation of team reasoning where what reasons we have depends on our social identities: insofar as I identify as an independent agent, I have reason to engage in utility maximisation. Insofar as I identify as a member of some social group, I have reason to engage in team reasoning. The concept of social identity is based on Korsgaard's (1996) concept of practical identity. Anderson's account has not received in-depth attention in the team reasoning literature. In this thesis, I examine whether Anderson manages to solve the justification problem.

I will argue that Anderson's proposal does not provide an adequate solution to the justification problem. I identify two main issues: first, she cannot motivate why we should identify with humanity and why this identity should be prioritised over other social practical identities. Second, she lacks an adequate account of the nature of the independent agent and its relation to social identities and the individual self. Further, I show that Korsgaard's account of practical identity can solve the first issue, but only at the cost of sacrificing team reasoning.

I start by introducing the Prisoner's Dilemma and present two arguments for why it is unsatisfactory from a normative perspective ([section 2](#)). Further, I introduce the concept of team reasoning and explain the justification problem ([section 3](#)). [Section 4](#) will be dedicated to an examination of the existing literature as well as the presentation of the research question. [Section 5](#) constitutes the core of the thesis, analysing Anderson's account and evaluating her solution to the justification problem. Subsequently, I am concerned with Korsgaard's account of practical identity ([section 6](#)), before offering a synthesis of the

¹ Henceforth, I will use 'RCT' for 'rational choice theory'.

findings (section 7). Finally, I will address some limitations (section 8) and conclude (section 9).

2. The Problem of Cooperation in Prisoner's Dilemmas

2.1. The standard Prisoner's Dilemma model

The standard form of the Prisoner's Dilemma (table 1) describes a strategic non-cooperative game between two players with symmetric payoffs, whereby $A > B > C > D$ (see Peterson 2015).

		player 2	
		cooperate	defect
player 1	cooperate	B / B	A / A
	defect	A / D	C / C

table 1: Prisoner's Dilemma, standard form

According to standard game theory, which is based on rational choice theory, rational players maximise their expected utility². Accordingly, 'defect' is the dominant strategy for both players: regardless of what the other player does, each player should defect, as $A > B$ and $C > D$. Therefore ('defect', 'defect') is the unique Nash Equilibrium and both players will end up with payoff C. However, this result is not Pareto-efficient: the outcome of ('cooperate', 'cooperate') would have been better for *both* players, as $B > C$.

2.2. Real-world applications

The structure of many collective endeavours is captured by the (n-person³) Prisoner's Dilemma (Hardin and Garrett 2020). Tragedies of the Commons are commonly understood in terms of a n-person Prisoner's Dilemma (Bovens 2015). The same applies to public good provision situations, where the impact of the individual's contribution on the socially desirable outcome is negligible.

Voting is a paradigmatic example for the latter (Anderson 2001, 26): consider the individual decision of a partisan of whether to vote or not. The election of the desired candidate is seen as a public good by this partisan and his fellow partisans. For the partisan, voting is costly: she has to sacrifice time and possibly money to go to the polling station. At the same time, the marginal impact of her vote is close to zero⁴. Now, this is not simply an individual

² I am not going to discuss the differences between rational choice theory and expected utility theory and treat the two as synonyms. In the philosophical literature, they are often used interchangeably.

³ I will not go into the details regarding the differences between n-person PD's (where n is very high so that an individual's impact is negligible) and 2-person PD's. Importantly, they both have a similar incentive structure: 'defect' is the dominant strategy, but if all (or most) players defect, we end up in a non-Pareto efficient outcome. For our purposes, we can therefore treat them analogously. For a discussion about how the size of n matters for in what sense 2-person PD's and n-person PD's are comparable, see Brennan and Brooks (2015).

⁴ It is of course true that there is a certain, albeit tiny, probability that my vote will make the difference. If this happens, the utility of my act of voting is considerable. One could argue that the expected utility of my act of voting is therefore positive and I should vote. Similar examples could be made with regard to pollution and

decision problem, but a strategic one: the outcome (which relates to the individual partisan's utility) depends on whether (and what) the fellow partisans vote. According to expected utility theory, the partisan should not vote: independently of how many of her fellow partisans vote and what they vote, her utility is higher if she does not vote, i.e. not voting is the dominant strategy. However, if all individuals act according to expected utility theory, no one will vote and we will end up in a non-Pareto-efficient situation⁵.

2.3. The normative problem

From a philosophical perspective, rational choice theory can be understood as a theory of practical rationality. Roughly speaking, this means that an agent faced with some choice situation should a) have preferences that fulfil the axioms of expected utility theory; and b) choose the action that maximises her expected utility⁶. Understanding expected utility theory as a theory of practical rationality has two distinct dimensions⁷: From the third-person perspective, it is about evaluating an agent's choices as rational or irrational. From the first-person perspective, the question is: how should I decide? In that sense, rational choice theory is action-guiding.

Concerning the Prisoner's Dilemma, rational choice theory says that there is a rational obligation⁸ to defect, or in other words that cooperating is irrational. One possible normative stance is to embrace this and hold on to the claim that agents in these situations should always defect. However, there is a long tradition of scholars who are dissatisfied with the fact that RCT considers cooperative behaviour in these situations as irrational. This was most famously expressed by Amartya Sen, who labelled the consistent follower of RCT a "rational fool" (1977, 336).

climate action, for example. Shelly Kagan presents such an argument for what she calls "triggering cases" (Kagan 2011, 120). I will not examine her proposal in detail in the context of this thesis. Prima facie, however, it does not convince me. First, because the probability in these cases is very low and it is hard to imagine that the possible benefit is high enough to rationalise the action. Second, it does not seem to do justice to the first-person perspective: it seems for instance implausible that the small probability that their vote might make the difference is the consideration in light of which people decide to go voting.

⁵ Importantly, these cases do not present a conflict between altruistic and egoistic behaviour. As Anderson shows, as long as the individual actions have a marginal impact of zero on the socially desirable outcome, the problem persists even if we assume altruistic preferences (2001, 25-6).

⁶ What 'utility' exactly means has been debated in the literature. For a differentiation between readings of 'utility', see Thoma (2019). For our purposes, I consider both a modern realist or a constructivist reading possible, provided that preferences are not just understood as choices. The scope of this thesis does not allow for a discussion about what notion of preferences fits best the normative use of RCT.

⁷ For a more detailed explanation of these two dimensions of practical rationality with specific regard to expected utility theory, see Thoma (2019).

⁸ I understand the rational obligation to do ϕ to be equivalent to having decisive reason to do ϕ and the rational permission to do ϕ to be equivalent to having sufficient reason to do ϕ . I will present the account of reasons resp. rationality Anderson employs in [section 5.1](#).

Within the debate, we can find two different – although closely related – lines of thought spelling out this dissatisfaction. The first refers to the possibility of action based on commitment. The second refers to our expectations of rationality as leading us to achieve our goals. In the following sections, I will elucidate each one of them.

2.3.1. Argument from the possibility of committed action

One of the most prominent critiques of rational choice theory has been put forth by Amartya Sen. In his paper “Rational Fools: A Critique of the behavioural foundations of economic theory” (1977), he introduced the concept of “action based on commitment” (Sen 1977, 326). The term refers to choice based on considerations that are *independent* of personal welfare and can therefore even be counter-preferential. Importantly, this kind of counter-preferential choice is supposed to be perfectly rational. The plausibility of the possibility of counter-preferential but rational choice depends on how we conceive of personal welfare⁹. Sen contrasts committed action with action based on sympathy. If I help someone because it makes me feel better, I act out of sympathy, which is plausibly part of personal welfare. If I do so while accepting that it will make me worse off because I think it is the right thing to do, I act out of commitment. This is plausibly not included in personal welfare.

One example of action based on commitment can be cooperation in Prisoner’s Dilemmas. The appeal to our intuition is the following: do we want to deem people who vote or refrain from flying to be irrational? While it might be too strong to claim that they are rationally obliged to vote or not to take the aeroplane, it seems at least reasonable to do so in many cases. If we accept the possibility of committed action, we can understand the cooperator’s choices as based on commitment and therefore rational. The difficult question is *what makes it rational* to do so¹⁰.

2.3.2. Argument from inefficiency

The obligation to defect in Prisoner’s Dilemma situations is based on the principle of strategy dominance: for every possible choice of the other agent, defecting is the best response with respect to my payoff. Interestingly, in this specific kind of situations, strategic rationality systematically leads us into non-Pareto-efficient outcomes: if all agents had acted irrationally, they would all be better off than if all agents had acted rationally. As Hans

⁹ It is an open question how the concepts of preferences and utility are understood. One point of contention relates to how preferences and personal welfare are connected and how narrowly or broadly ‘personal welfare’ is conceptualised. On a narrow reading (only including my own consumption, not including sympathy), one might argue that Prisoner’s Dilemmas arise only because some relevant motivations for action are not included. On a wide reading, where everything possibly influencing choice is included, Sen’s distinction between commitment and sympathy will lose its sense. Sen’s argument presupposes a middle-ground reading of ‘welfare’. A detailed discussion of this matter is beyond the scope of this thesis.

¹⁰ Sen’s explanations remain rather unspecific with regard to the term ‘commitment’ or the understanding of rationality associated with it, which has led to an extensive debate and countless interpretations of what committed action refers to. One of these interpretations comes from Elizabeth Anderson (2001) and forms the central text of this master’s thesis.

Bernhard Schmid points out: "It seems counterintuitive to say that it is rational to be inefficient" (Schmid 2021, 592). Efficiency is connected to another important aspect of rationality: it should somehow be conducive to our good, to what we want to achieve. In Prisoner's Dilemma situations, however, rationality seems to be impeding our effectiveness in the world. The payoff structure of the Prisoner's Dilemma drives a wedge between efficiency and strategy dominance (Schmid 2021).

2.3.3. Conclusion

Both the argument from the possibility of committed action and the argument from efficiency speak in favour of finding a way to render cooperation in Prisoner's Dilemma situation rational – at least in some circumstances. Any account doing so has to meet the challenge to show why it can be rational for me to choose a behaviour that is dominated by another strategy in every situation, i.e. show why it can be rational for me to act against my preferences.

One promising family of accounts that attempt to rationalise cooperation in Prisoner's Dilemmas is team reasoning. In the following chapter, I will present the main concept of team reasoning, demarcate it from social preference theories and show why for now, it only leads to a problem shift.

3. Team reasoning and the justification problem

3.1. Team Reasoning and the Prisoner's Dilemma

Team reasoning accounts aim to keep the standard rationality assumptions of rational choice theory while trying to solve the problem of cooperation in Prisoner's Dilemmas by introducing a new form of reasoning: team reasoning.

This involves two kinds of transformations (see Gold and Colman 2020): First, a payoff transformation: individual preferences are somehow aggregated to form team preferences. Second, an agency transformation: the players are not asking 'What should I do?' but rather 'What should we do?'. This means that the unit of agency is the group: players act *by virtue of being members of the team*, engaging in a joint action.

Consider the following example from Sugden and Gold (2007, 292) displaying the payoff transformation, in this case by averaging the individual utilities of the team members.

		player 2	
		cooperate	defect
player 1	cooperate	8 / 8	0 / 10
	defect	10 / 0	6 / 6

		player 2	
		cooperate	defect
player 1	cooperate	8 / 8	5 / 5
	defect	5 / 5	6 / 6

table 2: Prisoner's Dilemma, payoff transformation and agency transformation

We observe that, under the condition that the agent reasons as a member of the team, cooperating becomes the rational individual action from the point of view of each agent. Importantly, without the agency transformation in addition to the payoff transformation, we end up in a Hi-Lo game¹¹.

Team reasoning can thus render cooperative behaviour in Prisoner's Dilemmas rational, without having to extend or adjust our conception of rationality: it is in accordance with expected utility theory, except that now the teams utility is maximised. Further, the proponents of team reasoning do not have to sustain two distinct forms of reasoning: individual reasoning is simply team reasoning where the team consists of one member only (see Gold and Colman 2020).

3.2. Demarcation from social preference theories

Team reasoning approaches are to be distinguished from social preference theories. Team reasoning gives up the assumption that agents are only *individually* rational. Social preference theories hold on to this assumption, but change the preferences: they try to explain cooperation in Prisoner's Dilemma situations by the fact that the players not only

¹¹ The Hi-Lo game is a purely game theoretic problem and of no practical relevance (see Sugden and Gold 2007), which is why I will not explain it in detail. Note also that not all Prisoner's Dilemmas become a Hi-Lo game when the pay-offs are averaged, consider for instance B = 7, A = 10, C = 2 and D = 0.

care about their material pay-offs, but also take into account the utility of other players (altruistic preferences) or have an inequality aversion¹², for example. If the all-things-considered preferences of these agents are taken into account, it is no longer a Prisoner's Dilemma.

From a normative perspective, however, these kind of models are not helpful: they do not show how rational counter-preferential choice is possible, but merely show that the cooperators are in reality not cooperators but expected utility maximisers, because when including all pay-offs they do not face a Prisoner's Dilemma anymore.

3.3. The justification problem

Why should I engage in team reasoning rather than individual utility maximisation? Team reasoning offers an explanation for why it is rational to cooperate *given* that individuals ask 'What should we do?' rather than 'What should I do?'. This leaves us with two possible questions, and for each question, the other action is the rational choice:

Question	rational action (for the individual)
What should I do?	defect
What should we do?	cooperate

However, it is unclear why the second question should be asked in the first place.

The problem is therefore *shifted* from 'How can it be rational to act counter-preferentially?' to 'How can it be rational to act as a member of a team (rather than as an independent individual)?'. I call the latter *justification problem*.

¹² see for instance Fehr and Schmidt (1999)

4. State of the Art and Research Question

4.1. State of the Art

It is an open question how we should think about the justification problem. The main economic accounts of team reasoning stem from Michael Bacharach (2006) and Robert Sugden (2015)¹³.

Michael Bacharach (2006) explains the 'choice' of teams in terms of psychological framings of the decision problem. The idea is that there are non-rational factors that determine whether individuals engage in team reasoning or in individual reasoning. The question itself is therefore not an object of rational choice. Accordingly, although it might be an interesting account regarding the explanation or prediction of actual behaviour, it is relatively uninteresting from a normative perspective and thus for our purposes.

Robert Sugden's (2015) discussion of team reasoning deserves a somewhat deeper consideration. According to Sugden, the decision as to whether an agent engages in team reasoning or not *is* subject to rational choice and he attempts to account for it within rational choice theory.

Sugden presents team reasoning as cooperation for mutual benefit. Mutual benefit means that not only the team as a whole must benefit from cooperative behaviour, but also each individual. The core of Sugden's proposal is his "Schema of Cooperative Team Reasoning" (Sugden 2015, 161). Roughly, the schema states that *if* a mutually beneficial practice exists and the agent has reason to believe that all other involved agents are aware of this and will engage in cooperative behaviour, *then* the agent should also act cooperatively.

The schema is a deductive model according to which it follows from certain conditions that the agent is rationally obliged to cooperate. However, why endorsing the schema in the first place? Sugden is elusively brief on this aspect: "I as author am not asserting that this schema 'really' is valid. Rather, it is a schema that any player might endorse. Were she to do so, she would take it to be valid." (Sugden 2015, 161)

Two aspects of Sugden's account are particularly important for our purposes:

First, the fact that the basic willingness to engage in a joint action is not enough to actually rationalise cooperative action. According to Sugden, this requires that the agent has reason to believe that the other players will also endorse the scheme and cooperate. Sugden is very explicit about this:

"I suggest, a person who is motivated to seek cooperation need not be committed to act on the prescriptions of team reasoning *unless she has*

¹³ Sugden has presented and developed his account in the course of various articles (e.g. 1993, 2000, 2003). Given the limited scope of this thesis, I will exclusively rely on Sugden (2015) for the discussion of his account.

adequate assurance that other members of the team will do so too” (Sugden 2015, 154, my emphasis).

This means that the initial situation is slightly different from the classic Prisoner’s Dilemma: If I have reason to believe that the other agent will be cooperating, the only question for me is whether I am participating in the mutually beneficial practice or am betraying her. If I cooperate, however, I do not risk being betrayed by the other agent (as long as my belief is true). This limits the scope of the possibility of team reasoning. If I have no justified belief about the behaviour of others or even believe that they will not cooperate, it is not rational for me to do so.

Second, Sugden does not claim that the agent is rationally obliged to endorse the schema in the first place. He merely speaks of ‘might’, which I take to mean rational permission. However, Sugden does not provide any justification as to why it is rationally permissible to endorse the schema. He is therefore unable to answer the fundamental question of whether and to what extent it can be rational to cooperate in Prisoner’s Dilemma situations or to engage in team reasoning. The justification problem remains unsolved.

Gold and Colman (2020) address the search for principles of rational team identification. They claim that the most promising place to look for them lies outside the realm of rationality, namely in ethics or morality – “except maybe in a thick Kantian sense of rationality where the moral law is a truth that can be discovered by reason” (Gold and Colman 2020, 314)¹⁴. However, the latter would no longer be a theory of purely instrumental rationality, as is the case with RCT¹⁵. Gold and Colman argue that there are “insuperable obstacles to solving the problem of rational group identification *within* a theory of purely instrumental rationality” (Gold and Colman 2020, 314, my emphasis). The reason for this is that instrumental rationality can only provide the rational action relative to a given goal. However, every goal must be someone’s goal, which presupposes a unit of agency. This means that the question of what the unit of agency should rationally be already assumes a unit of agency. In view of these considerations, an examination of more substantial theories of rationality is certainly promising.

In that context, Elizabeth Anderson’s (2001) approach is interesting: she attempts to make a concept of social identity based on Korsgaard’s notion of practical identity (1996) fruitful for the rational justification of cooperation in Prisoner’s Dilemma situations. In doing so, she extends the concept of rationality and reasons for action employed by RCT. Her proposal has not received much attention in the team reasoning literature. It is for instance mentioned in

¹⁴ It is noteworthy that Gold and Colman mention Korsgaard (1989), but only with regard to her arguments on personal identity and the rationality of pursuing projects over time. They do not refer to her account of practical identity developed in Korsgaard (1996) and later works.

¹⁵ RCT is concerned with instrumental rationality only in the sense that it considers ends (or preferences) to be given and therefore not up for rational evaluation.

Gold and Sugden (2007), Smerilli (2012), Schmid (2021), Roth (2017) or Gold (2022), but none of the authors deal with her proposal in depth.

Before proceeding, I should mention an important caveat: there is a debate about whether it is even possible to rationally intentionally decide whether to ask ‘What should we do?’ or ‘What should I do?’. While Tuomela (2013), for example, speaks in favour of this possibility, other authors like Townsend (2015) and Pacherie (2011) are sceptical. I will not discuss this aspect in more detail and assume that it is possible that the question of the mode of reasoning is an object of rational choice.

4.2. Research Question

In this thesis, I ask the following question:

RQ: Can the concept of social, practical identity solve the justification problem?

To address this question, I shall examine and evaluate Anderson’s account with her proposed solution for the justification problem.

4.3. Analysis of the justification problem

In order to analyse and evaluate Anderson’s account, we need to structure the justification problem in more detail. Recall the formulation from [section 3.3](#):

How can it be rational to act as a member of a team (rather than as an independent individual)?¹⁶

In view of the considerations regarding Sugden’s account from [section 4.1](#), we must differentiate between two steps. The first step is the endorsement of the schema. This means to “dispose oneself to treat N as a unit of agency and to play one’s part in its joint actions” (Sugden 2015, 162). Sugden explicitly equates this step with the identification with a team: “The concept of ‘endorsing and acting on the Schema of Cooperative Team Reasoning’ is the analogue of group identification in Bacharach’s theory” (Sugden 2015, 162). The second step presupposes the first one: once the schema is endorsed, the question still stands whether it is rational to engage in cooperative behaviour. For this, according to Sugden, the conditions elucidated in [section 4.1](#) have to be fulfilled: there has to be the possibility of mutual interest and the agent needs to have reason to believe that the other players will cooperate.

I will structure the justification problem according to the following two steps:

- step 1:* rationality of team identification
- step 2:* rationality of team reasoning

¹⁶ Strictly speaking, there is a conceptual difference between *acting* as a member of a team and *reasoning* as a member of a team, i.e. asking ‘What should we do?’. For simplicity, I will not distinguish between the two, which amounts to the assumption that if I engage in team reasoning, I will also act accordingly and vice versa.

The *rationality of team reasoning* pertains to the question whether it is rational to cooperate *given* that I already identify with a team¹⁷. Importantly, the connection between the two steps has implications for the *rationality of team identification*. If I claim for instance that team identification already commits me to cooperate (and act against my preferences), this identification might be harder to justify than if further conditions¹⁸ still have to be fulfilled to commit me to cooperate once I have identified. Therefore, the connection between *step 1* and *step 2* is also relevant to the justification of *step 1*.

The structure of the *rationality of team identification (step 1)* can be refined further, i.e. broken down into three levels¹⁹:

Level 1: Why is it rational to identify as a member of any team in the first place?

It is not clear whether I am rationally permitted or even obliged to act by virtue of being a member of any team.

Level 2: Assuming that it can be rational to identify as a member of a team in the first place, which concrete teams is it rational to identify with?

It might be that I am rationally permitted to be a member of the local wind band, but not of some dangerous cult that expects me to leave my beloved ones. Assuming that I justifiably identify with a set of several teams, e.g. the local wind band, some political party and the local football club, a further question arises:

Level 3: How can I rationally prioritise between the teams of my set (as determined by *level 2*) and my personal utility maximisation in cases of conflict like the Prisoner's Dilemma?

Imagine I have a football game at exactly the time when the ballots are open. My personal utility maximisation as well as team reasoning from the perspective of the football club speak in favour of not voting and instead going to the game. However, from the perspective of me being a party member, I have to vote. *Level 3* asks for rational prioritisation between teams which it is already rational for me to identify with in the first place. Therefore, with respect to the rationality of team identification, *level 1* is prior to *level 2*, which is prior to *level 3*. This subsection leaves us with the following structure for the justification problem (figure 1):

¹⁷ The steps are often not explicitly delineated in other team reasoning accounts. For example, in Bacharach's account, "once an individual has identified with a team, his willingness to act on team reasoning is not conditional on any assurance that other team members will do the same" (Sugden 2015, 154). Still, as Sugden's account shows, there is a conceptual difference between team identification and team reasoning.

¹⁸ like for example those from Sugden's account

¹⁹ For clarificatory purposes, the levels are formulated as questions. However, they are not intended as proper sub-questions of my RQ, but rather as a means to structure the problem.

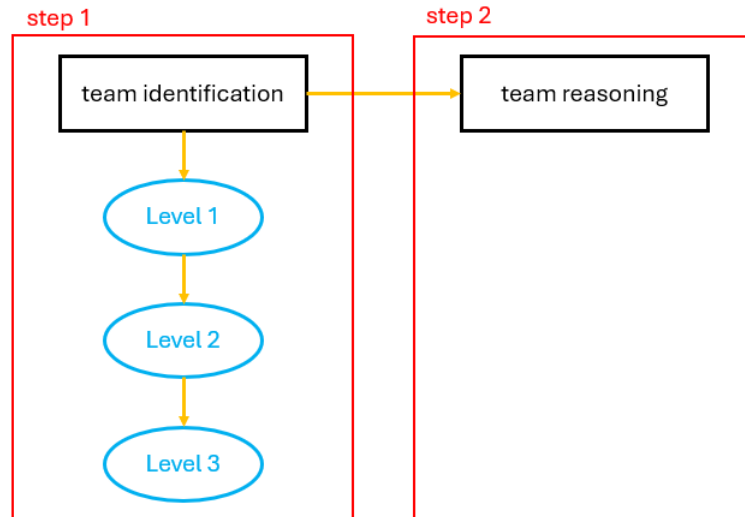


figure 1: structure of the justification problem

The question how it can be rational to act as a member of a team instead of as an individual first pertains to the rationality of team identification and then to the rationality of team reasoning (depending on how team identification and team reasoning are connected). Further, the rationality of team identification can be structured according to three consecutive levels.

5. Elizabeth Anderson's proposal

According to Anderson, in order to understand committed action as rational, a theory of rationality must be extended to include a non-act-consequentialist principle of rational choice. I claim that her proposal for the latter is best understood within a Korsgaardian framework of reasons. Accordingly, I will briefly present it in the following subsection before introducing Anderson's choice principle and conception of social identity.

5.1. A rough reconstruction of the Korsgaardian conception of a reason

Korsgaard's account (1996) of the concept of a reason starts with the statement of the reflective nature of our mind: we do not just follow our impulses and desires but have a distance to our mental activities. We therefore need to endorse our desire in order to act upon it, which requires a justification (i.e. a reason). What happens is that the potential reason is put through a reflective endorsement test. This evaluation is done from the standpoint of a conception of our (practical) identity: "the principle or law by which you determine your actions is one that you regard as being expressive of yourself" (Korsgaard 1996, 100). Given that I identify as a student, can I endorse the desire to skip class to go to a party? If the answer is yes, I have a sufficient reason to go to the party (insofar as I identify as a student). If it is not, I have a rational obligation not to go to the party. 'Reason' thus refers to "reflective success" (Korsgaard 1996, 93) and is derived from rationality. figure 2 illustrates the framework:

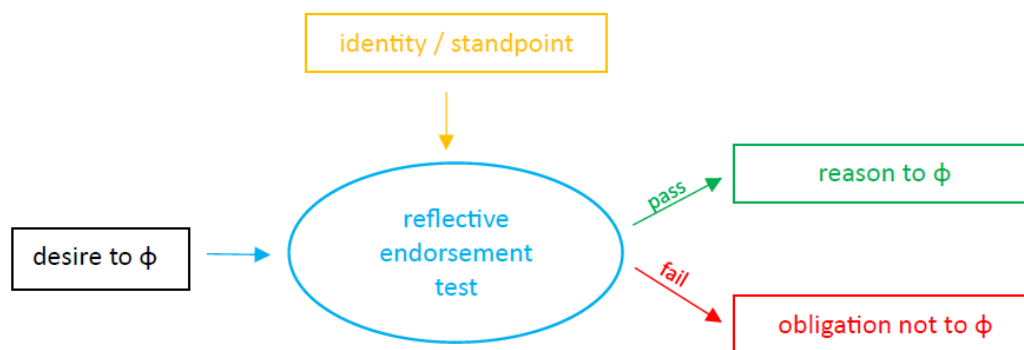


figure 2: the reflective endorsement test

Based on this conception of rationality and reasons, I will present Anderson's team reasoning account.

5.2. Social Identity and the Universalisability Test

Anderson understands practical identity as social in the sense that it is (at least partly) "constituted by membership in various social groups" (Anderson 2001, 28), whereby she refers to diverse writings on identity of Amartya Sen. The idea is the following: when I deliberate in the face of a choice situation, my decision is based on how I conceive of myself.

This practical self-conception is further organised in such a way that it is composed of affiliations to different social groups. Anderson writes:

“People may also identify with their occupation, the firm where they are employed, various associations and clubs, their nation, caste, religion and so forth, and choose on the basis of these identities” (Anderson 2001, 28).

Anderson insists that this identity is practical, not ascriptive: it refers to how the agent conceives of herself, not to how one is conceived by others²⁰.

When I act by virtue of being a member of some social group, I regard myself “as acting in concert with the other parties, as a single body” (Anderson 2001, 28). Referring to Gilbert’s (1990) concept of collective agency, Anderson argues that the member’s commitment to only act on reasons that are universalisable to all members of the group is a constitutive principle of a collective agent.

Interpreted within the Korsgaardian framework this means: from the standpoint of my identity as a member of the local wind band, the reflective endorsement test amounts to checking whether my potential reason (e.g. skipping a repetition for a friend’s birthday party) could be a reason for all members of the group. figure 3 illustrates the process:

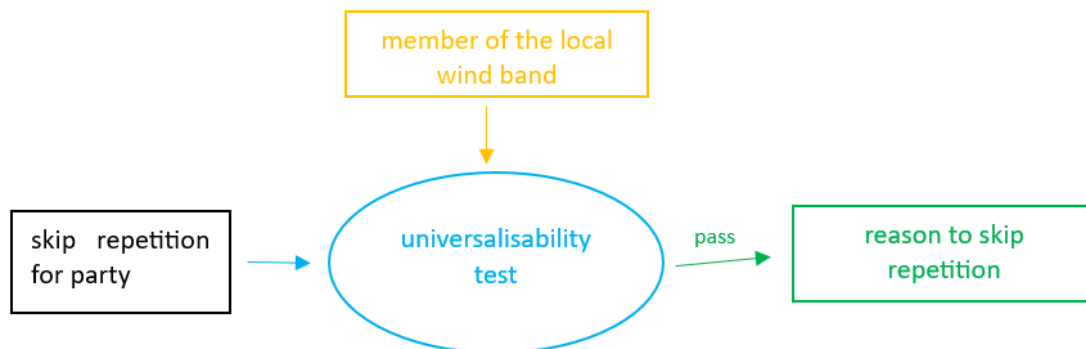


figure 3: universalisability test; standpoint of the identity as a local wind band member

Assume that not attending a repetition once due to a friend’s birthday party is universalisable among all the members of the local wind band. Therefore, I am rationally permitted to skip the repetition and attend a friend’s party instead.

²⁰ Anderson does not elaborate further on the extent to which someone can decide on their identity and the extent to which this depends on ascriptive identities. There are two points worth mentioning at this point from skimming Sen's writings on identity: First, he makes it clear that there is an important role for reasoning and deliberation in social identities, on the one hand with regard to the choice of identities, on the other hand with regard to their relative importance (which is contextual) (see for instance Sen (2006, chapter 2), Sen (1999)). Second, however, he also makes it clear that the choice takes place within certain constraints. An important type of constraint is how others see us or insist to take us (Sen 2006, 31). This suggests that ascriptive and practical identities are probably not so clearly distinguishable.

Anderson claims that the identification with the team constitutively commits me to act on the universalisability principle. I understand her saying that applying the universalisability principle is part of what it means to look at a decision situation from the perspective of the identity as a member of a team. This means that for the justification problem only the rationality of team identification (step 1) is relevant for Anderson, as it implies the rationality of team reasoning (step 2)²¹. Before the presented account can be applied to Prisoner's Dilemma situations, I would like to make two comments.

First, Anderson's account does not imply that expected utility theory is discredited as a principle of choice. Insofar as I identify as "independent agent" (Anderson 2001, 32), it is still rational to maximise my expected utility. It is not clear how this independent agent relates to the group identities, as Anderson does not give further explanation. I take it to refer to the individual standpoint abstracting from all social identities, along the following lines: independently of all my commitments to social groups, what do I prefer? An example would be: if there is an event of the local wind band today where a lot of friends are coming too, from the standpoints of me being a friend group member and me being a wind band member, I should go there. But abstracting from those identities, I just do not feel like going out tonight and would prefer to stay home. At this point, two comments should be addressed in relation to the independent agent, the importance of which will become apparent in due course. Firstly, given that the independent agent is the utility maximiser, the question of what choice-relevant considerations are included in the independent agent's perspective relates to the question of how broadly we understand 'utility' and 'preferences' mentioned in [section 2.3.1](#). Secondly, it is unclear whether Anderson understands this standpoint of the utility maximiser to simply be the smallest group with only one member (as the economist proponents of team reasoning generally do) or to be a different kind of reasoning.

Second, it might seem that Anderson's Kantian approach differs from the economic team reasoning accounts because it is about universalisability rather than expected utility maximisation – whereas the economic accounts aimed to keep expected utility maximisation as the rational choice principle. However, they are compatible: universalisability is the reflective endorsement test on the individual level insofar as the

²¹ Nevertheless, Anderson is concerned with the role of the knowledge that others are cooperating. For her, however, this takes place at the level of the rationality of identification, not with identification already present: "Common knowledge of everyone's (rational) conditional willingness to join together is sufficient, but not necessary, to make identification with the others rational. Where such common knowledge is absent, it could still make sense to take others on trust" (Anderson 2001, 32) and regard oneself as part of an "imagined common agency" (Anderson 2001, 32). Anderson therefore seems less strict than Sugden, but it seems plausible that trust requires reasons to trust as well. It remains unclear what the implications of these considerations are: one possible answer would be to say that if there is not enough reason to trust, it is irrational to identify with the team (if we were to speak of a team at all). I will not discuss this further. For a discussion regarding the role and rationality of beliefs about what others will do in the context of team reasoning, see Roth (2014).

individual identifies with the team. On the team level, it might well be that the joint action leading to the maximisation of the team's utility is endorsed by this test. For this reason, Anderson's approach does not undermine the aim to keep expected utility maximisation as principle of choice for the respective unit of agency.

5.3. Social Identity and Prisoner's Dilemmas

In what follows, I will apply the account of social identity and the universalisability test to Prisoner's Dilemma situations.

If I do not identify with the team, I am still rationally required to defect, as expected utility maximisation asks me to. If I do identify with the team, I have to apply the universalisability principle. In that case, it depends on whether maximising my expected personal utility is compatible with the team's goal in the relevant context. In Prisoner's Dilemma situations, this is never the case: by definition, the payoff structure is such that collective and individual utility conflict. Let me return to the problem of voting from [section 2.2](#). The members of a political party agree that the party's goal is to get their candidate elected. Nonetheless, each of their individual votes is negligible: from the individual's perspective, the costs of voting exceed its potential payoff. However, this way of reasoning is not universalisable: if every member of the party were to act based on that principle, this would undermine the party's goal, as no one would vote and the candidate would not get elected. This can be illustrated as in figure 4²²:

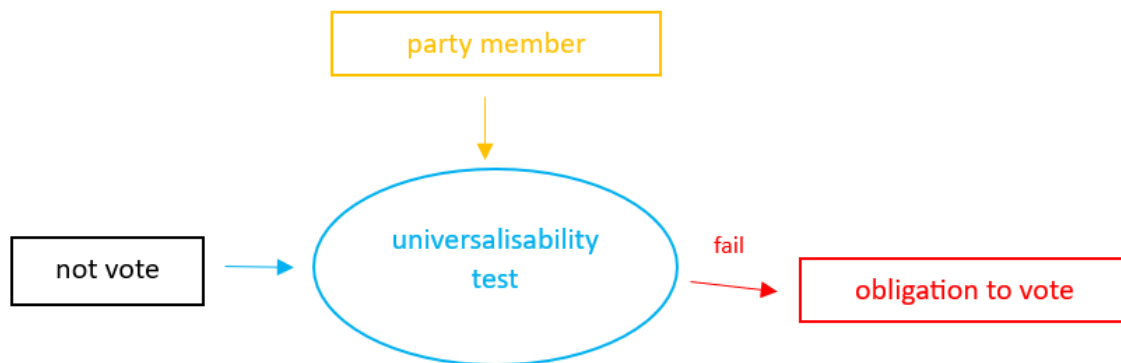


figure 4: universalisability; the standpoint of the identity as a party member

However, at the same time, the relevant individual is also an independent agent. As just seen, from this perspective, voting is irrational. This yields (figure 5):

²² The obligation to vote must be understood with respect to the given circumstances. It does not exclude that there might be some circumstances in which it would be acceptable to refrain from voting even from the perspective of the identity as a party member. For instance, not voting due to severe illness is plausibly universalisable amongst all party members.



figure 5: universalisability test; standpoint of the identity as an independent agent

Thus we can represent the Prisoner's Dilemma according to Anderson's account as follows (figure 6):

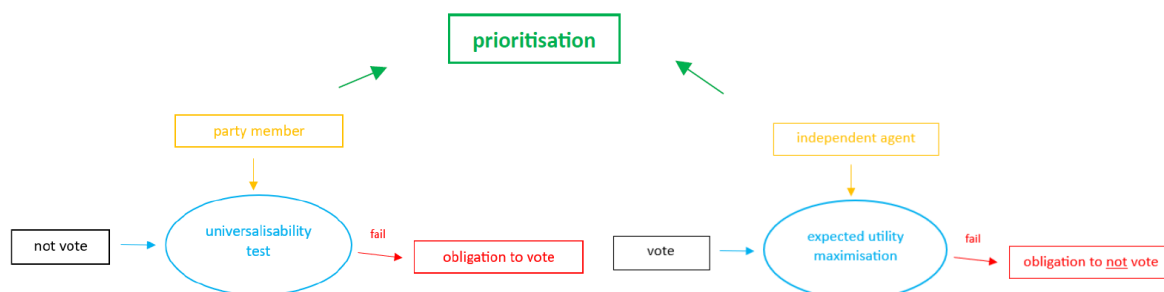


figure 6: Prisoner's Dilemma in Anderson's framework

5.4. Back to the Rationality of Team Identification

After elucidating Anderson's framework and conception of social identity, I now turn to the core of my thesis. I will examine, based on the structure developed in [section 4.3](#), whether Anderson can solve the justification problem. As explained in [section 5.2](#), for Anderson, the rationality of team identification (*step 1*) implies the rationality of team reasoning (*step 2*). Accordingly, I will only examine her account with regard to step 1. Recall the structure of the rationality of team identification from [section 4.3](#): *level 1* pertains to the question as to why it is rational to identify as a member of any team in the first place. Assuming an affirmative answer to *level 1*, *level 2* asks which concrete teams it is rational to identify with. Finally, *level 3* is concerned with the rational prioritisation between (rational) identities in the Prisoner's Dilemma.

I will first examine Anderson's account with regard to *level 1* ([section 5.4.1](#)). I will then examine *level 3*, since Anderson's proposed solution for prioritisation also includes her perspective pertaining to *level 2* ([section 5.4.2](#) and [section 5.4.3](#)).

5.4.1. The priority principle

Anderson addresses *level 1* with what I call the *priority principle*: she accords priority to social identity over the mode of reasoning: 'what principle of choice it is rational to act on

depends on a *prior* determination of personal identity, of who one is' (Anderson 2001, 30, my emphasis). This 'prior' refers to the order of justification. To act according to expected utility maximisation is rational *conditional* on me identifying as an independent agent.²³

The priority principle explicitly constitutes an answer to *level 1* in the following way:

"What the priority of identity to rational principle [*i.e. the priority principle*] does do is establish a rational permission to identify with others and join in a common agency" (Anderson 2001, 31).

Anderson does not explain in detail why this principle establishes a rational permission. As I understand it, the priority principle is intended as a response to scepticism about team identification from the RCT perspective: it is implausible that it is rational for me to identify with a team *because it goes against my utility maximisation*. The priority principle refutes this argument because it establishes that utility maximisation as a choice principle is only valid under the condition of a prior identification as an independent agent. It can therefore not be used to justify or criticise the choice of the identity itself. This consideration seems convincing, but only provides argument against the criticism from RCT. However, Anderson's account lacks an argument for why it is rational to conceive of oneself as a member of some team.

Nevertheless, her critique of the economic concept of the individual seems to be a possible answer to *level 1*. Within economic theory, Anderson argues, "the individual is assumed to be given, with interests and preferences defined independently of and prior to her joining any group" (Anderson 2001, 36). Anderson rejects this conception, claiming that "acquiring a self-conception as an individual requires actually conceiving of oneself as a committed member of multiple social groups" (Anderson 2001, 36). Seen in this way, social identities (which are constituted by group membership) are a prerequisite for the question of the rationality of identification to be asked at all. Therefore, it is possible to ask which teams it is rational to identify with (*level 2*) and how to rationally prioritise them (*level 3*), but not whether it is rational to identify with a team at all (*level 1*)²⁴.

These considerations are *prima facie* plausible, a more in-depth elaboration of this aspect would go beyond the scope of this thesis. Consequently, I assume (in line with Anderson)

²³ A brief comparison with Sugden is interesting. He writes: "the question of whether a particular action is rational is ill-formed unless the unit of agency has been specified: an action is rational for an agent to the extent that it can be expected to achieve that agent's objectives. Thus, the question 'Who am I?' (or 'Who are we?') is *logically* prior to rational choice" (Sugden 2015, 153, my emphasis). Sugden's claim is thus more radical than Anderson's: the priority is not only rational, but logical. This seems in line with the conceptual difficulty raised by Gold and Colman (recall [section 4.1](#)) of trying to solve team identification within purely instrumental rationality.

²⁴ Strictly speaking, this would mean that we cannot give a positive answer to the *level 1* question. Instead, it would remove *level 1* from the scope of the question about the rationality of team identification. However, this is of little consequence for the further course of the analysis, as it still permits us to proceed to *level 2* and *level 3*.

that we can motivate respectively presuppose *level 1* and proceed to examine *level 2* and *level 3*.

I now turn to *level 3*: how rationally prioritise between the two identities (see figure 6)?.

In line with Korsgaard's framework presented above, I need two ingredients to 'produce' a reason: a standpoint and a reflective endorsement test. The question therefore calls for two distinct but related considerations: first, from which standpoint do I deliberate when adjudicating between identities? Second, which considerations are valid as a basis for rational decision-making, i.e. which choice principle do I apply? In what follows, I will consider those two questions in turn.

5.4.2. The standpoint of the individual self

Anderson refers to the "identity of an individual self, who is authorized to adjudicate the conflicts among its various constitutive collective identities" (Anderson 2001, 35). She understands this as "individuality – identification as an individual person" (Anderson 2001, 36). As aforementioned, although this identity is supposed to adjudicate between different social identities, it does not exist prior to the individual joining any specific group, but emerges out of an individual being part of many different social spheres. This allows people to understand themselves as "agents authorized to set their own priorities, on their own, according to an autonomously defined self-conception" (Anderson 2001, 36). Anderson does not elaborate further on the notion of the individual self.

The relevant question is how this individual self relates to the group identities on the one hand and to the identity as an independent agent on the other. Given that it has to adjudicate between all the other identities, it has to be able to transcend them in some sense. At the same time, it is supposed to emerge from these other identities. Anderson does not elaborate on the process by which it reaches the meta-level, nor explain in what sense it is independent of other identities and how it can still be 'close' enough in order to mediate between the other identities.

The individual self also has to be different from the independent agent, because the latter is as expected utility maximiser on the 'same level' as the group identities, as illustrated in the conflict between them in Prisoner's Dilemmas (figure 6). Therefore, the individual self has to transcend the independent agent (who is already an abstraction from all other social identities) as well. Seen in this way, the individual self would abstract from all group identities. It would also abstract from my personal preferences (that are independent of my group identities) because they belong to the independent agent, which has to be transcended as well. Again, the question is connected to how broadly we conceive of 'utility' respectively what considerations are factored into 'preferences'. For instance, if we consider personal values to be included in preferences, they belong to the independent agent. However, if we do not, it seems plausible that they belong to the realm of the individual self, and therefore form a basis for decisions for and between different social identities.

5.4.3. The choice principle

How does the individual self rationally prioritise between different identities?

First of all, from the *priority principle* we can derive some indication about what kind of considerations one is allowed to take into account when deliberating about how to adjudicate between different conflicting identities in some choice situation. Namely, I cannot take into account the features of the specific choice situation that presuppose one of the identities for their evaluation: the fact that the event of the local wind band is very boring cannot influence whether I decide from the perspective of my identity as an independent agent (finding it boring) or whether I decide from the perspective of me being a wind band member.

On the one hand, this makes sense, as it would be undermining what we try to do by referring to social identities: the fact that voting has costs would not only be a reason not to vote *from the perspective of the independent agent*, but also a reason to give priority to the independent agent rather than to the membership of some political party. This would lead us back to expected utility maximisation and into the Prisoner's Dilemma trap. In that sense, I take it that the deliberation regarding the prioritisation of conflicting identities should be somewhat independent of the specific identities' viewpoint on the choice situation at hand.

On the other hand, the worry is that with this prioritisation of identity over choice, Anderson cannot sufficiently address the specific features of the choice situation: if my identity as a party member 'wins' over my identity as an independent agent, I will have to vote even in situations with exorbitantly high voting costs, for instance if I would have to fly back from a foreign country in order to go to the polls. It seems implausible that this would be rational. Anderson might counter this by saying that in the case of exorbitantly high costs, 'not voting' could even be universalised from the perspective of the party member (assuming that it affects enough few people so that it does not undermine the group goal even if everyone were to act according to this reason).

However, this is not quite my point. It is not a question of whether the renunciation of voting can be universalised in the case of exorbitantly high costs, but of whether the costs should be allowed to constitute a reason for not giving priority to an identity in a particular choice situation. The question is whether the priority principle excludes some good reasons for prioritising identities. It seems reasonable that the action demanded from the perspective of an identity may have a say in whether it is rational for me in a situation to prioritise that identity. If flying to a vacation destination is not universalisable among members of the green party, it might still be a valid reason not to prioritize my identity as a member of the green party if going to this destination is somehow really important to me. However, as explained, this line of thought threatens to undermine the point of the priority principle²⁵.

²⁵ This also seems to be about integrity. On the one hand, the adoption and prioritization of an identity is context-dependent (see footnote 20), on the other hand, it makes sense that both nevertheless have to be

Anderson herself does not explicitly deal with the consequences of the *priority principle* for the rational deliberation of the individual self. She proposes the following Kantian principle of choice for adjudicating between conflicting claims of social identities: we have to transcend all of these social identities and “identify with a community that comprehends them all” (Anderson 2001, 37). Her elaboration of this aspect is very sketchy. A footnote gives us a hint as to how she might see this. She seems to conceptualise it as a process where the agent goes up from more specific identities to more general ones (i.e. to ‘larger’ groups):

“we begin with our current, parochial identities and show how they generate practical problems that cannot be solved within the confines of those identities, or experiences that cannot be understood in their terms. The quest for a perspective that can make sense of our experiences and solve our problems leads to more and more expansive, cosmopolitan identifications” (Anderson 2001, 37).

I may see myself as a member of the local wind band and the local football club. If there is a conflict between the two (e.g. an event at the same time), I look at the problem from the next higher point of view, which includes both identities, namely being a club member. This encompassing perspective is given priority and the universalisability test is applied from its point of view. I understand this process as bringing otherwise incommensurable identities to a common denominator. The last, highest and most encompassing of these identities is ultimately the community of all people:

“To harmonize the demands of different groups requires adoption of a perspective that can coordinate them all. [...] it requires that we transcend our various parochial identities and identify with a community that comprehends them all” (Anderson 2001, 37).

Regarding the proposed procedure, two questions arise: first, to what extent can this way of reasoning contribute to conflict resolution if it is adopted? Second, why should it be adopted at all?

I begin by considering the question of how the proposed procedure helps in conflict cases. A distinction must be made here between two types of conflict. I call a conflict *vertical* when one of the two identities includes the other, such as being a human includes being a woman. Conflicts are *horizontal* when this is not the case.

subject to a certain stability: It seems okay for me to decide not to give too much weight to my identity as a member of the green party (or not to adopt it at all) because I like to go on vacation by plane. But if I give it a very high priority in all areas of my life, except when I'm about to go on vacation, that seems more of a problem. This point would be interesting to pursue further, but it is beyond the scope of this thesis.

With respect to vertical conflicts, the adjudication works by definition. Suppose my identity as a mother conflicts with my identity as a woman²⁶: as the latter is encompassing the former, it has priority. With respect to horizontal conflicts, the matter is complicated. Let me consider the conflict between the local football club and the local wind band. From the standpoint of being a member of the football club, I should go to the football club event and vice versa. Now, what does my identity as a club person, encompassing my particular club identities, have to say about this? The general affirmation that I have to take part in the clubs' events does not solve the problem. One could go further and consider even larger groups until ending at humanity. But I do not see how this could help.

It must be said, however, that the problem of horizontal conflicts is not central to the present thesis, since Prisoner's Dilemma situations by default show a conflict between the independent agent and a social identity²⁷. At this point, the question of the status of the independent agent becomes relevant: is it merely the smallest possible group or something distinct and apart from the group identities? If the former is the case, then the independent agent's group is encompassed by every other social identity and therefore every conflict involving it is a vertical conflict. The prioritisation of 'larger' groups in the resolution of conflicts seems to lead to the independent agent 'losing' in any case²⁸. It seems highly counterintuitive that this would be a rational requirement: if my identity as an employee of some company requires that I put my personal needs and desires aside (and work overtime, for example), it does not seem rational to me to prioritise this identity as a matter of principle. This applies regardless of whether it would sometimes be acceptable to prioritise my needs from the perspective of my identity as an employee (for example, because otherwise I would no longer be equally productive).

In contrast, since the independent agent as an individual is understood as abstracted, isolated, from all other identities, it could also be understood as fundamentally different and alongside these groups instead of as part of all of them. This way of conceiving it leads to other difficulties: the question arises as to how the independent agent is commensurable with the other identities and how the individual self can mediate between them. The recourse to larger groups no longer works as long as it is not clear what relationship the independent agent has to the other social identities.

²⁶ One could imagine that the role of mother requires certain forms of behaviour that are detrimental to women as a social group. For example: me being a mother might require me to be a housewife, but me being a woman might require that I earn my own money.

²⁷ I say by default because in the case of real prisoners, for example, it can also be a conflict between different social identities, see footnote 28.

²⁸ Importantly, this does not mean that I should always cooperate in Prisoner's Dilemmas. If we consider a game with real prisoners, although cooperating might be rational from the perspective of the identity as a member of the team of prisoners, confessing one's crimes might be rational from the identity as a citizen or a human being (see Anderson 2001, 32).

I will now turn to the other question: supposing for the moment that this process of considering 'larger' identities and finally the community of all humans *does* help adjudicate between conflicting identities, why should we adopt it in the first place?

The first question is why we should identify with humanity at all (*level 2*). At least *prima facie*, it seems plausible that I identify *as* a human being. But why should I identify *with* the group of all human beings, especially in light of the fact that this commits me to act only upon reasons that are universalisable among all people? Similar points can be made for other abstract identities as the identity as a woman or as an European person. It becomes even more difficult when we consider what Anderson said about the conditions of identification and the joint agency involved: is it really rational to take all other human beings on trust? In the same sense as for humanity, these considerations also apply to other such as 'woman' or 'nation'. There is a gap between *identity as* and the *identification with* regarding abstract identities.

Anderson admits that she cannot ultimately justify the identification with humanity: "I have no argument that would show that identification as a member of a universal community of humanity – a kingdom of ends – is rationally required." (Anderson 2001, 37) However, she proposes a pragmatic consideration: the procedure seems necessary if we want to resolve these conflicts. This is an instrumental consideration and leads us back into the Prisoner's Dilemma trap: independently of whether the other people involved in a collective action problem identify with humanity or not, I am better off not identifying with humanity. Anderson's argument does not solve the problem²⁹.

What is more: even if we grant that we identify with humanity, the proposed procedure is not yet justified: why should we give priority to the 'larger' group (*level 3*)? This is doubtful for two reasons: first, it would lead to the independent agent losing in any case, which is hardly the rational thing to do³⁰. Second, it might be that some of the 'smaller' groups (e.g. interpersonal relationships, friendships) are far more important for my self-conception, than belonging to more abstract groups such as women or human beings – but according to Anderson's procedure, I still would have to prioritise the abstract groups. The importance of an identity for my self-conception is plausibly a parameter that should be included in the rational weighting or prioritisation between identities. The fact that Anderson does not address this aspect adequately constitutes a lacuna in her account.

²⁹ Furthermore, the question arises as to whether it is even possible to identify oneself on the basis of purely instrumental considerations, given Anderson's concept of social identity. It seems plausible that social identity also involves a sense of belonging as well as personal values and it is unclear whether one can choose these for strategic reasons. This aspect seems interesting to me, but I will not pursue it further in this thesis.

³⁰ Under the assumption that the independent agent is considered to be a team with only the respective agent as a member. If it is not, the procedure is not even applicable in the first place, as showed above.

6. Christine Korsgaard's practical identity

We identified amongst others two main problems with Anderson's proposal. First, pertaining to *level 2*, she cannot provide adequate reason for identifying with the community of humanity³¹. Second, pertaining to *level 3* she cannot show why this identity should take priority over more local identities.

Anderson's concept of social identity is heavily based on Korsgaard's concept of practical identity. She writes: "I borrow the concept of practical identity from Christine Korsgaard" (Anderson 2001, 31). Given this, the question arises as to whether Korsgaard can tell us anything about why one should identify with humanity and, if so, why this identity should take priority over more local identities. In this section, I examine whether Korsgaard can solve these problems. I shall argue that although Korsgaard provides an argument for why we have to identify as human beings and why this identity is prioritised over more local identities, this comes at the cost of losing team reasoning.

6.1. Why the identity as a human being is necessary

According to Korsgaard, a conception of your practical identity is

"a description under which you value yourself, a description under which you find your life to be worth living and your actions to be worth undertaking" (Korsgaard 1996, 101).

In that way, I can conceive of myself as a student, as a psychiatrist, as someone's friend, as a member of a certain family, a club, a nation – or as the "steward of my own interests" (Korsgaard 1996, 101). The latter seems to be the counterpart to Anderson's independent agent.

Korsgaard claims that although we have many contingent identities, it is not contingent that we have *some* practical identity. This is a consequence of the reflective structure of our mind (see [section 5.1](#)): there must be a standpoint from which we can endorse a desire or impulse, otherwise we cannot act. Korsgaard links this point about the nature of our consciousness to our identity as a human being. The idea is that part of our identity as a human being is that we need practical identities:

"Most of the time, our reasons for action spring from our more contingent and local identities. But part of the normative force of those reasons springs from the value we place on ourselves as human beings who need such identities. In this way all value depends on the value of humanity: other forms

³¹ As explained above, this justification difficulty also applies to other abstract identities such as gender or race. I will only consider identification with humanity here. First, because Korsgaard provides an argument for it. Second, if the problem can be solved for humanity, which is the most abstract of the identities under consideration, it can plausibly be solved for all other abstract identities.

of practical identity matter in part because humanity requires them” (Korsgaard 1996, 122).

Korsgaard’s argument has two implications. First, the identity as a human being (which Korsgaard denotes as moral identity) is necessary: as all contingent identities get part of their normative force from the moral identity, endorsing contingent identities presupposes the endorsement of the moral identity. The argument is transcendental: if one adopts any practical identity (and one has to do so in order to act), one has to adopt the identity as a human being. This enables Korsgaard to solve the question of identification with humanity: we not only have sufficient reason, but even a rational obligation to identify as a human being.

Second, the identity as a human being is the most fundamental of all identities. This provides a reason why this identity has priority in cases of conflict³². One consequence of this is that in all Prisoner’s Dilemma situations, in which from the perspective of a human being the cooperative action is required, we have a rational obligation to cooperate³³.

So, whereas Anderson cannot motivate why we should identify with humanity and why this identity should be prioritised, Korsgaard provides an argument pertaining to both points: Identifying with humanity is a rational obligation because it is the normative foundation of all other practical identities. Accordingly, this identity must also be prioritised in cases of conflict³⁴.

6.2. Korsgaard and team reasoning

Korsgaard sometimes speaks of identity *as* (e.g. a student, a psychiatrist), sometimes of identity *as a member of* (e.g. a nation): she does not conceptually differentiate the two. Although she once speaks of “groups” (Korsgaard 1996, 101) and refers to the identity as a “member of the party of humanity” (Korsgaard 1996, 127), she does not understand practical identity as fundamentally constituted by membership in social groups. Even when considering the identity *as a member of*, this is not about joint action. Rather, action based on a practical identity pertains to some in-built normativity in specific roles. For example,

³² Whereas my identity as a member of the local wind band gets part of its normative force from my identity as a human being, it is unclear whether it also gets part of its normative force from other more general identities like the identity as a club member. Intuitively, it seems plausible that part of the value of being a member of some specific club comes from the fact that club membership itself has value. However, this would have to be worked out in more detail, which would go beyond the scope of this master thesis.

³³ It should be noted that Korsgaard’s Kantian arguments require more substantive philosophical commitments about identity, rationality and its connection to morality than Anderson’s. Anderson explicitly refrains from adopting a ‘stronger’ Kantianism. However, she also does equate reasons based on the identification with humanity with morality. For an in-depth discussion of the differences between Anderson and Korsgaard with respect to the connection between morality and rationality as well as practical identity, see Pauer-Studer (2006).

³⁴ Korsgaard seems to soften this point somewhat when she talks about relationships and their importance to our practical identity (see Korsgaard 1996, 126). I will not discuss this further.

she refers to some action being “an essential part of the idea of being a student” (Korsgaard 1996, 106). The reflective endorsement test is “seeing whether the maxim of acting on it can be willed as a law by a being with the identity in question.” (Korsgaard 1996, 113).

Accordingly, although we can rationalise ‘cooperative’ actions within Korsgaard’s account this is not joint action: go voting is simply part of what it means to be a citizen. Whereas in Anderson’s account the identities provide a rational foundation of team reasoning and thus cooperation, Korsgaard ‘skips’ the team part. Therefore, for Korsgaard, acting based on a practical identity has nothing to do with team reasoning. So, we face a trade-off: Korsgaard can give us a reason for the identity as a human being and its priority, but in return we sacrifice team reasoning.

Can we still use Korsgaard’s arguments to solve Anderson’s problems? The difficulty lies in that there is a gap between *identity as* (some social role) and *identification with* a team. If we want to transfer Korsgaard’s arguments for the necessity of identity as a human being and its priority to Anderson, we have to clarify how those two notions relate to each other. Given that Korsgaard does not conceptually differentiate between them, her account cannot shed light on their connection. These points can be illustrated with figure 7:

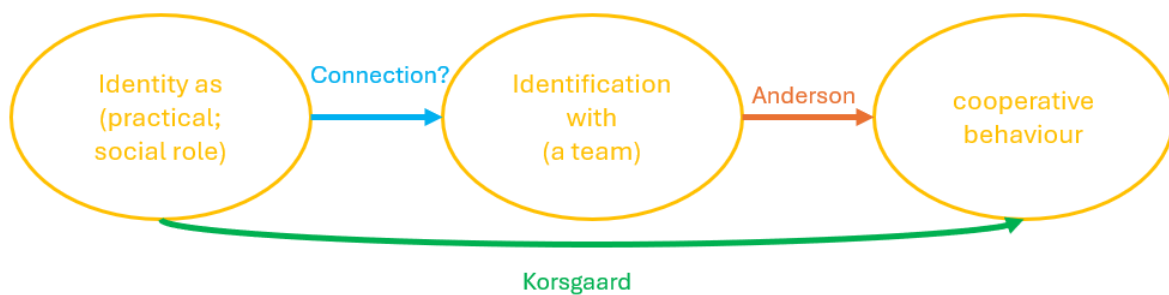


figure 7: comparison Anderson and Korsgaard

To sum up, Korsgaard provides an argument for why we have to identify as human beings and why this identity is prioritised over more local identities. However, this comes at the cost of losing team reasoning. If we wanted to make Korsgaard fruitful for team reasoning, we would need to offer an account of how *identity as* and *identification with* are connected.

7. Summary and further thinking

Let me return to the research question.

RQ: Can the concept of social, practical identity solve the justification problem?

Anderson's proposal regarding the justification problem can be synthesized as follows: *step 1* constitutively involves *step 2*, as team identification commits the agent to act on the universalisability principle (i.e. engage in team reasoning). She therefore has to argue for the rationality of *step 1* in order to address the justification problem. With respect to *level 1*, she presents the priority principle, which works against the challenge from RCT, but not as an argument for the rationality of team identification tout court. With respect to *level 2* and *level 3*, she claims that we should identify with humanity and that the adjudication process should consist of going up to 'larger' groups whereas the larger groups have priority. She cannot adequately justify either point. Further, there is also a problem regarding the priority principle: whereas it seems important to avoid the Prisoner's Dilemma trap, it might be too restrictive with respect to the reasons it allows for identity prioritisation (*level 3*).

Thus Anderson's concept of social, practical identity cannot satisfactorily solve the justification problem. The recourse to Korsgaard does provide us with an argument why we are rationally obligated to identify with humanity and why this identity takes priority – but only at the cost of sacrificing team reasoning.

The most important structural insights gained from this thesis on the justification problem can be summarised using figure 8:

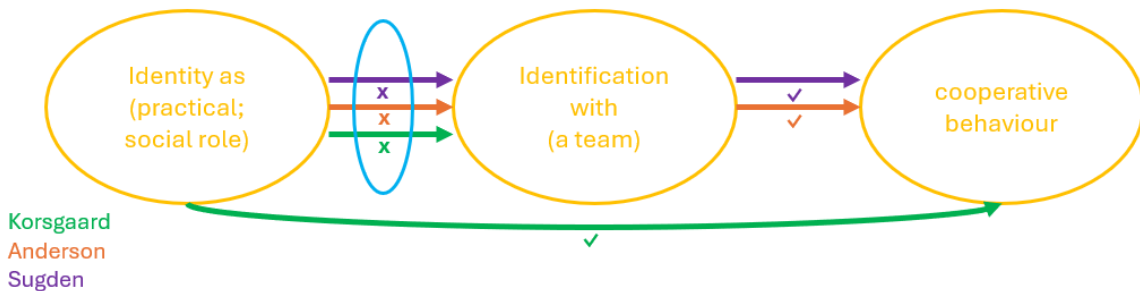


figure 8: comparison Sugden, Anderson and Korsgaard

Sugden (violet) can show under what circumstances it is rational to cooperate given that the identification with the team has taken place (i.e. the agent endorses the schema). However, he cannot adequately show why this identification with the team is rational in the first place³⁵. Anderson (orange) can show why cooperative behaviour is rational given that I

³⁵ Strictly speaking, Sugden is not talking about social practical identity. I have nevertheless placed the arrow in this way because I am concerned with this connection in this thesis. The cross therefore does not mean that Sugden's account of the relation between *identity as* and *identification with* is flawed, but that he is not concerned with it at all and therefore does not provide an explanation.

identify with the team, but it is very unclear why I should do this in the first place and how this relates to *identity as*, especially for abstract identities like being a human. Korsgaard (green) can show why *identity as* can commit me to cooperate, but *identification with* is skipped, as she does not conceptually differentiate between the two and her account does not involve team identification in the strict sense. The rationality conditions of team identification and their connection with *identity as* thus remain open (blue circle). Whereas this connection is plausibly easy to bridge for small, actual teams (such as the local wind band), it becomes more difficult for abstract identities. However, teams connected to abstract identities like humanity is what we need to tackle many of the important collective action problems like for instance climate change. The nature and rationality of team identification and its connection to practical identity are therefore open for further research³⁶.

Prisoner's Dilemmas mostly refer to a conflict between the independent agent and some possible social identity. The clarification of the nature and role of the independent agent is therefore crucial to solve the conflict. In Anderson's proposal, we identified two open problems in this respect. First, the interpretation of utility and preferences. The independent agent is the expected utility maximiser. Depending on how 'broad' our reading of the terms is, different choice relevant considerations (like for instance personal values) belong to the independent agent or not. This is also relevant for understanding the standpoint of the individual self and the adjudication process. Second, relatedly, it is unclear how the independent agent relates to the other social identities, i.e. whether it is simply the smallest group or something different. As some commensurability seems to be presupposed for any adjudication, this question is crucial to solving conflict situations.

³⁶ After cursory research, there seems to be some recent philosophical work on group identification in the context of shared agency, for instance Salice and Miyazono (2020). Also Björn Petersson (2017) and Blomberg and Petersson (2024) do speak of group identification, in the sense of "viewing the decision situation from the collective's perspective" (2024, 15). However, upon skimming the literature, I could not find a discussion of the *rationality* of group identification nor its connection to practical identity. With respect to Petersson, this might be due to the fact that he seems to be very sceptical regarding the possibility of rationally choosing to group identify at all (see Petersson 2017, 19-20).

8. Limitations

Finally, I would like to briefly discuss four selected limitations of the present thesis.

First, Anderson introduces in her account a more substantial concept of rationality than is the case with RCT. It is not just about instrumental rationality, but the universalisability criterion also involves the rational evaluation of preferences from the standpoint of different social identities. Proponents of RCT may object that this leaves the RCT-ballpark and that the proposal is therefore unmotivated. At this point, the extent to which and at what cost it might make sense to adopt a more substantial conception of rationality and reasons must be discussed. On the one hand, a concept of rationality that involves as little philosophical commitments as possible is a plausible desideratum; on the other hand, as explained above, it is probably not sufficient to solve prisoner's dilemma situations and collective action problems. A detailed discussion of this matter could not be conducted in the present thesis.

Second, it might be argued that the influence of identity on choice should simply be factored into the pay-offs of the respective agents. If this works, the question arises as to why practical identities should be understood as different standpoints associated with agency transformations instead of simply integrating them into preferences. In economics, there has been some work on the influence of identity on economic decisions, most prominently by George Akerlof and Rachel Kranton (2000). The scope of this thesis does not allow for an adequate engagement with their theory. I would just like to offer a brief consideration: from a normative perspective, such a theory seems *prima facie* uninteresting, because the decision problem shifts to the determination of one's own preferences and, since RCT takes these as given, outside the realm of rational decision³⁷.

Third, I want to raise a potential worry with respect to team reasoning more generally. Recall the argument from inefficiency ([section 2.3.2](#)) that motivated the search for a way to rationalise cooperative behaviour in Prisoner's Dilemma situations. Even if we can show that I can rationally identify with a team and that this in turn rationalises cooperative action, we do not seem to be able to do justice to the problem of inefficiency: if we do not identify with the team, it is still rational to defect and the inefficiency remains. If we do identify with the team, we are efficient, but no longer relative to our personal preferences but relative to those of the group. Team reasoning thus seems to miss the point of the problem of inefficiency. Further thinking would be needed to examine whether this is actually the case.

Fourth, team reasoning is commonly understood as an instance of joint action, involving an agency transformation from the individual to the team. An in-depth analysis would therefore require to engage with the body of literature surrounding questions such as what it means to intend something collectively, to be a plural subject or to have team reasons. Especially important would for instance be the question of what it means to have an

³⁷ see also comments on social preference theories from [section 3.2](#)

individual reason for action by virtue of being member of a team. Two influential approaches within the joint action literature are considered to be related to team reasoning (see Petersson 2017): ‘we-mode of intending’ by Raimo Tuomela (2013) and ‘shared intentions’ by Michael Bratman (2014). An adequate examination of this literature would have gone beyond the scope of this master thesis, which is why I have not considered the connection between team reasoning and joint agency in detail.

9. Conclusion

This thesis has examined whether the concept of social practical identity can solve the justification problem. Anderson's proposal in this regard has been analysed and I have shown why it cannot provide an adequate solution. I thereby identified two main issues: first, she cannot motivate why we should identify with humanity and why this identity should be prioritised over other social practical identities. Second, she lacks an adequate account of the independent agent. Further, I argued that Korsgaard can offer an argument for the first issue, but only at the cost of sacrificing team reasoning.

The major open questions that remain are the question of rational identification with a team, in particular its connection with *identity as* and the nature and role of the independent agent. The latter concerns the conception of utility and preferences on the one hand, and the connection between the independent agent and the other social identities as well as the individual self on the other.

Bibliography

- Akerlof, G. A., & Kranton, R. E. (2000). Economics and identity. *The Quarterly Journal of Economics*, 115(3), 715–753.
- Anderson, E. (2001). Symposium on Amartya Sen's philosophy: Unstrapping the straitjacket of 'preference': A comment on Amartya Sen's contributions to philosophy and economics. *Economics and Philosophy*, 17(1), 21-38.
- Bacharach, M. (2006). *Beyond individual choice: Teams and frames in game theory*. Princeton University Press.
- Bratman, M. (2014). *Shared agency: A planning theory of acting together*. Oxford University Press.
- Brennan, G., & Brooks, M. (2015). The role of numbers in prisoner's dilemmas and public good situations. In M. Peterson (Ed.), *The prisoner's dilemma: Classic philosophical arguments* (pp. 177-198). Cambridge University Press.
- Blomberg, O., & Petersson, B. (2024). Team reasoning and collective moral obligation. *Social Theory and Practice*, 50(3), 483-516.
- Bovens, L. (2015). The tragedy of the commons as a voting game. In M. Peterson (Ed.), *The prisoner's dilemma: Classic philosophical arguments* (pp. 156-176). Cambridge University Press.
- Fehr, E., & Schmidt, K. (1999). A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics*, 114, 817–868.
- Gilbert, M. (1990). Walking together: A paradigmatic social phenomenon. *Midwest Studies in Philosophy*, 15(1), 1-14.
- Gold, N., & Sugden, R. (2007). Collective intentions and team agency. *Journal of Philosophy*, 104(3), 109-137.
- Gold, N., & Colman, A. M. (2020). Team reasoning and the rational choice of payoff-dominant outcomes in games. *Topoi*, 39(2), 305-316.
- Gold, N. (2022). Guard against temptation: Intrapersonal team reasoning and the role of intentions in exercising willpower. **Noûs*, 56*(3), 554-569.
- Hardin, R., & Cullity, G. (2020). The free rider problem. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Winter 2020 Edition). Retrieved from
- Kagan, S. (2011). Do I make a difference? *Philosophy and Public Affairs*, 39(2), 105-141.
- Korsgaard, C. (1989). Personal identity and the unity of agency: A Kantian response to Parfit. *Philosophy and Public Affairs*, 18(2), 103-131.
- Korsgaard, C. M. (1996). *The sources of normativity*. Cambridge University Press.
- Pacherie, E. (2011). Framing joint action. *Review of Philosophy and Psychology*, 2, 173–192.
- Pauer-Studer, H. (2006). Identity, commitment and morality. *Journal of Economic Methodology*, 13(3), 349–369.

- Peterson, M. (2015). Introduction. In M. Peterson (Ed.), *The prisoner's dilemma: Classic philosophical arguments* (pp. 1-6). Cambridge University Press.
- Petersson, B. (2017). Team reasoning and collective intentionality. *Review of Philosophy and Psychology*, 8(2), 199-218.
- Roth, A. S. (2014). Team reasoning and shared intention. In A. Konzelmann Ziv & H. B. Schmid (Eds.), *Institutions, emotions, and group agents: Contributions to social ontology* (pp. 279–295). Dordrecht: Springer.
- Roth, A. S. (2017). Shared agency. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Summer 2017 Edition).
- Salice, A., & Miyazono, K. (2020). Being one of us: Group identification, joint actions, and collective intentionality. *Philosophical Psychology*, 33(1), 42-63.
- Schmid, H. B. (2021). Collective rationality. In M. Knauff & W. Spohn (Eds.), *The handbook of rationality*.
- Sen, A. (1977). Rational fools: A critique of the behavioral foundations of economic theory. *Philosophy and Public Affairs*, 6(4), 317-344.
- Sen, A. (1999). *Reason before identity: The Romanes lecture*. Oxford University Press.
- Sen, A. (2006). *Identity and violence: The illusion of destiny*. Norton.
- Smerilli, A. (2012). We-thinking and vacillation between frames: Filling a gap in Bacharach's theory. *Theory and Decision*, 73(4), 539-560.
- Sugden, R. (1993). Thinking as a team: Towards an explanation of nonselfish behavior. *Social Philosophy and Policy*, 10(1), 69-89.
- Sugden, R. (2000). Team preferences. *Economics and Philosophy*, 16(2), 175-204.
- Sugden, R. (2003). The logic of team reasoning. *Philosophical Explorations*, 6(3), 165–181.
- Sugden, R., & Gold, N. (2007). Theories of team agency. In F. Peter & H. B. Schmid (Eds.), *Rationality and commitment*. Oxford University Press.
- Sugden, R. (2015). Team reasoning and intentional cooperation for mutual benefit. *Journal of Social Ontology*, 1(1), 143–166.
- Thoma, J. (2019). Decision theory. In R. Pettigrew & J. Weisberg (Eds.), *The open handbook of formal epistemology* (pp. 57-106). PhilPapers Foundation.
- Townsend, L. (2015). Review of B. Raimo Tuomela: Social ontology: Collective intentionality and group agents. *Journal of Social Ontology*, 1(1).
- Tuomela, R. (2013). *Social ontology: Collective intentionality and group agents*. Oxford University Press.

List of figures

figure 1: structure of the justification problem	14
figure 2: the reflective endorsement test.....	15
figure 3: universalisability test; standpoint of the identity as a local wind band member ..	16
figure 4: universalisability; the standpoint of the identity as a party member	18
figure 5: universalisability test; standpoint of the identity as an independent agent	19
figure 6: Prisoner's Dilemma in Anderson's framework.....	19
figure 7: comparison Anderson and Korsgaard	28
figure 8: comparison Sugden, Anderson and Korsgaard	29

List of tables

table 1: Prisoner's Dilemma, standard form	4
table 2: Prisoner's Dilemma, payoff transformation and agency transformation	8

Appendix: Abstracts

Abstract (English)

Team reasoning constitutes a prominent solution in economics to explain cooperation in Prisoner's Dilemma situations: agents cooperate insofar as they reason as members of a team. From a normative perspective, this raises the question of why the agent should engage in team reasoning rather than individual utility maximisation (justification problem). Team reasoning accounts have either ignored the normative problem (Bacharach 2006) or unsuccessfully attempted to solve it within the realm of instrumental rationality (Sugden 2015). The problem is therefore still unsolved (see Gold and Colman 2020).

Elisabeth Anderson (2001) proposed an interpretation of team reasoning where what reasons we have depends on our social identities. Her concept of social identity is based on Christine Korsgaard's (1996) concept of practical identity. Anderson's account has not received in-depth attention in the team reasoning literature. In my master thesis, I will examine whether it can solve the justification problem.

I argue that it cannot and identify two main issues: first, she cannot motivate why we should identify with humanity and why this identity should be prioritised over other social practical identities. Second, she lacks an adequate account of the nature of the independent agent and its relation to social identities and the individual self. Further, I show that Korsgaard's account of practical identity can solve the first issue, but only at the cost of sacrificing team reasoning.

Abstract (Deutsch)

Team Reasoning stellt eine bedeutende Lösung in der Ökonomie dar, um Kooperation in Gefangenendilemmas zu erklären: Akteure kooperieren, insofern sie als Mitglieder eines Teams denken. Aus normativer Perspektive wirft dies die Frage auf, warum die Akteurin sich für *Team Reasoning* anstelle der individuellen Nutzenmaximierung entscheiden sollte (Rechtfertigungsproblem).

Team Reasoning Theorien haben entweder das normative Problem ignoriert (Bacharach 2006) oder erfolglos versucht, es im Rahmen der instrumentellen Rationalität zu lösen (Sugden 2015). Das Problem ist daher nach wie vor ungelöst (vgl. Gold und Colman 2020).

Elizabeth Anderson (2001) schlägt eine Interpretation von *Team Reasoning* vor, bei der unsere Handlungsgründe von unseren sozialen Identitäten abhängen. Ihr Konzept der sozialen Identität basiert auf Christine Korsgaards (1996) Konzept der praktischen Identität. Andersons Ansatz hat in der Literatur zu *Team Reasoning* bislang keine vertiefte Aufmerksamkeit erhalten. In meiner Masterarbeit werde ich untersuchen, ob Anderson das Rechtfertigungsproblem lösen kann.

Ich argumentiere, dass dies nicht der Fall ist, und identifiziere zwei wesentliche Probleme: Erstens kann sie nicht überzeugend darlegen, warum wir uns mit der Menschheit identifizieren sollten und warum diese Identität Vorrang vor anderen sozialen praktischen Identitäten haben sollte. Zweitens fehlt ihr eine adäquate Erklärung für die Natur des unabhängigen Akteurs und dessen Verhältnis zu sozialen Identitäten und dem individuellen Selbst. Darüber hinaus zeige ich, dass Korsgaards Ansatz der praktischen Identität das erste Problem lösen kann, jedoch nur auf Kosten von *Team Reasoning*.

Acknowledgements

I would like to thank my supervisor, Professor Hans Bernhard Schmid, for his continuous support, stimulating discussions, and valuable time throughout the course of this thesis. His insightful guidance has been crucial to the development of my ideas, and I am particularly grateful for his encouragement to apply for the Como Summer School, an experience that significantly enriched my academic journey.

I am also indebted to my good friend, Damian Diethelm, for his numerous incisive comments on earlier drafts of this work. His flexibility, commitment, and dedication to providing constructive feedback were indispensable in refining my arguments and improving the overall quality of the thesis.

Furthermore, I would like to extend my sincere thanks to the following individuals: Dr Felix Pinkert, whose helpful comments during the Thesis Colloquium provided me with valuable perspectives, and whose encouragement further motivated me to pursue my project. I am also grateful to Professor Natalie Gold, whose insightful feedback on my project at the Como Summer School in Philosophy of Economics deepened my understanding. Lastly, I would like to thank my brother, Dominic Schinner, for his careful proofreading of the manuscript.