Children’s Developing Use of Reasons for Joint Decisions

A thesis submitted to the University of Manchester for the degree of Doctor of Philosophy in the Faculty of Biology, Medicine and Health

2018

Andreas Domberg

School of Health Sciences
# Table of Contents

- List of Figures and Tables   5
- Abstract         7
- Declaration      9
- Copyright Statement 11
- Rationale for alternative format   13
- Acknowledgements 15

**Chapter One** 17
*General Introduction*

1 A Social Turn in Reasoning..............................................................17
2 Social theories of reasoning..........................................................20
   2.1 The Argumentative Theory of Reasoning....................................21
   2.2 The Intention Alignment Model................................................22
   2.3 The two accounts in comparison............................................22
   2.4 Evidence for social theories of reasoning..................................27
3 Development of evaluative and productive reasoning skills....................29
   3.1 The relationship between individual and social reasoning in development.. 30
   3.2 The development of selective trust..........................................32
   3.3 Children’s non-compliance, justification and reasoning...................38
   3.4 Summary of the developmental literature..................................43
4 Focus of the dissertation..............................................................43

**Chapter Two** 45
*Children prefer to reason with partners who respond to good reasons*

1 Introduction....................................................................................45
2 Method............................................................................................47
   2.1 Participants..............................................................................47
   2.2 Materials..................................................................................48
   2.3 Procedure................................................................................49
   2.4 Coding......................................................................................52
3 Results............................................................................................52
4 Discussion......................................................................................53

**Chapter Three** 57
*Children’s Reasoning with Peers in Cooperative and Competitive Contexts*

1 Introduction....................................................................................57
Chapter Four
Cooperative Reasoning in the Context of Group Competition

Chapter Five
Summary and general discussion

Bibliography
Appendix A: Narration Materials for Chapter Two
Appendix B: Additional material pictures and plots for Chapter Four

Word count: 27,792
List of Figures and Tables

Figure 2.1: Cooperative warm-up game.............................................................. 48
Figure 2.2: Reasonable partner choices and their explanations.......................... 53
Figure 3.1: Warm-up and testing materials.......................................................... 60
Figure 3.2: Animal items to be placed in the zoo............................................... 61
Figure 3.3: Quantity of children’s arguments....................................................... 64
Figure 3.4: Reason types and self-servingness..................................................... 66
Figure 3.5: The setup and materials of Study 2................................................... 69
Figure 3.6: Trained argument reproduction......................................................... 72
Figure 4.1: Token selection before and during test phase.................................... 80
Figure 4.2: Picnic, zoo and doll-house tokens..................................................... 81
Figure 4.3: Number of overall arguments produced per trial............................. 87
Figure B.1: Background picture of the zoo material............................................ 117
Figure B.2: Background picture of the doll-house material................................. 117
Figure B.3: Number of interactive arguments produced per trial....................... 118
Figure B.4: Trial durations across conditions and age groups............................ 118

Table 4.1: Example arguments................................................................................. 84
Abstract

A rapidly growing literature characterises reasoning as a social phenomenon. On the one hand, Mercier and Sperber (2011) argue that the main function of reasoning is for speakers to persuade group members, e.g., of a given belief or decision. On the other hand, Tomasello (2014) emphasises that reasoning is also critical in cooperative contexts, when groups need to find a good and justified solution benefiting all group members, rather than members competing for individual argumentative success. The literature on children’s reasoning, however, focuses on their use of justifications in either conflictual/competitive or cooperative contexts without a systematic comparison of how young children’s arguments change across contexts. Also, investigating how children judge arguments and use these judgements as informative input to further tasks would deepen our knowledge about their reasoning abilities. In my dissertation, I therefore investigated with whom children prefer to reason and how they reason with same-age peers as partners in cooperative and competitive contexts.

In the first study, I investigated whether children who can choose between cooperative partners penalise candidates that accept poor reasons. 5- and 7-year-old children were presented with two candidates for solving a cooperative problem. In the experimental condition, one acceded to a good reason, the other to a poor reason. In the control condition, each agreed to a different good reason. Crucially, in both conditions, both arrived at a wrong conclusion. Results suggest that 7-year-olds, and 5-year-olds to a lesser degree, preferred the partner who followed the good reason in the experimental condition, but they showed no preference in the control condition. Thus, young children prefer partners who respond to good reasons over partners who comply more easily. This provides evidence that children evaluate reasons even in absence of one objectively correct informant, and use this evaluation to judge cooperation partners.

In the second study, I used a more interactive paradigm and investigated how 5- and 7-year-old children reason with peer partners in cooperative and competitive contexts. Peer dyads were asked to place animals in various places in a zoo, which included unusual items such as an alarm clock. One child received training, playing the game with the experimenter before playing with his/her naïve peer and learned a set of critical arguments (e.g., “Bears hibernate so they need the alarm clock to wake up when winter is over”). When the trained child played the game with a naïve peer, reproducing these arguments would help both of them in the cooperative condition but worsen the trained child’s chances of winning in the competitive condition. The results suggest that 7-year-olds, but not 5-year-olds withheld the trained arguments more often in the competitive condition than in the cooperative condition. Thus, this study suggests that children’s reasoning is less biased or less “inhibited” in cooperative settings in which a joint decision benefits both parties.

In the third study, I investigated whether combining cooperative and competitive contexts (namely, cooperating as a team against competitor) would facilitate children’s reasoning further. Similar to the second study, 5- and 7-year-olds were asked to cooperate in a task (e.g., decorating a zoo). In the main condition, children were competing against another group, whereas in the control condition, they were not. The prediction was that children would show more productive reasoning and richer discussions in the main condition. However, this prediction did not bear out. Out-group competition did not trigger richer reasoning between partners. I discuss this result in addition to a set of theoretical remarks.

As a whole, this set of studies suggests that starting in their sixth year, children start to identify and prefer to reason with people who submit to reason. They display less biased reasoning in cooperative contexts, in which the goal is to reach good joint decisions that benefit all group members. And out-group competition itself does not have a specific positive influence on children’s cooperative reasoning.

The University of Manchester
PhD Psychology
Andreas Domberg
May 2018
Declaration

No portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.
Copyright Statement

i. The author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the “Copyright”) and s/he has given The University of Manchester certain rights to use such Copyright, including for administrative purposes.

ii. Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made only in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from time to time. This page must form part of any such copies made.

iii. The ownership of certain Copyright, patents, designs, trade marks and other intellectual property (the “Intellectual Property”) and any reproductions of copyright works in the thesis, for example graphs and tables (“Reproductions”), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.

iv. Further information on the conditions under which disclosure, publication and commercialisation of this thesis, the Copyright and any Intellectual Property University IP Policy (see http://documents.manchester.ac.uk/display.aspx?DocID=24420, in any relevant Thesis restriction declarations deposited in the University Library, The University Library’s regulations (see http://www.library.manchester.ac.uk/about/regulations/) and in The University’s policy on Presentation of Theses.
Rationale for alternative format
And an account of how the thesis format has been constructed

This thesis is submitted in the Alternative Format, which incorporates sections that are in a format suitable for submission for publication in a peer-reviewed journal. The formatting of papers has been altered to make them consistent throughout the thesis but in all other respects they are as they were published or submitted for publication.

This alternative format thesis conforms to the same standards expected for a standard thesis. As in the traditional format, the introduction reviews previous research (Chapter One) and the summary draws together the various outcomes of the work (Chapter Five). In addition, three empirical papers are also presented. As of the date of this thesis submission, one paper has been published (Chapter Three), one paper is under review (Chapter Two), and one paper is in preparation for submission (Chapter Four).

My own contribution to these investigations was as follows. I researched the relevant literature for each paper; designed the experimental methodology including procedure and materials (apparatus, printed material, electronic material); conducted piloting and data collection; transcribed and coded recordings, prepared and analysed data statistically; wrote, submitted and revised the papers. In all of these matters, I received advice and supervision from my two co-authors and supervisors, Bahar Köymen and Michael Tomasello. In carrying out these experiments, I received support from research assistants and student assistants at the Max Planck Institute for Evolutionary Anthropology, Leipzig (Germany).
Acknowledgements

This dissertation would not have been possible, were it not for the inspiring, passionate and wonderful people that have accompanied me over the past couple of years. I am especially grateful to Mike Tomasello and Bahar Köymen for first giving me a shot at a challenging Masters thesis, and entrusting me with the unique opportunity to tackle this exciting dissertation project. I have immensely profited from Mike’s trust, ideas and big picture perspective, and Bahar’s extraordinary amount of guidance, motivation and expertise that have bridged 1,000 kilometres as if they were a quick walk down the MPI hall. I am also grateful to Elena Lieven for having me at the Manchester Child Study Centre and making the intricacies of doing a PhD seem like another walk up the MPI hall.

I would like to thank the MPI child lab for world-class support, seemingly effortless organisational magic, and patience. Especially Elena Rossi, Isabelle de Gaillande-Mustoe, Jana Jurkat, Susanne Mauritz and Christian Biermann have made these studies feasible. Also, Petra Jahn, Roger Mundry, Sylvio Tüpke, Ronny Barr, Henriette Zeidler, Cristina Zickert, Susann Hannemann, Elvira Möhring and Elvira Portner have been extraordinarily engaged and vital to this project in so many important ways.

I have learned a great deal in discussions at the Social Cognition Meeting, especially from Maria Mammen, Ulrike Kachel, Gregor Kachel, Susanne Hardecker, Marina Proft, Avi Benozio, Cathal O’Madagain, Alex Sánchez, Sebastian Grüneisen, Katharina Helming, Richard Moore, Barbora Siposova, Hagen Knöfe, Rebecca Koomen, Jan Engelmann, Marie Schäfer, Sonja Ebel and Johanna Eckert. These scientists have given me essential insights about kids, the universe and everything.

I am also indebted to Bahar, Lea and Ulli for reading all or portions of earlier versions of this thesis and making it into a much more comprehensible and better text.

A special thank-you goes to my wonderful friends Maria, Jana, Sonja, Lea, Susann, Gregor and Ulli and HHK for their cheerfulness, support and perspective over this long period. Finally, I am deeply indebted to my family for more enduring and generous support and love than you could ask for.

This research was funded by the Max Planck Institute for Evolutionary Anthropology, Leipzig (Germany).
Chapter One
General Introduction

1 A Social Turn in Reasoning

Reasoning is a quite general phenomenon to study. And it is an important one to study, too, given the role it has in the many decisions that we make in everyday life. Also, reasoning counts among the things that are commonly taken to make humans special—a quality that it shares with language. The development of reasoning is a similarly important phenomenon to study. For one, if we want to understand the functioning of the fully developed capacity and how it is used, it is worth focusing on how children build it up. Also, given the importance of good reasoning, it makes sense to target interventions at the development of this skill in childhood and education. Incidentally, reasoning shares this property with language as well.

A theoretical and terminological distinction is necessary: The term “reasoning” is commonly used to refer to one of two separate meanings. One is to do with individual thinking, with making up our mind. The other usage is about reasoning with one another, exchanging reasons in communication. In the second half of the 20th century, the psychology of reasoning has largely focused on the individual perspective. It has approached reasoning as the capacity of going beyond perception in information processing. Given a set of observations, it involves, for instance, the induction of general principles, deduction of necessary consequences, or inference to the best explanation of a set of observations. Wider construals also include probabilistic inferences and predictions. As a consequence, the dominant framework in the psychology of reasoning has relied on tasks in logic or probability. Participants’ performance in these tasks has been compared to the objective benchmarks of logic and probability calculus, and the func-
tioning of human reasoning has been characterised based on the observed patterns of failure (Manktelow, 2012). Incidentally, an important approach to the study of linguistic performance has also been relying a great deal on where it fails. The collection and analysis of speech errors and slips of tongue, as well as investigating the comprehension of garden-path sentences have been important early tools in research in language use (Fromkin, 1968).

However, the point at which the similarity between reasoning and language breaks down is the following. As an evolved capability, language is well-adapted to its domain. Speech errors and garden-path sentences are so rare as to be meaningless for the success of day-to-day communication. In comparison, reasoning in the individual shows a poor fit to quite general problems. The study of decision-making, thinking and reasoning has identified countless biases and errors that violate probabilistic and logical norms. Mercier and Sperber (2011) suggest that

“[…] in fact, humans reason rather poorly, failing at simple logical tasks […], committing egregious mistakes in probabilistic reasoning […], and being subject to sundry irrational biases in decision making […].” (p. 58)

Importantly, these problems do not seem to come up only under boundary conditions, such as garden-path sentences, which are very rare. An example illustrating the generality of the problem is a study by Bonnefon and Villejoubert (2007). Participants in this study failed to apply the quite general modus tollens argument form:

If \( p \), then \( q \).
It is not the case that \( q \).
(conclusion) Therefore, not \( p \).

In this type of argument, the first (major) premise and second (minor) premise together deductively lead to the conclusion. So given that \( p \) always leads to \( q \), if \( q \) is not the case, \( p \) cannot be either. Participants were presented vignettes in this form with various

---

1 A famous example of a garden-path sentence is “The horse raced past the barn fell.” The garden-path aspect means that a sentence structure and the roles of individual words in it can be ambiguous. The last word disambiguates the structure and triggers a reanalysis of the sentence as a whole. In some sense, thus, the analysis up to that point was misled by a “bias” towards early syntactic closure.
semantic contents filled in and largely expressed that they considered it likely that \( p \) follows. This conclusion is not warranted following the rules of logic. Considering the generality and ubiquity of argument forms of this kind, it is unusual that a cognitive ability should be found to be deficient at this scale. One explanation for this might be that for some reason, the regularity of good fit between evolved capacity and domain does not apply straightforwardly to reasoning. Another, more realistic explanation is that understanding and studying reasoning as an individual activity misses its domain of adaptation. Logical or probabilistic accuracy may not be its goal.

Recent inquiry into reasoning performs a radical shift in perspective towards the other usage of the term “reasoning”, namely, “reasoning with each other”. Theorists acknowledge that social life is too important a shaper of thought to study reasoning in the individual (Mercier & Sperber, 2011; Norman, 2016; Tomasello, 2010, 2014). Proponents of a social turn aim to account for this fact by suggesting that reasoning is closely tied to communication.

Flexible communication is highly beneficial for species living in groups, as it allows similarly flexible cooperation. This, in turn, empowers cooperators to better meet the challenges presented by life. It also gives group members access to rich second-hand information, allows them to align their goals and beliefs, and facilitates teaching as a means to convey cultural knowledge across generations. But it also follows that contradicting information, incompatible beliefs and motivations may be present at any time, leading to disagreement. In response to that, social views on reasoning share the idea that the main function of reasoning is to preserve the benefits of group life by resolving these disagreements (Mercier & Sperber, 2011, 2017; Norman, 2016). When individuals disagree about their past or future course of action, reasoning means that each points out reasons that motivate or justify their own position. These reasons can be circumstances in the present situation, world knowledge, or shared values of what group members take to be true and authoritative.

This way, reasons can have a stabilising effect. For one, they exert their force on hearers to some extent independently of speakers and their dominance or reputation.
Typically, reasons do not lose this force in absence of the speaker, unlike direct requests or threats, for instance. And secondly, reasons are public. Once made explicit, parties to an argument can engage and scrutinize each other’s reasons and figure out which one is superior and which one is flawed. Consequently, potentially harmful stalemates can be avoided. Social theories of reasoning propose that this reason-based regulation of social interaction is indeed a more coherent phenomenon to study under the umbrella of reasoning than logic puzzles. If we think of reasoning as dealing with premises and conclusions, the direction in which individuals move between them is reversed in the social account: Whereas traditionally, conclusions would be computed from given information, now, individuals would start with a conclusion in hand and identify suitable premises.

With an updated picture of what reasoning is for, the importance of the developmental approach translates straightforwardly into the social perspective. To better understand how the capacity to reason with others is made up and how it can possibly be supported, it is worth investigating how children’s reasoning skills develop. This is what the present thesis is about.

2 Social theories of reasoning

Two of the foremost social theories of reasoning currently being discussed are the Argumentative Theory of Reasoning formulated by Hugo Mercier and Dan Sperber (2011, 2017) and the Intention Alignment Model by Andy Norman (2016), which I will abbreviate as ATR and IAM. Both posit that the function of reasoning is to overcome disagreements between individuals. This is a clear departure from traditional views on reasoning. Recall that so far, the psychology of reasoning focused on individuals and the inferences, predictions and decisions they make, given certain kinds of information. Below, I outline both of the social theories in turn, how they compare, and how they explain some empirical evidence of how people reason. Notwithstanding some interesting differences between both theories and their behavioural predictions, the empirical part of this thesis focuses on their shared view that reasoning is best understood as social. I raise
ideas for specifying the social view more precisely experimentally in the final discussion (Chapter 5).

2.1 The Argumentative Theory of Reasoning

The starting point for the ATR is the emergence of flexible, informative communication among humans organised in groups. In everyday life, humans relied more and more on second-hand information as a rich source of knowledge and necessity for mutually beneficial cooperation. However, reliance on second-hand information also introduces reliance on the source. In this situation, Sperber et al. (2010) argue, individuals have an incentive to deceive others. That is, a communicator is incentivised to give information to others not only when this is beneficial to all, but also when it only benefits themselves, and is even disadvantageous for others. As a result, the incentive on the hearer side to remain open to communication diminishes, since hearers would reasonably expect less and less benefit, or even damage if they trust speakers. This threatens to lead to a breakdown of communication altogether. Against this background, Sperber et al. (2010) propose the emergence of Epistemic Vigilance, a set of behaviours that hearers exercise towards speakers. This means on the one hand that hearers calibrate the trust they assign to what they are being told to the speaker’s credibility. On the other hand, they judge how plausible the communication is both internally and with regard to their world knowledge. The upshot of this is an incentive for speakers to come up with reasons demonstrating that they can be trusted, that ignoring them would be unreasonable, and thus to overcome their interlocutors’ Epistemic Vigilance.

In the ATR, the function of reasoning is to help speakers overcome hearers’ Epistemic Vigilance. While communicative skill certainly still plays a role for speakers’ success, the most decisive factor is how compelling available reasons are that they can draw from. The immediate motivation for individuals to reason under this view is to justify their views, plans and past actions. Speakers intend to show how they are reasonable, i.e., supported by sound reasons and in line with shared norms. On an aggregate level, this motivation overall mitigates the stability problem for communication. Deceptive signals
backed by poor evidence end up being filtered from communication by vigilant hearers, and speakers’ success is bounded in part by the evidence they can present.

2.2 The Intention Alignment Model

Norman’s (2016) point of departure for the IAM is similarly the emergence of informative, cooperative communication in groups of early humans (see Tomasello, 2010, 2014). Under this model, disagreement or critical attitudes towards communicative acts do not arise primarily as a response to the perils of deception. Rather, individuals engaged in cooperative activities may come to different conclusions about their further course of action. They may have perceived different evidence or have diverging world knowledge, so their perceptions of the situation are misaligned.

The function of reasoning for a cooperative group (e.g., a band of hunters) is to ensure that in a situation in which different individuals prefer different strategies, the best justified one succeeds. For each individual, then, the function of reasoning is to justify their own strategy choice for the cooperative enterprise. To this end, they base it on the best available evidence, making transparent to peers exactly why it is the best course of action. Again, skill in solving this communicative task helps, but the quality of the evidence for an individual’s conclusions is key. It systematically influences their leverage in discussing the cooperative activity, in which all are interdependent (Tomasello, Melis, Tennie, Wyman, & Herrmann, 2012). The adaptive value of this is that groups maintain alignment around publicly accessible evidence, which is often a more reliable indicator of the optimal solution than individual beliefs. More importantly, groups end up united over their joint endeavour (Norman, 2016). The rationale is that in case the evidence is misleading—prediction is never perfect—, being misled together is still better than a group breaking up over an irremediable controversy.

2.3 The two accounts in comparison

The ATR and IAM share key assumptions about the nature of reasoning, especially how reasoning differs from the traditional, individual view. Their authors point out rich
Social theories of reasoning

evidence, which is complemented by empirical findings in the present thesis. Both theories characterise reasoning in terms of pragmatics, of interacting with communication partners, and both depart from individual accounts with their focus on semantic, logical, and probabilistic thought. In both theories, reasoners start from a conclusion that they already hold, then evaluate evidence as to how well it supports their conclusion, and try to make this support transparent to their interlocutors. Also, in both theories, the evidence supporting reasoners’ conclusions plays an important role, because it co-determines their success in the dispute. However, there are two points in which the two approaches take different perspectives.

The adaptive goal of reasoning. A first point of difference between the two theories is the environmental challenge to which each is taken to be a response. For the ATR, it is hearers’ Epistemic Vigilance, which in turn is motivated by the incentive to deceive arising from the importance of flexible communication. For the IAM, it is the risk that a group may break down if members do not resolve important disagreements. In the ATR, reasoning first emerged to convince distrustful peers, but its use extends to situations in which individuals have cooperative goals; whereas in the IAM, it emerged in the social infrastructure of collaboration, and its mind-changing force has been co-opted for persuasive motives. As a consequence, the theories differ in what motivates reasoners to change their mind when facing superior reasons. In the ATR, the argument ends when I have successfully demonstrated to you, and ideally to bystanders, how reasons support my conclusion, and how my reasons and conclusions together are superior to yours. Your motivation for complying is to not appear unreasonable to the community, thus balancing a possible defeat in the matter of the discussion against the danger of losing social capital. After all, unreasonable individuals are not attractive partners for cooperation or good sources of information (the experiment in Chapter 2 bears on this). In the IAM, in turn, the convincing occurs because the reasons I produce better show how my conclusion helps attain our shared goals. Your own investment in our shared goal motivates you to comply without it feeling or appearing to others like defeat.
The source of reasons’ force of conviction. The second point of difference between the two theories can to some extent be seen as an extension or spelling-out of the IAM. It concerns the question of objectivity and what underlies the ability of reasons to change hearers’ minds. In both views, confirmation bias is not a failure of reasoning, but part of its proper functioning. It is important to make a distinction between separate phenomena that are commonly subsumed under ‘confirmation bias’ (Klayman & Ha, 1987; Kunda, 1990; Nickerson, 1998), namely its motivated and unmotivated form. The motivated bias is what is usually identified as case-building, the rational tendency to justify and defend own beliefs and actions. It is what one would expect from lawyers and debaters, for instance. The unmotivated confirmation bias is not in fact seen by Mercier and Sperber as a bias in a relevant sense. In line with Klayman & Ha (1987), they characterise it as a positive test strategy: a tendency to test a given belief by looking for instances of fulfilment of its predictions. Under many, although not all realistic conditions, it is an effective heuristic.

Having made the motivated–unmotivated distinction, Mercier and Sperber focus on motivated confirmation bias. In the ATR, it is a key feature of our faculty of argumentative reasoning. Reasoners are, in this view, well advised to reason with others in a self-serving way. This way, bias in reasoning helps speakers overcome hearers’ Epistemic Vigilance, so both balance each other and stabilise communication. Accordingly, the ATR makes, among others, a quite central prediction about confirmation bias:

“[...] that it should occur only in the production of arguments. The rationale for a confirmation bias in the production of arguments to support a given claim does not extend to the evaluation of arguments by an audience that is just aiming to be well informed.”

(Mercier & Sperber, 2011, p. 64)

In contrasting the production of reasons with their evaluation, Mercier (2016, p. 691) similarly suggests that “[...] reasoners are biased and lazy when they produce arguments, but objective and demanding when they evaluate others’ arguments.” A toy example should show how these assumptions can play out:
Little Alex expresses that he wants a puppy for Christmas. Mum rejects this idea and gives two reasons: (1) that Alex would probably lose interest to do the necessary work after four weeks; and (2) that a puppy would be too expensive to have.

Recall that Alex should evaluate these reasons in an objective manner as “ [...] an audience that is just aiming to be well informed”. However, what Alex thinks upon hearing this likely differs from what a disinterested observer of this conversation would think. From a logical perspective, both reasons are conclusions that rest on certain premises: (1a) if one loses interest it a puppy quickly, it should not be bought; (1b) Alex will lose interest quickly. (2a) If the costs of something exceed a relevant threshold, it should not be bought; (2b) the costs of a puppy exceed a relevant threshold. The premises (1a) and (2a) are also called major premises. While these major premises that underlie the reasons given may not be in doubt, Alex likely contests (1b) and (2b), also called minor premises. The present two minor premises are subject to personal evaluation and further premises that are so far implicit. Premise (1b), e.g., is probably contentious: Alex should have no motivation to critically introspect and reflect on the possibility that he might lose interest in a puppy. That is certainly more obvious to his mother. Regarding (2b), Alex should have no motivation to inquire as critically as his mother how many expenses come with a puppy (the dog’s price, food, bed or blankets, dog lead, collar, medicine, toys, annual vaccination, tax, insurance, or a dog sitter once in a while, to give a non-exhaustive overview). This premise, too, is likely contentious.

The point is that the evaluation of reasons is objective only to the extent to which reasoners share and commit to underlying premises. When underlying premises are shared (such as the rules working as major premises above), their role in reasoning is shared as well and thus in some sense intersubjective or objective. For instance, if the costs of a puppy exceed some threshold, it is too expensive and cannot be bought. This premise is hard to argue with. But other premises underlying the final conclusion may differ between speaker and hearer and lead to disagreement about the validity of the whole reason. And in fact, even the above major premises are not safe from being dis-
puted: One can buy items in absence of money, if one drops the shared commitment to the deeper tacit belief that a loan is not an option for a puppy.

If objectivity in evaluating reasons rests upon shared beliefs and commitments (see Tomasello, 2014, 2016 for their role in cooperation and normativity), it is only as stable as reasoners’ motivation to cooperate and honour them. This applies to commitments in the present situation, in the longer-term perspective of personal interaction, and with respect to norms and the group. So the central decision in reasoning-as-evaluating-reasons may as well be characterised as being about how to proceed in the face of incompatible beliefs. Reasoners can (a) resolve contradictions pointed out to them by accepting the reasons offered and giving up or modifying their own beliefs and intentions. Or they (b) identify counterarguments that defeat these reasons. This step is recursive: several counterarguments from all sides may be brought into the discussion. At a later step this can lead either to (a), i.e., admissions by one side, or to the *ultima ratio*: (c) Reasoners can drop an underlying joint commitment on which the incompatible reason presented to them rests.

Both endpoints, (a) and (c), come with changes that persist beyond the exchange of reasons, and reasoning means pondering the desirability of these changes: (a) Either reasoners have to give up their position. This can be perfectly uncostly, such as acknowledging a piece of information presented by a collaborator for a shared goal. It can come with a more noticeable cost, as when the information presented disproves a prediction that the reasoner has publicly committed to. Or it could have a much greater cost, as in acknowledging the evidence of having lost an election, or dispensing with a puppy. Alternatively, if reasoners cannot adduce counterarguments, but decide to stand their ground, they face the cost of (c), i.e., dropping the commitment that underlies a reason that clashes with their standpoint. For instance, they can negate the legitimacy of an election and the institutions and people that organised it; dismiss critical journalism by calling it “fake news”; or dismiss the assumption that a loan is not an option for a puppy. This affects reasoners’ relationship to their immediate collaborators and the institutions of their group. Since joint commitments are a basis for cooperation, i.e., for the possibil-
ity that group members can exert coordinative influence on others, this can mean a considerable cost.

While the ATR does not spell out these contingencies, the IAM captures this dilemma in part, by relating reasoning to group cohesion. It suggests that the function of reasoning on an ultimate level is to maintain a shared outlook for cooperative groups. Group members who are not successful giving arguments or counterarguments but still choose to stand their ground are bad for the group’s functioning. On the basis that early human group members were strictly interdependent (Tomasello et al., 2012), the cost of cancelling a commitment would exert a strong force towards group unity and compliance, according to the IAM. However, modern societies are characterised by many overlapping subgroups, each of which can be aligned around a specific topic—sports teams, politics, values, attitudes to climate change, and so on. In this context, giving up commitments to a group and membership in it is not an absolute question, where renegades are left to their own devices with little hope of survival. For the objectivity of reasons, it means that it hinges on group cohesion.

Regardless whether reasoning emerged as a safeguard against deception, as an arbiter between different paths to collaborative success, or as balancing immediate preferences against group belonging, social theories of reasoning predict that it is part and parcel of communicative development. This is what the three studies in this thesis are concerned with, building on the empirical literature on social reasoning.

2.4 Evidence for social theories of reasoning

This section aims at giving an idea of how a social perspective makes sense of seemingly flawed reasoning shown in experiments on individual reasoning (see Mercier & Sperber, 2011, and Mercier, 2016 for evidence favouring the ATR or a social view on reasoning in general; Nickerson, 1998 and Kunda, 1990 for extensive explorations of normatively deviant individual reasoning). The basic idea is that individual reasoning fails so often because critical evaluation of reasons can only be done by a partner.
A study reported by Evans and Wason (1976) presented participants with a logic puzzle, the widely used Wason card selection task (Wason, 1966), which ca. 10% of adult participants had previously been shown to solve correctly. In this study, participants were additionally presented with a solution to the puzzle and asked to justify it, not knowing in some conditions that the solution was false. Participants overall justified the solutions readily with high confidence, irrespective of their accuracy. Both Evans and Wason (1976) and Mercier and Sperber (2011) explain this by suggesting that participants did not engage in logical reasoning at all. Rather, they did what reasoning does best, justify their given belief, regardless how they got it.

Moshman and Geil (1998) administered the same task to groups of five to six participants and compared their performance to that of individuals. They observed that while individuals, as expected, had a low success rate of 9%, groups solved the puzzle 75% of the time. It suggests that those group members that individually would have responded correctly convinced their peers successfully, and even that does not explain the whole increase in correct responses. Proponents of social accounts of reasoning trace this “truth wins” effect back to the idea that a better fit between evidence and speakers’ goals allows for more compelling reasons and more success in discussions. Participants with a solution in mind that is easy to justify have better resources to convince their peers. This effect has also been observed with dyads of second-graders made up of one child who previously solved a Piagetian conservation task, the other not. In one such task, for instance, children essentially respond to the question if a clay ball, if broken down into many, conserves weight. Here, children who subscribed to this view were more often able to convince their peers that weight was conserved than the other way around (Miller & Brownell, 1975).

In a recent experiment by Trouche, Johansson, Hall, and Mercier (2015), participants answered a series of logical reasoning problems. These were, among others, of the following form:

2 In the original abstract form of this task, subjects are presented with four cards, e.g., A, B, 4 and 5, and with the rule that every card with a vowel on one side has an even number on its flipside. They are asked to name only those cards that need to be turned over to demonstrate the rule holds.
“The [...] fruit and vegetable shop carries, among other products, apples. None of the apples are organic. What can you say for sure about whether fruits are organic in this shop?” (Trouche et al., 2015 p. 4)

Participants answered in a multiple choice format and gave written justifications for their responses. In a subsequent step, they had to accept or reject the responses and explanations of, as they were told, other participants. At this point, experimenters tricked participants into evaluating one of their own prior responses, believing that it is someone else’s. They saw the question, the response that they had given but were led to believe was not theirs, and an incompatible response that, it was indicated, had been theirs. Over half of those who had not discovered the manipulation rejected their own prior response and the reasons for it. Their critical stance towards what was marked as someone else’s response illustrated Epistemic Vigilance, which is always directed outward, not towards what reasoners (think they) hold as their own belief.

Two key ideas are reflected by the reasoning shown in these studies. Firstly, reasoners do not subject their personally held beliefs to critical scrutiny, whereas they do critically evaluate the beliefs of others. Secondly, the evidence available to a speaker determines how compelling reasons based on it are. This means that speakers have the power to convince and align others to their own view conditional on the evidential support they can adduce.

3 Development of evaluative and productive reasoning skills

How are developmental studies relevant for understanding reasoning? As children begin to engage in reasoning, they naturally make mistakes. Investigating what tasks they master at a given age and with what aspects they struggle allows an insight into how our social reasoning capacity is structured. Research in language development has demonstrated that looking at children’s patterns of language use and patterns of errors can be very informative about how language competence is structured more generally (Tomasello, 2003). Also, given that argumentative thinking and reasoning are key objectives
in education, a developmental perspective is essential to providing a theoretical foundation for practical applications.

In this respect it is worth pointing out that somebody’s reasoning is only as good as their reasons. As outlined in Section 2, for social accounts of reasoning, the strength of a given reason depends on the mutually accepted beliefs (facts) that the speaker can refer to. The immediate logical cogency of “don’t stick that pen into the fan, you know what happened last time” is not an achievement on the speaker’s part. Recalling and pointing out that past event, which we all agree was unpleasant, is the actual piece of work that a speaker can get right or wrong. Therefore, characterising what it is that children need to learn to become good social reasoners, it is identifying the beliefs they share with an interlocutor to which they can appeal. This has little to do with logic and much with memory or world knowledge, and knowledge of the addressee.

In the following, I outline how the present research connects to classical and current research in child development.

3.1 The relationship between individual and social reasoning in development

If we follow the social turn in reasoning, then we are looking at a different activity and skill set than early theorists like Piaget and Vygotsky had in mind when talking about reasoning as the goal of cognitive development. They focused on individual reasoning, but as outlined above, individual and social reasoning are different. What is commonly subsumed under individual reasoning is in fact a large collection of domain-specific skills. Many of these have given rise to specialised research in (developmental) psychology, such as spatial, quantitative, verbal, or logical reasoning. Mercier and Sperber further elaborate this view in endorsing a massively modular cognitive architecture (for more detail, see Mercier & Sperber, 2017). The point is that the classical models aim to explain the acquisition of these individual reasoning skills and presuppose communication as one channel over which children receive input. Social theories of reasoning, in turn, suggest that communication itself is the very domain of reasoning. So the most interesting insights for the
present purposes of the classical models of cognitive development concern less their goals of development, but rather what they have to say about communication.

An important idea in this respect is Piaget’s hypothesis that social reasoning proceeds differently among peers than between child and adult, especially in moral judgement (Piaget, 1932). An adult’s authority determines to a large extent the course of an exchange, whereas peers are on an equal footing, able and motivated to try and change each other’s mind. It is in these peer contexts that children can perform reasoning in the sense of pondering different sides to a question. The fact that two can look at the same matter and both be justified to get to different conclusions becomes most salient with peers (see also Kuhn (1999) who puts this insight into a lifespan perspective). And it suggests that peer studies are an especially suitable method for eliciting children’s social reasoning.

Vygotsky emphasises how higher cognitive skills are inherently social, first encountered by the child as social behaviours, and then internalised. So what is initially the exchange of reasons in dialogue translates more and more over time into the child’s individual handling of reasons. Two insights emerge from this socio-cultural view. On the one hand, this anticipates the view espoused later by the social approaches to reasoning presented above, that reasoning serves the exchange of reasons. Granted that Vygotsky likely took social exchange to be rather the medium for how cultural truths are brought to the child’s “zone of proximal development” for assimilation into individual skills. But with every demonstration of why one should believe or do something the culturally sanctioned way, children also receive lessons in achieving communicative goals by changing perspectives, and in changing perspectives by pointing out reasons.

On the other hand, Vygotsky emphasised culture as the provider of developmental goals. In the discussion of objectivity above, the proposal was already made that the basis and source for beliefs warranting some decisions but not others is the set of shared commitments in a group. This involves a sense of cultural relativity: What is objectively true depends on our group’s (including our culture’s) shared commitments. To even begin to interact effectively with the environment, of course, there needs to be considerable overlap with the physical and statistical structure of reality. A culture that passes on cultural
knowledge in ignorance of the rules of addition will be at a disadvantage. But most cultures do rather well without traditions that help solve Wason’s selection task correctly.

Taken together, the key ideas from Piaget and Vygotsky to which the present research connects are, firstly, the importance of peer discussion as opposed to parental guidance. Interlocutors must be on equal terms for there to be even the possibility of reasoning in the sense of changing minds. Secondly, child development is also an entry into a culture. Cultural knowledge, including knowledge of norms, beliefs, but also the physical-statistical structure of the environment, need to be acquired to be available as reasons in discussions. Thirdly, every time an interlocutor makes an argument or demonstration, this can be an opportunity for the child to watch and practice the ins and outs of social reasoning.

In recent research in child development, two lines of inquiry have produced especially relevant bodies of literature, which I review in the following. The first covers studies on selective trust in young children. It addresses the question of when and how children learn to judge the quality of speakers, information and reasons. The second covers studies on how children articulate opposition and coordinate desires early in development and gradually produce more and more elaborate reasons in both oppositional and cooperative interactions over time.

3.2 The development of selective trust

Children actively explore their environment and seek information and explanation (Harris, 2007). However, a large part of what is there to learn is not directly accessible, such as the name of a bird or tree, what is important to Santa Claus, or the answer to the question: “Why?” Second-hand information is vitally important in many instances, and children need to learn to judge who and what is credible. This literature basically asks what information children take into consideration to decide when they trust an informant.

Trust based on speaker attributes. In one line of research, studies on selective trust have had children judge informants with regard to relevant personal attributes such
Development of evaluative and productive reasoning skills

as prior accuracy or benevolence. Many of these experiments overlap broadly in their procedures: During a familiarisation phase, typically realised with videos or puppets, children see two informants name a number of familiar items, with one of them being correct and the other incorrect all or most of the time. In a subsequent explicit judgement trial, children are asked which informant has been right or wrong. This is followed by a succession of test trials in which children can demonstrate their selective trust by choosing between informants, upon being shown a novel object whose name is to be found out. In “ask” trials, the question that children answer is which informant they would like to ask about the object’s name. In “endorse” trials, each informant names the object using a novel word, and children are asked which name they think is right.

Using this paradigm, Koenig, Clément and Harris (2004) showed that 4-year-old children assign trust selectively. More often than a random selection would suggest, children preferred the previously accurate character over the inaccurate one as a source for new information. In a similar study by Clément, Koenig and Harris (2004), the information children could ask or endorse was the location of a hidden object (instead of the name of a novel object), the result being comparable. In addition, 4-year-olds in this latter study were shown to also ignore the previously correct informant when in the test trials he contradicted children’s first-hand knowledge.

Children also assign selective trust when inquiring procedural information, as shown by Koenig and Harris (2005). Here, in one experiment, after the familiarisation, children were asked to find out the “correct” use of a novel object. They again could select which informant to ask, and for the “endorse” trial, each demonstrated a way of using the object. Children aged four again assigned trust more often than chance would suggest to the previously correct informant. Also, in this study, even 3-year-olds assigned trust to the correct informant above chance level when the suboptimal informant, instead of giving false information, admitted ignorance. Birch, Vauthier and Bloom (2008) extended the paradigm by investigating whether 3- and 4-year-olds could use the principle of exclusion when learning verbal information: Upon familiarisation, and after the accurate and inaccurate informant both named a novel object as “wug”, if the experimenter asked
for the “dax”, children gave what the inaccurate informant had apparently misnamed. Thus, children concluded that the accurate informant named the novel object correctly, so that this object was excluded as a possible referent of the name “dax”. Pasquini, Corriveau, Koenig and Harris (2007) added the difficulty that in the familiarisation, the difference between informants was not absolute, but one was correct 75% of the time, the other 25%. Under these circumstances, only 4-year-olds, but not 3-year-olds selected the better informant more often than chance level.

Other studies have investigated how stable children’s preference for a more accurate informant is if other variables besides accuracy are manipulated. This is important insofar as in the above studies, all that varies across informants is accuracy, so even a very weak preference for it can produce an effect. Varying a second factor, however, provides a measure of the strength of children’s preference for accuracy. Jaswal and Neely (2006) presented 3- and 4-year-olds with a child and an adult as informants. In one experimental condition, the child informant named objects accurately and the adult inaccurately throughout the familiarisation. Children reliably preferred the child over the adult as informant. Corriveau and Harris (2009) offset informants’ accuracy with familiarity. Children aged three to five years saw two informants of whom one was a familiar teacher, and the other an unfamiliar adult. Before familiarisation, a majority of children preferred the familiar informant. During familiarisation, in one condition, the familiar adult was the inaccurate informant, and children’s preferences in this condition changed towards the unfamiliar accurate adult, slightly in 4-year-olds, and more strongly in 5-year-olds.

Apart from informants being accurate or inaccurate previously, the role of their benevolence or malevolence has been investigated. Mascaro and Sperber (2009) varied for each of two puppet informants whether they were introduced as being nice or mean. They then made statements about what was in a container, and 3-year-olds were asked which of both statements was true. After two trials in this procedure, children were to select which of the puppets should give them a present. Of those twelve children who preferred to get a present from the benevolent puppet, ten endorsed this puppet’s information in both trials.
Overall, children at the age of four years, under some circumstances already at three, show a preference for informants who have a history of being accurate. They take further informant attributes such as familiarity, adult status and benevolence into account to calibrate their trust. And in experiments that have pitted these attributes against accuracy, children focused on prior accuracy to inform their choices.

**Trust based on the quality of reasons.** In the above studies on selective trust, informants’ accuracy was established mostly by them naming familiar objects, so children could verify their accuracy directly, assign trust, and predict future informativeness. More recently, research has focused also on children’s responsiveness to how well informants justify their claims, judging which informant provides better reasons. Many of these studies have used circularity as the mark of the bad reason. Children typically compare this against a non-circular argument, e.g., a perceptual argument, so the respective informant appears as an eye-witness.

For instance, Corriveau and Kurkul (2014) presented 3- and 5-year-old children with pairs of informants who gave either circular or non-circular explanations for familiar phenomena. For example, given the question “Why does rain fall?”:

- **circular explanation:** “Sometimes it rains because it is wet and cloudy outside, and water falls from the sky. When water falls from the sky it is called rain and it gets us all wet.”

- **non-circular explanation:** “Sometimes it rains because there are clouds in the sky that are filled with water. When there is too much water in the clouds it falls to the ground and gets us all wet.”

When informants’ explanatory tendencies were thus established in the training phase, they each offered a new explanation for a novel object, and children were to respond which one was correct. Five- but not 3-year-olds showed a preference for the informant who had given the non-circular argument. When the explanations were shortened, both age groups tended to prefer this informant.

Similarly, Mercier, Bernard, and Clément (2014) found a preference in one task for information from non-circular over circular informants in 3- to 5-year-olds and adults. In
another task, subjects chose between an informant giving a circular argument and another giving no reasoning at all for their information. Four- and 5-year-olds in this study preferred the circular informant, while 3-year-olds showed no preference and adults preferred the informant giving no reason. The authors interpret the outcome in the first task as an indication that children around 3 years start to disprefer informants who give circular arguments. Based on the second task, they suggest that 4- and 5-year-olds interpret circularity as a cue to speaker dominance.

Castelain, Bernard, Van der Henst, and Mercier (2016) investigated children’s sensitivity to the quality of reasons in a Mayan sample and found that 4- to 6-year-olds’ preference for non-circular arguments trumps a speaker’s dominance status. Children in this study preferred to endorse a subordinate over a dominant speaker’s information concerning the whereabouts of a lost item when the subordinate speaker’s reasoning was non-circular. In a study by Castelain, Bernard, and Mercier (2018), children as young as 1;11-2;8 were shown a novel, hybrid item that, for instance, could be a bird, but also a fish. Upon their naming of the item, e.g. as bird, the experimenter used the opposite label for it, supported by a circular, non-circular, or no argument. For instance:

- **strong argument condition:** “It’s a fish, because I saw it swimming in the water.”
- **circular argument condition:** “It’s a fish, because I saw that is a fish.”
- **absence of argument condition:** “It’s a fish, hmm it’s a fish.”

While the experimenter was still in the room, the majority of children endorsed his label. After he left, however, they went back to the label they had used originally if the experimenter had given a circular argument more often than when he had used a non-circular one or none.

On a methodological note, circular reasoning has been frequently used as an example of bad reasoning. In adult discourse, circularity is normally diagnosed when an argument stretches over a succession of inference steps, at the end of which follows the point of departure, or no relevant addition. Indeed, speakers are the more likely to succumb to this fallacy, the more complex their reasoning is. On the hearer side, circularity
can be difficult to spot even for experienced hearers, as circularity is a global property of the argument as a whole. Consequently, to make the discovery of circularity tractable for children, the arguments used in experiments are generally of minimal length in terms of successive inference steps (see the above examples). This should be kept in mind when interpreting the results: These arguments can in fact be defective to hearers not just due to their global circularity, but also because they violate an expectation of non-redundant information on a linguistic level. Therefore, circular arguments of minimal length may tap into children’s linguistic competence rather than, or in addition to, their evaluation of the quality of a reason. Another point worth noting is that what was used as a circular argument in Castelain et al. (2018), namely saying in principle “ϕ, because I saw that ϕ” falls precisely into the category of strong eye-witness argument in, e.g., Mercier et al. (2014) and Koenig (2012, see below). To summarise this point, the quality of arguments can be hard to dissociate from linguistic quality when their structure needs to be kept minimal for experimental reasons.

Another way besides circularity to manipulate the quality of reasons is to manipulate the propositional attitude towards the information that speakers express. For instance, given the proposition “this is the right box”, speakers can express the attitude of desire (“I wish this was the right box,”) or certainty (“I know this is the right box,”) or questioning (“Is this the right box?”) among many other things. Choosing a specific propositional attitude for a statement can render it inappropriate as a reason for its context. In a study by Koenig (2012), informants offered reasons for their information about what might be in an opaque container. The propositional attitude into which they embedded a statement about the content of the container determined the quality of the reason. “Good” reasons had the following form:

teacher’s testimony: “My teacher told me that there’s a cookie in the box.”
looking: “Before, when I came in here, I looked and I saw a cookie in the box.”
informative container: “It’s a cookie jar. Cookie jars hold cookies.”

“Poor” reasons, in turn, involved these propositional attitudes:
**Chapter One**

**desire:** “I like cookies. I want there to be cookies in the box.”

**pretence:** “I like to pretend. I’m going to pretend that there’s a cookie in the bag.”

**guessing:** “I don’t know. I’m going to guess that there’s a cookie in the jar.”

The latter are three propositional attitudes that are incompatible with the function of justifying a claim about the container’s content. Three- to 5-year-olds, after hearing two informants present a good or poor argument, were asked which one “has the best way of thinking”. All age groups, with increasing certainty over age, preferred the giver of an appropriate reason more often than chance level.

**Summary of the trust literature.** The picture that results from the literature on selective trust is that by four years of age, children use various cues about their informants, such as accuracy, to assign trust. Accuracy receives more weight than the fact that informants might be adults, familiar, or benevolent, instead of children, unfamiliar, or malevolent. Also, children prefer information that is supported by non-circular arguments, although with the caveat that short circular arguments also resemble linguistic anomaly. Finally, children also prefer informants that express a propositional attitude such as knowing or concluding from direct experience in support of their information, instead of desire or guessing. All this suggests that preschoolers do care about and keep a record of their informants’ likely reliability, based on which they guide their social learning. Relevant to the present thesis, these studies show that children before the age of five make principled judgements about how well some claims are justified. In what follows, the focus is on their ability and motivation to justify their own claims.

### 3.3 Children’s non-compliance, justification and reasoning

Children’s first justifications emerge as an integral part of their language and communication skills. Early justifications appear in interaction with their parents, their most available interaction partners early in childhood. Parents make these interactions easy by scaffolding conversations, structuring situations and maintaining the topic. Later, with
siblings and peers outside home, conversation partners become more diverse and importantly share less common ground, raising the requirements of making justifications work.

**Interaction with parents.** Kuczynski, Kochanska, Radke-Yarrow, and Girnius-Brown (1987) observed children between ages 1;6 and 3;8 in naturalistic interactions with their mothers, focusing on occurrences of different forms of non-compliance. Younger children relied more on acts of passive non-compliance (not responding to directions, but not openly opposing them either) and direct defiance (overt refusal, with negative affect). But with age, this gave way to negotiation and justification, i.e., trying to elicit a new mutually agreed upon directive, giving or asking for explanations. Mothers’ strategies also shifted from direct control to reasoning and suggestion (e.g., referring to norms or consequences). This bidirectional relationship between mothers’ and children’s behaviour in conflict has been repeatedly captured in various observation studies.

Towards three years, children use justifications in about 30% of family disputes (Dunn & Munn, 1987), mostly drawing on their own feelings as reasons, less often on social rules. An example for reference to own feelings is responding to a parent’s request with, “No. I tired”. Reference to social rules can be about possession, when the child holds a sibling’s toy and the sibling takes it saying “That doesn’t belong to you”. The topic of the dispute predicted justification use. For example, acts of aggression or destruction rather elicited prohibitions instead of justifications. This may reflect a recognition that some contexts allow dispute to really achieve communicative goals, while in others, nothing can be won by discussion. This underscores children’s developing skill to put communication to use to serve their own goals (Dunn & Munn, 1987). In their fourth year, children increasingly focus interactions with their mothers upon this goal (Tesla & Dunn, 1992).

Goetz (2010) analysed four children’s interactions recorded longitudinally from 2;6 to 4;11 and observed increases both in the quantity of justifications and the breadth of situations in which they were used. In their third year, children used fewer justifications than their parents and mostly in the context of refusals and conflicts. At that age, parents had an important role in eliciting these interactions from children: 27.1% of children’s jus-
Justifications in that age range were in response to parents’ questions. In their fourth year, these children went through a phase of asking “why” questions, eliciting numerous explanations from their parents, and producing already about as many justifications themselves. At this age, they also engaged in self-expansions, proactively recognising when their parents needed more explicit information. The significance of refusals and conflict, in turn, decreased sharply in this period. In their fifth year, children’s self-expansions gained further prevalence, becoming an important context for children’s use of justifications.

While parents are naturally the foremost interlocutors in early language development, speakers do not meet on an equal footing. The asymmetry in power, knowledge and discursive role in this relationship is not characteristic of how equals, such as peers, approach each other’s contributions to a dispute (Mammen, Köymen, & Tomasello, submitted). Studies on reason giving among peers and siblings therefore provide different insights into children’s reasoning.

Interaction with peers and siblings. Giving reasons turns out to be just as successful a strategy with play mates as it is with parents—indeed, the most successful one in peer disputes. Eisenberg and Garvey (1981) observed children between 2;10 and 5;7 in peer play, analysing adversarial episodes. Here, giving reasons for their opposition was among children’s most frequently used strategies, most often leading to conflict resolution. Also, the authors note that children increasingly expect to get more than a simple “no” from opposing peers, as well as that reason leads to more reason, more compromise and less insistence.

Phinney’s (1986) observations of naturally occurring disputes around five years of age added the variable of whether same-age peers were siblings or friends. Disputes are strongly influenced by who children are talking to, also beyond the parent–child distinction. After all, children do not pick their parents or siblings; they are present a large proportion of the time at home, so there are more situations in which dispute is inevitable. As a result, in these observations, disputes at home were more repetitive, so that minor quarrels quantitatively played a larger role here than in conversations with peers. In other
words, disputes with peers, to the extent that they occurred, were more likely to be motivated by substantial disagreements. Similar to what Dunn and Munn (1987; see above) observed in talk with mothers, peer talk also had fewer justifications where there was little to win, namely, in conflicts around possession or name calling. Talk about the truth of some belief and about how “we” proceed yielded more frequent use of justifications. Additionally, Phinney (1986) made the common observation that children tend to respond to reasons with more reasons.

An important aspect in peer discussions is how children deal with the requirements of discourse. Knowing what interlocutors do or do not know, where they are mistaken and how to communicate this, is key for reasoning to succeed. Adults tend to give a lot of scaffolding in this respect (Kruger & Tomasello, 1986), but for peer discussions to work, children need to learn to manage common ground (Clark, 1996). Common ground refers to the set of beliefs that interlocutors know they share, so communication can build upon these beliefs. These shared beliefs can for instance result from the immediate context (personal common ground) or the fact that we are part of the same cultural group (cultural common ground). Köymen, Mammen and Tomasello (2016) found that both 3- and 5-year-old children are able to adjust the informativeness of their explanations to the personal common ground that they share with their peer. In this experiment, dyads in each age group had to make a joint decision about a fantastical animal, about which they learned some important information (e.g., that it eats rocks). This information was either learned together and shared in common ground, or learned individually and unshared. When information was unshared, knowledgeable children justified their vote in the joint decision more often explicitly, tailoring their reasoning to the informational needs of their peer (see also Köymen, Rosenbaum, & Tomasello, 2014). Mammen, Köymen and Tomasello (2018) showed this ability in the same two age groups also applied to cultural common ground. The type of cultural knowledge used in this study were moral norms, which were contrasted with more arbitrary conventional norms. Peers had to make and justify to each other a joint decision about two characters, one of whom had committed a transgression of either a moral or conventional norm. For instance, in one story, a charac-
ter would put a yellow toy into a green box. This was presented as stealing (moral) in one condition, and as breaking a rule that only permits green items in the green box (conventional) in another condition. Children were overall less explicit when the transgression was moral, suggesting that they recognised the absence of a need to explain the rule that they knew they shared with their peer.

While managing common ground is important for producing and understanding reasons, it is also essential to understand that an interlocutor’s view towards a belief can differ in the first place. In a theoretical paper, O’Madagain and Tomasello (submitted) call this joint attention to mental contents, or JAM: Two individuals can attend to the same belief while recognising that each does not necessarily have the same perspective; one may believe that a box is empty, another that it is not empty. The authors examine evidence of the influence of language input to the development of this key skill and how it is necessary for reasoning. Specifically, they identify sentential complement clauses (SCC, such as “I know/guess/hope that [there is an apple in the box]”) as an early context that confronts children with this type of attention. On this proposal, encountering SCC in their language input and accumulating practice in their comprehension and use gives children a grasp of JAM. This grasp requires children to both attend to the idea that “there is an apple in the box” and evaluate whether this statement is true or false. Experimental evidence suggests that exposure to SCC can improve children’s performance in explicit false belief tasks (Lohmann & Tomasello, 2003).

In addition, O’Madagain and Tomasello (submitted) propose that mastering justified sentential complement clauses (JSCC, such as “I believe there is an apple in the box, [because I saw it in there]”) is especially important for reasoning development. Being exposed to these justified sentential complement clauses gives children the opportunity to learn that beliefs can be justified, and differently well so. So they can form the motivation to change their interlocutor’s beliefs as they think they should.
3.4 Summary of the developmental literature

Reasoning as a social phenomenon is quite likely to originate in the context of opposition that toddlers experience early on when their intentions are misaligned with their parents’. As children’s language competence develops gradually, so does their ability to pinpoint their disagreements and to make their reasons for them explicit. Social reasoning relies on a range of skills:

- mustering the necessary affective control to formulate opposition verbally, instead of through direct defiance or other, simpler strategies;
- knowing what counts as reason generally and in a particular domain; for instance, “I saw that X” generally indicates better justifications than “I guess that X”; “I’m not tired” can justify opposition against bedtime, but not “I’m not hungry”;
- recognising that people might have different attitudes towards one belief;
- recognising that each attitude can be differently well supported by reasons;
- identifying in common ground what beliefs can be leveraged and what beliefs need to be addressed to change an interlocutor’s attitude.

The literature shows that by five years of age, children master these skills sufficiently to engage in meaningful reasoning, making adept attempts at changing interlocutors’ minds based on relevant reasons.

4 Focus of the dissertation

Two main ideas sum up the introductory part of this thesis: the view that reasoning is a social phenomenon that amounts to the practice of changing hearers’ minds with the help of reasons that are relevant and compelling; and evidence that children around five years of age can identify and express such reasons. The focus of this thesis is to extend in two important dimensions the evidence for the social view on reasoning and for how this skill develops.
Chapter One

One dimension is the flexibility with which children put their evaluations of reasons to use, to go beyond this evaluation and make further inferences. Chapter Two reports a partner choice experiment in which children, in addition to evaluating reasons, also judge partners based on how they evaluate them.

The second dimension is children’s motivation for producing reasons. Chapter Three presents a study that investigates how children reason with each other when they are motivated to cooperate or to compete. The article that underlies this chapter, published as Domberg et al. (2018), reports two studies on this question. Study 1 of this paper is not among the experiments encompassed by this dissertation, but is reported in the chapter nevertheless to provide introductory background and motivate Study 2. Chapter Four presents a study that investigates whether children that share a joint goal increase their efforts in pursuit of good reasoning when they compete together against others.

Finally, Chapter Five reviews and sums up the findings in this thesis, puts them into perspective against the theoretical background and highlights future directions.
Chapter Two

Children prefer to reason with partners who respond to good reasons

A near-identical version of this chapter is currently under review as a research article at Child Development as “Domberg, A., Köymen, B., & Tomasello, M. (under review). Children prefer to reason with partners who respond to good reasons.”

1 Introduction

A rapidly growing literature characterizes reasoning as a social phenomenon. On the one hand, Mercier and Sperber (2011) argue that the main function of reasoning is for speakers to think of reasons that persuade group members, e.g., of a given belief or decision, and for hearers to be critical of reasons when they contradict their own interests. On the other hand, Tomasello (2014) emphasizes that reasoning is also critical in collaborative decision-making, when groups need to find a good and justified solution benefiting all group members, rather than competing for individual argumentative success. In cooperative contexts, where the paramount goal is to come to a good conclusion, people should prefer reasoning with a partner who can distinguish good reasons from bad ones, even if that means that they might sometimes lose the argument, and who values good practical outcomes above all.

Children are vigilant social learners who selectively trust some information sources more than others. Four-year-olds prefer accurate informants over inaccurate ones (Birch et al., 2008; Corriveau & Harris, 2009; Jaswal & Neely, 2006; Koenig & Harris, 2005). When accuracy cannot be verified and children need to attend instead to how informants justify their conclusions, 2- to 3-year-olds prefer those who give good reasons for their proposals (e.g., “The dog went this way because I saw it go this way.”) as opposed to circular ones (e.g., “The dog went this way because it went this way.”); Castelain, Bernard, & Mercier, 2018; Castelain, Bernard, Van der Henst, & Mercier, 2016; Corriveau & Kurkul,
Similarly, Koenig (2012) showed that 4-year-olds trust conclusions that are based on good epistemic grounds such as perceptual access, (e.g., “I looked and I saw an apple in the box.”) more than those based on wishful thinking (e.g., “I like apples. I want there to be apples in the box.”).

However, in all these studies, children were asked to choose between informants, where the informant providing the good reason was likely perceived as simply “right”, and the other informant providing the poor reason was likely perceived as “wrong”. A more conservative test of children’s sensitivity to the quality of reasons would thus be to investigate whether children still prefer an informant who produces good reasons, when both informants are wrong or have a false belief (e.g., “I saw that there is an apple in the box”—but the apple is no longer there). Moreover, in all of these studies, children were often asked to choose between individuals who produce reasons, instead of individuals who evaluate these reasons. Children’s ability to identify and trust individuals who are able to distinguish good reasons from poor ones is a more advanced skill, because it requires children to not only evaluate the quality of the reasons and detach it from their own beliefs about the truth, but also to evaluate another person’s evaluation of reasons.

In this study, therefore, we presented 5- and 7-year-old children with a cooperative game for which they needed a partner. The purpose of this game was to give children a profile of their ideal partner: This person would have to reliably tell good from bad reasons and consequently act in a well-reasoned manner. Once these requirements in a partner were established, children observed two candidates together in a problem-solving task, namely, finding a lost paintbrush. Candidates heard together two reasons from two disagreeing informants about where the brush is. Crucially, the candidates were ignorant and the informants had a false belief about the location of the paintbrush. In the experimental condition, one informant produced a good reason based on an eye-witness account (“Check the red box [incorrect location] because I saw it there yesterday”), and the other informant produced a poor reason based on preference (“Check the yellow box [incorrect location] because yellow is my favourite colour”). In the control condition, one informant produced the same good reason based on an eye-witness account for the same
incorrect proposal, and the other informant produced another good reason based on rules (“Check the yellow box [incorrect location] because we always put it there”). In both conditions, one candidate followed the good reason based on eye-witness and the other followed the other reason. We predicted that in the experimental condition, children would choose the candidate who followed the good reason over the candidate who followed the poor reason; whereas in the control condition, children would not prefer one candidate to the other, because both candidates followed a good reason. We explored if there were any age differences. Although 5-year-olds are good at evaluating reasons (Corriveau & Kurkul, 2014; Mercier et al., 2014) and show false belief understanding (Wellman, Cross, & Watson, 2001; Wimmer & Perner, 1983), 7-year-olds have been observed to be more strategic in their reasoning than 5-year-olds (Domberg et al., 2018). For instance, their withholding of reasons when this brings an advantage in a competitive situation shows that their pragmatic understanding of reasons improves clearly between the two age levels.

2 Method

2.1 Participants

Ninety-three 5-year-olds ($M = 5;9$, $Range = 5;3–6;2$, 50 girls) and 69 7-year-olds ($M = 7;6$, $Range = 7;3–7;9$, 32 girls) participated in this study in a between-subjects design. All children were native German speakers with various socio-economic backgrounds, recruited and tested at 15 urban daycares and six schools distributed across the city. Thirty-two 5-year-olds could not be included in the analysis, because they gave incorrect responses to false belief questions (24 children), failed to recall one or both reasons after cues (5 children), or were uncooperative, e.g., refusing to choose a partner (3 children). Eight 7-year-olds could not be included in the analysis due to incorrect responses to false belief questions (5 children), failure to recall one or both reasons after cues (2 children), or experimenter error (1 child).
2.2 Materials

For the cooperative game, we used an 85 cm × 31 cm laminated background with six panels with different colours, each depicting two sets of footprints leading up to two different animals (see Figure 2.1). At the beginning, each animal was covered by a cardboard box with a slot, and only its footprints were visible. The child and experimenter had identical chips, depicting a desired item for an animal. They had to infer what kind of animal would need the depicted item, which of both sets of footprints belonged to this animal, and coordinate their decisions by cashing their chips into the same correct box. If
correct, the child would win marbles to stack in an upright transparent tube with a finish line. For the partner choice, the story describing the candidates was told using a 15-page booklet (see Appendix A for the pictures and the exact narration in English).

2.3 Procedure

**Warm-up phase.** For the cooperative game, the experimenter (E) explained to the child (C) the two were playing a detective game. E and C would have to infer together what animals were hidden under the two boxes based on their reading of the footprints, and place the two identical “gifts” (food items printed on each chip, see Figure 2.1) in the box of the correct animal. If both E and C placed their chips in the same correct box, C would get two marbles. For instance, in the first round, E and C’s chips each depicted a glass of honey. E asked C what animal might be happy about honey (proposed bear, if necessary). Then E and C had to decide which set of footprints belonged to the bear. Given that one set of footprints belonged to a bear, the other to a mouse, E asked where C wanted to put her/his chip (reliably the box with bear traces) and asked C to put the chip there. E then inserted his own chip into the same box. They verified together that the animal under that box was a bear. They won two marbles.

In the second and third round, the procedure was parallel to the first, except that E acted incompetently. After C entered her/his chip into the correct box, E first asked where he should put his, but upon C’s response remained doubtful about the right choice and inserted his chip into the incorrect box, against any protest. After the third round, E pointed out that C seemed to be a much better detective and would need a better partner than E. E introduced the storybook about the two candidates.

**Story phase and partner choice.** In the storybook, page 1 depicted the photos of the two candidates side by side. E explained that one was a good detective, just like C, and the other was not, and that C had to pay attention to what these two said in the story in order to know which was the good one (see Appendix A for the narration). On pages 2–3, the informants were introduced as drawings so they were distinguishable from the can-
didates. The informants were painting with a brush (p. 4), then tidied up and put the brush into one of two boxes, red or yellow, without making explicit which one. When informants left the room, a thief stole the brush (p. 5). From this point forward, the thief with the brush sticking out of his bag was permanently visible next to the storybook. The next morning, the two candidates entered the room to do some painting but could not find the brush (p. 6). The informants reappeared to help them out (pp. 7–8).

On page 8, the first false-belief question (out of six) was introduced, which concerned the two informants collectively (“Do these two know where the brush really is?”) to make sure that children do not attribute knowledge to the informants. Any incorrect responses to this first question were corrected to establish the norm that answers should be guided by information from the story, and not, e.g., guessing. Any incorrect responses to any of the remaining five false belief questions led to exclusion from the analyses.

On pages 9–12, one informant proposed to search in one box and the other informant the other box. One informant gave a good reason based on an eye-witness account, “because that is where we put the brush yesterday, and that is also where I last saw it”. In the experimental condition, the other informant gave a poor reason based on preference “because [colour of this box] is my favourite colour, and I like that colour the most.” In the control condition, the other informant gave another good reason based on rules “because that is where we actually always put the brush, and that’s where it also actually belongs.” Then, the next two false belief questions were asked about each of the two informants (“Does [this informant] know where the brush really is?”).

Next, one candidate followed the good reason based on an eye-witness account. E asked C to recall the reason for this proposal (memory question), followed by the fourth false belief question about this candidate (“But does [this candidate] know where the brush really is?”). The other candidate followed the other reason based on preference/rules. E asked C the same memory question about the reason for the proposal, and the fifth false belief question about this candidate (“But does [this candidate] know where the brush really is?”). Children who responded incorrectly or not at all to either of the two memory questions (34 children, 27.9%), received clues for giving the correct an-
answer. For the reason based on an eye-witness account, the clue was that it was about the previous day; for the rule-based reason, that it was something that is always the case, and for the preference-based reason, the question whether the informant disliked the colour. All but seven children gave the correct response after the clues (these seven were excluded from the analyses). Then, given a page showing all five characters including the thief, E asked the final, collective false belief question, “If you look at these five guys, which of them is the only one who knows where the brush really is?”

On a summary page (p. 14) with the pairs of boxes, informants, and candidates arranged such that they reflected who referred to what and whom in the story, E repeated the informants’ and candidates’ statements and asked C, “Which of these two [point to the candidates] is the better detective for our game?” Upon C’s pointing or naming, E asked why. Throughout the story and its summary, we counterbalanced the colour of the boxes, identity of the informants giving a reason based on an eye-witness account, the identity of the candidates agreeing to the reason based on this account, and the presentation order of the two reasons.

For eight children, there were minor experimenter errors but these children were nevertheless included in the analyses. Five children did not receive the first non-critical false belief question, which just reminded children not to guess. Two children were told (instead of asked) about the first informant’s false belief. One final child was told (instead of asked) about both informants’ and a candidate’s false belief. However, this final child gave a very sophisticated explanation for her choice, showing that she paid attention to the false belief of the characters and the quality of the reasons (“Because he rather responded to what happened last. You can’t just go after your favourite colour ...”). Crucially, all of these eight children responded correctly to all remaining questions, so we opted to include them in the analyses.

After that, E explained that neither of the candidates was able to make it to the school/daycare, so C could finish the game with E and won the reward. The sessions lasted about 15 minutes and were video-recorded.
2.4 Coding

First, we coded which candidate children chose. Then, we assigned children’s explanations for their choices to one of three categories:

1. reason: children made reference to the reason, identifying the critical difference between both options with regard to the quality of the reasons and/or false belief of the informants/candidates (e.g., “Because the favourite colour has nothing to do with it. But if you’ve last seen it, it’s actually better”);
2. simple: children only referred to the candidate/informant/the box, without any reference to the reasons (e.g., “Because he listened to little Max [the informant]”);
3. none: children gave irrelevant responses or no response (e.g., “Because blue is my favourite colour”).

A second rater coded 32 of the 122 responses to the request for explanation, and the agreement was $\kappa = .83$.

3 Results

As Figure 2.2 shows, in the experimental condition, 70% of the 5-year-olds and 87% of the 7-year-olds chose the candidate who followed the good reason based on an eye-witness account; whereas in the control condition, this was the case with 35% of the 5-year-olds and 57% of the 7-year-olds. In both age groups, the number of children who preferred the candidate following the good reason based on the eye-witness account was significantly higher in the experimental condition than in the control condition (Fisher’s exact, $p = .010$, odds ratio = 4.13 for 5-year-olds, and $p = .011$, odds ratio = 5.02 for 7-year-olds).

We also compared the number of children who chose the candidate following the good reason based on the eye-witness account to chance in each condition using binomial tests. In the experimental condition, children in both age groups chose this candidate significantly above chance (5-year-olds: $p = .043$, 7-year-olds: $p < .001$), whereas in the con-
Results

In the control condition, their choice between candidates was at chance (5-year-olds: \( p = .150 \), 7-year-olds: \( p = .585 \)).

Also, Figure 2.2 shows that in both conditions 7-year-olds produced more explanations for their choice that referred to the relevant reason than did 5-year-olds (Fisher’s exact, \( p = .006 \)). The amount to which children made reference to the reasons (56 out of 122) also suggests that distinguishing the reasons was not an issue. In addition, the lower part of the plot indicates how many children failed the task with regard to the correct comprehension of the narration and reconstruction of the facts and beliefs. Of the 32 5-year-olds and eight 7-year-olds who were excluded, 24 and five, respectively, were because they failed in this aspect of the task.

4 Discussion

Our results suggest that in the experimental condition, 7-year-olds, and 5-year-olds less reliably due partly to high drop-out rates, preferred the candidate who followed the
good reason over the candidate who followed the poor reason. However, in the control condition, where both candidates followed good reasons, they showed no preference. Crucially in both conditions, both candidates arrived at the wrong decision. Thus, our study is first to show that children are able to evaluate 1) the quality of reasons, even if these lead to incorrect decisions, and 2) how others evaluate reasons.

Our study specifically rules out that children, if they selected the follower of the good reason, did so based on the assumption that this person is more knowledgeable or simply right. Through the false belief questions, we made sure that children knew that neither candidate knew better. Moreover, the control condition rules out the alternative interpretation that children’s preference for the candidate following the good reason based on an eye-witness account in the experimental condition was due to an enduring belief that the brush was still in the indicated box. If this were the case, children would have preferred this candidate in the control condition as well.

When explaining their choice, 7-year-olds more often than 5-year-olds produced explanations that referred to the relevant reason. This is in line with the finding that at school age children start to produce more complex reasons for their claims (Domberg et al., 2018; Köymen & Tomasello, 2018). Similarly, Piaget (1932) reports (for a task of comparing stories, however) difficulty eliciting judgments from children younger than six years. It should be noted that explanations categorized as “simple” (e.g., “Because he agreed with X”) were ambiguous in terms of what guided children’s decision, but these explanations were, in a way, optimally informative. The research on children’s use of common ground supports this view (Köymen et al., 2016; Mammen et al., 2018). That is, since the experimenter knew the story, children might have assumed some common ground with him and thought that it was obvious to the experimenter that the candidate who agreed with informant A was better than the other candidate.

Despite showing the same pattern as 7-year-olds, 5-year-olds did have some difficulty with the task. Our story was indeed a long story with five characters. Twenty-five 5-year-olds, who could not be included in the analyses, answered at least one false belief question incorrectly and thus attributed knowledge to one of the four characters (two in-
formants and two candidates). This high rate of incorrect responses to false belief ques-
tions by 5-year-olds is surprising given the literature on false belief understanding (Well-
man et al., 2001). One potential explanation could be that when children answered five false-belief questions in which the correct answer was always “no”, they might have found it pragmatically odd to answer “no” to all questions and switched some of their an-
swers to “yes”.

To conclude, starting at age five, and quite reliably at age seven, children distinguish between speakers who submit to “reason” and those who do not. They distinguish good reasons from poor ones, even if neither reason leads to a better solution, and prefer those partners who do the same.
Chapter Three
Children’s Reasoning with Peers in Cooperative and Competitive Contexts

This chapter is published as:

As the published paper, it reports two studies. Only Study 2 has been part of this dissertation project, Study 1 has been carried out in the context of my Master’s thesis. Together with the Introduction, however, it provides the necessary groundwork that motivates and contextualises Study 2.

1 Introduction

Reasoning is a core aspect of decision-making. In their Argumentative Theory, Mercier and Sperber (2011) characterize reasoning – in the sense of giving and evaluating reasons – as a social skill, with which speakers win others over to their point of view. The theory is motivated by the concept of Epistemic Vigilance (Sperber et al., 2010), which guards hearers from being deceived and requires speakers to support their claims with mutually accessible evidence to overcome this mistrust. Tomasello (2014), on the other hand, emphasizes that reasoning is also important in cooperative contexts. Because all parties share an interest in reaching the best decision, who produces the ‘winning’ argument becomes less important.

Children’s use of arguments has been mostly studied in the context of peer conflicts (Dunn & Munn, 1987; Eisenberg & Garvey, 1981; Phinney, 1986). It has been observed that two- to five-year-old children justify opposition, and use of conciliatory justifications increases with age. Recent studies investigated how children reason cooperatively with peers, finding that in justifying their proposals, three- and five-year-olds adjusted the in-
formativeness of justifications to their mutual knowledge with their partners to reach the correct joint decisions (Köymen et al., 2016, 2014). However, studies on young children’s use of justifications focused either entirely on conflictual/competitive or entirely on cooperative contexts without a systematic comparison of how young children’s arguments change across both contexts. Such comparison would reveal more about children’s ability to pursue communicative goals argumentatively.

Research with adolescents compared speakers’ argumentation across contexts, however. Felton and Kuhn (2001) had adolescents discuss controversial topics in dyads of consenting or dissenting partners. Consenters were asked to defend their view, and dissenters were asked to reach agreement. Compared to adults, adolescents showed much less strategic argumentation, for example weakening a dissenter’s argument. Kuhn and Udell (2003) found that only after training, adolescents produced such powerful arguments as refutations, refuting potentially opposing arguments, in contrast to affirmations, which advance the speaker’s point of view. More importantly, when adolescents were instructed to reach agreement, in contrast to winning over their interlocutor, they produced more complex arguments (e.g., refutations) and less biased argumentation, showing ‘two-sided reasoning’ (Garcia-Mila, Gilabert, Erduran, & Felton, 2013; see Kuhn, 2015 for a review).

Our main question here is whether the differences in argumentation observed with adolescents across cooperative and competitive contexts emerge earlier in development, in five- and seven-year-old peers’ conversations. We chose these age groups because around age 5, children were observed to produce and critically evaluate arguments (Corriveau & Kurkul, 2014; Köymen et al., 2016, 2014; Mercier et al., 2014). In Study 1, dyads of five- and seven-year-old peers were asked to reach agreement about placing toy animals in a zoo. Each child owned one half of the zoo. The rewarding scheme either favoured finding nice homes for every animal (‘finding the best solution’ – cooperative condition), or having more animals on one’s own side (‘winning’ – competitive condition). We predicted that the cooperative condition would elicit more objective two-sided argumentation than the competitive condition. We explored any differences between conditions or age groups in the overall frequency, content, and especially type of arguments, that is, whether they
were affirmations or refutations, as the literature suggests that even adolescent children tend to focus on the explanations for ‘why’ (justifying a desired outcome), rather than ‘why not’ (refuting an undesired outcome; Kuhn & Udell, 2003).

In Study 2, children played a similar game, but we made sure children knew a set of arguments to test a specific interpretation of the results from Study 1, namely, whether the bias observed in the competitive condition had cognitive or strategic reasons. One child per dyad first played the game with the experimenter and learned critical arguments. Later, he/she played with a naïve peer either in the cooperative or competitive condition. Critically, in the competitive condition, by reproducing the critical arguments, the trained child would benefit the naïve child and worsen his/her own chances of winning. We predicted that in the competitive condition, trained children would withhold these arguments more often. As in Study 1, we explored any age differences in cooperative and competitive contexts.

2 Study One

2.1 Method

Participants. Forty-eight 5-year-olds (M = 5;8, range = 5;5–5;11, 24 girls) and 48 seven-year-olds (M = 7;6, range = 7;0–7;11, 24 girls) in 48 same-age and same-sex dyads participated in the study in their kindergartens and schools. Children in dyads knew each other (although their degree of familiarity with each other varied) and were randomly assigned to the cooperative or competitive condition in equal numbers. All children were native German speakers with various socio-economic backgrounds. Participation was subject to prior parental consent, and the procedure was approved by the departmental Ethics Committee at the Max Planck Institute for Evolutionary Anthropology.

Materials. Children placed toy animals onto two backgrounds. One was used as warm-up, comprising two pictures. The other was used in the testing phase, depicting a zoo with four cages (Figure 3.1). The two cages in each half were the following: (1) a ‘warm cage’, suitable for animals in warm climates; (2) a ‘cold cage’, suitable for animals...
in cold climates. The cold and warm cages in each half differed so that children could produce different arguments. Moreover, the orientation of the cages on each half was such that each child would see their own cages upright, but their peer’s upside down.

The toy animals were released from a cardboard dispenser with 17 slots. Pulling the lowest card would release the animal at the bottom slot into a transparent box with a lid (see Figure 3.2 for the list of animals). Finally, a ‘surprise box’ held stickers as rewards.

**Warm-up Phase.** The study took place in quiet rooms of preschools in a German city. The procedure took c. 15 min. The experimenter (E) seated both children facing one another with the dispenser to their side. E first introduced how the dispenser works by asking children to pull out the lowest card together. When the animal was released into
the transparent box, E explained the game rules: ‘First we say what animal this is, then to which cage it should go and most importantly why we put it there’. In the warm-up phase, E primed children to give reasons and exemplified three argument topics: one about habitat, two about animal safety/interaction within cages.

The first animal, introducing the habitat argument, was a fish. Almost all children said that it should go to the ocean, because it lives there. If children’s justifications did not involve the habitat topic, E provided the justification: ‘because fish live in water’. The second item was a group of baby chicks. Again, almost all children said they were chicks, they should go to the farm because they live in farms. If they did not, E provided the justification. The third animal was a baby fox to prime the safety argument. All children said that it was a baby fox, and it should go to the farm. E primed the safety argument by saying, ‘And foxes hunt chicks, that is why it also wants to go to the farm. Now the chicks get scared, right? But, it’s a baby fox. It can’t hurt the chicks yet, so let’s place them together’. The safety arguments alerted children to attend to what animals they place together in cages.

Figure 3.2: Animal items to be placed in the zoo.
The toy animals used in the warm-up and testing phases in the order of appearance. This order was constant for all dyads. Items 5 and 8 were pairs of usually conflicting animals and meant to elicit discussion.
**Experimental Phase.** After the warm-up, E introduced the zoo, placing the zoo background with four cages (see Figure 3.1 b) and said,

“Our zoo has four cages, they have sand like this [point: warm cages], or an iceberg like this [point: cold cages]. This is your half, [A], and this is yours, [B]. Now you play the game like before. First, say what animal it is; second, where it should go; and then why it goes there. And all animals go into cages and not outside, because otherwise they scare the visitors.”

Children’s behaviour indicated no difficulties in understanding the meaning of the materials.

The final part of the instructions differed across conditions. In the cooperative condition, E said, ‘If you find a nice home for each animal, you will both get rewards from this box [presents the gift box], but remember: they have to be nice homes. So, find a nice home for each animal, and you will both get the rewards from this box’. In the competitive condition, E said, ‘If you, [Child A], have more animals on your side, only you, only [Child A], can get the reward from the box. And if you, [Child B] have more animals on your side, only you, only [Child B] can get the reward’. From an individual perspective, both children were motivated to win in both conditions. In the competitive condition, ‘winning’ meant ‘winning individually’ (i.e., one child gets a reward), whereas in the cooperative condition, it meant ‘winning together’ (i.e., each child gets a reward). Therefore, in case of winning (and losing), the individual pay-off was the same in both conditions.

After the instructions, E left the room, and re-entered after the fourth animal was released to remind children of the game rules, ‘Remember you always have to say why each animal should go there’. After that, E did not enter until the last animal was placed, unless an animal was stuck in the dispenser, children placed an animal outside the cages, or got distracted.

**Coding.** Children’s conversations during the experimental phase were transcribed verbatim, each line containing a clause with a single proposition. The transcript was divided into segments pertaining to the discussion of each of the 14 items.
We first identified children’s game-related utterances addressed to one another. Within the game-related utterances, we identified arguments, defined as justifications for why an animal should or should not go into a cage. We excluded from the analyses those arguments that repeated or rephrased a previous argument on the same item. For each argument, we coded three properties: its content, favoured side, and type.

First, we coded arguments into five content categories:

1. **habitat**: arguments about the suitability of the cage (“It [:the polar bear] needs a cave to hide”);
2. **family**: arguments construing adult and young animals of a species as kin who go together (“Because its mother is there.”);
3. **safety**: arguments about animal safety (“He [:the lion] would eat the zebra.”);
4. **species**: arguments about placing the same species of animals together (“There are already some [:penguins] there.”);
5. **game affordances**: arguments about the number of animals (“It’s too crowded there!”). This last category was not primed in the warm-up and children spontaneously referred to this in their justifications.

Next, we coded each argument in terms of which cage it favoured: the speaker’s cages (‘own’ arguments), the peer’s cages (‘other’ arguments), or unclear/both (e.g., “Let’s put it on an iceberg”—which both children had).

Finally, we coded arguments in terms of their type: affirmations explained why an item should go to a cage (e.g., ‘Because it can swim there’) and refutations explained why it should not go to another cage (e.g., ‘Otherwise the snake [: in his peer’s cage] would bite it’).

A second rater coded 25% of transcripts (12 dyads: 3 dyads per age group and condition). Agreement was \( \kappa = .63 \) for game-related utterances, \( \kappa = .81 \) for being arguments, \( \kappa = .83 \) for argument content, \( \kappa = .95 \) for the targeted cage, and \( \kappa = .90 \) for argument type.
Forty-eight dyads produced a total of 7,177 utterances (94% of these utterances (6,760) were game-related). These included a total of 1,241 arguments. Within these arguments, 1,169 (94%, $M_{per\ dyad}$: 24.4, SD: 18.3) were (i) not ambiguous in terms of their goal, and (ii) not repetitions.

There were four sets of analyses. In the first three, the unit of analysis was the dyad. We did not distinguish between child speakers within dyads, because observations were not independent. First, we analysed whether the frequency of arguments among game-related utterances differed across age groups and conditions with a two-way ANOVA with age (5- vs. 7- year-olds) and condition (cooperative vs. competitive) as between-subject factors, and gender as control factor. The response variable was the proportion of arguments to total game-related utterances. This analysis revealed two significant main effects of age and condition (Figure 3.3). 7-year-olds produced more arguments than 5-year-olds ($F(1,43) = 13.71, p < .001, \eta_p^2 = .24$), and both age groups produced more arguments in the cooperative condition than in the competitive condition ($F(1,43) = 6.31, p = .016, \eta_p^2 = .13$). There were no gender differences.
Second, we ran a two-way ANOVA with the same factors and the response variable of number of topics that each dyad based their arguments on, averaged across items. Two 5-year-old dyads in the competitive condition were excluded from this analysis, because they did not produce any arguments. This analysis revealed a trend towards more diverse arguments in the cooperative condition than in the competitive condition ($F(1,41) = 2.87, p = .10, \eta_p^2 = .13$), and significantly more diverse arguments in 7-year-olds’ conversations than in 5-year-olds’ ($F(1,41) = 13.71, p = .001, \eta_p^2 = .24$).

Third, we investigated how often children favoured their own cages. We ran a two-way ANOVA with the same factors and the response variable of proportion of ‘own’ arguments to the total number of arguments produced per dyad. Again, the two 5-year-old dyads that did not produce any arguments were excluded. There was only a significant main effect of condition: both age groups favoured their own cages significantly more often in the competitive condition than in the cooperative condition ($F(1,41) = 12.46, p = .001, \eta_p^2 = .23$). There were no age or gender differences.

Finally, we investigated to what extent children used the two types of arguments (affirmations and refutations) in their ‘own’ vs. ‘other’ arguments, across age groups and conditions, using a Generalized Linear Mixed Model (GLMM) with binomial error distribution. The unit of analysis was individual arguments and the response variable was the binary measure of whether an argument favoured the speaker’s cages (‘own’ arguments) or their peer’s (‘other’ arguments). The full model included the predictors: age group (5- vs. 7-year-olds), condition (cooperative vs. competitive), argument type (affirmation vs. refutation), and their 3-way interaction, the control factor of gender, and the random factor of dyad (as we had repeated observations per dyad). To test the significance of the full model, we compared its fit to that of a null model, which only included the gender control factor and the dyad random factor. The full model improved the fit ($\chi^2 = 24.74, df = 7, p < .001$). Compared to a reduced model without the 3-way interaction, the full model still improved the fit, suggesting a significant 3-way interaction ($\chi^2 = 5.30, df = 1, p = .021$; Figure 3.4).
To better understand the 3-way interaction, we ran post-hoc analyses on the subsets that correspond to combinations of age groups and argument types. These analyses corresponded to investigating separately for affirmations and refutations how 5- and 7-year-olds produced more self-serving arguments in the competitive condition:

- Using affirmations, five-year-olds favoured their own side significantly more in the competitive than in the cooperative condition ($\chi^2 = 7.12$, $df = 1$, $p = .008$).
- Using refutations, five-year-olds showed no condition difference as to which side they favoured ($\chi^2 = 1.04$, $df = 1$, $p = .307$).
- Using affirmations, seven-year-olds favoured their own side significantly more in the competitive than in the cooperative condition ($\chi^2 = 4.83$, $df = 1$, $p = .028$).
- Using refutations, seven-year-olds favoured their own side significantly more in the competitive than in the cooperative condition ($\chi^2 = 16.13$, $df = 1$, $p < .001$).

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure3.4.png}
\caption{Reason types and self-servingness. \newline The proportion of ‘own’ arguments in overall arguments, distinguished by argument type. \newline The difference marked in the proportions of refutations in the competitive condition refers to seven-year-olds’ better grasp of this argument type as compared to five-year-olds.}
\end{figure}
2.3 Discussion

Our results suggest that children in both age groups produced systematically different samples of arguments across the two contexts. In the cooperative condition, they produced not only more arguments, but also, the set of arguments that they produced was balanced in the sense of favouring the speaker’s and the peer’s side equally often. In the competitive condition, children favoured their own side, but did so differently across ages. Five-year-olds mostly relied on affirmations, explaining to their peer why a toy animal should go to their own side. In contrast, seven-year-olds relied not only on affirmations but also on refutations, explaining to their peer why a toy animal should not go to their peer’s side. Moreover, seven-year-olds overall produced more arguments and covered more topics in their arguments than did five-year-olds.

Since in the competitive condition, children’s arguments were fewer and favoured their own side, one possible explanation is that even though they noticed the evidence supporting arguments for their peer’s side, they ignored it. An alternative explanation is that they restricted the search for evidence to their own side and paid no attention to the evidence on their peer’s side. Considering that seven-year-olds in both conditions produced about an equal number of refutations in favour of their own side – arguments that are based on evidence on their peer’s side – they seemed to have considered the evidence on their peer’s side, but possibly abstained from producing the arguments that would use it. Five-year-olds, in contrast, produced fewer refutations and showed no condition difference with respect to refutations favouring either side. In their case, it is less clear whether they were aware of the evidence on their peer’s side.

Therefore, in Study 2, we controlled for children’s knowledge of specific arguments. One child, the trained child, played the game with the experimenter before playing with his/her naïve peer and learned a set of critical arguments that were unusual and easy to remember. These arguments were affirmations, more easily used by five-year-olds. In the testing phase, these critical arguments would support placing animals on the naïve peer’s side, benefiting the trained child in the cooperative condition, but worsening his or her
chances of winning in the competitive condition. Moreover, forming any arguments apart from the trained ones was made difficult for the trained children, since their cage offered only unusual objects (e.g., piano). We predicted that seven-year-olds would reproduce the critical arguments in the cooperative condition, but withhold them more often in the competitive condition. The reason is that their behaviour in Study 1, especially their production of self-serving refutations, suggested that they had considered both sides of the zoo in search for arguments, but seemed to withhold the ‘other’-serving arguments in the competitive context. We explored whether five-year-olds distinguish between conditions when having learned explicitly about their peer’s side encouraged them to attend to it.

3 Study Two

3.1 Method

Participants. Seventy-two five-year-olds (M = 5;5, range = 5;0–5;11, 36 girls) and 72 seven-year-olds (M = 7;5, range = 7;0–7;11, 36 girls) in 72 same-age and same-sex dyads, who did not participate in Study 1, participated in the study. Children in dyads knew each other (although their degree of familiarity with each other varied) and were randomly assigned to the cooperative or competitive condition in equal numbers. All children were native German speakers with various socio-economic backgrounds. This study fulfilled the same ethics and participation standards as Study 1.

Materials. The same dispenser, warm-up pictures, and surprise box were used as in Study 1. The animals in the warm-up were the fish and baby chicks. In the training and testing phase, the playing ground was a picture of a zoo (Figure 3.5). Each cage contained three objects, each of which had a circle by it, indicating where an animal should be placed. The “rich” cage contained an alarm clock, a toy brush, and a miniature handkerchief. The other, “poor” cage contained a toy lamp, piano, and refrigerator.

There were five animals to place (Figure 3.5): three critical ones, and two fillers. In the training phase, the three critical animals were a baby bear, baby lion, and baby ele-
Study Two

The filler animals were a zebra and penguin. In the testing phase, the three critical animals were a bear, lion, and elephant; the filler animals were a chimp and snake. We used baby and adult versions of the critical animals and different filler animals across the training and testing phase because our pilot data suggested that when the exact same animals were used in both phases, children treated the trained arguments normatively and did not act strategically.

**Warm-up.** The warm-up phase was identical to that of Study 1, but comprised only two items, a fish and a pair of baby chicks. The baby fox was not included because we did not want to prime children about the relations of animals placed in each cage. Another difference to the Study 1 warm-up was that when asking children for an argument, E1 accepted (or suggested if necessary) any arguments about what the animal could do at each location. For example, E1 said, ‘If I say ‘the fish goes into the water because it can breathe there’, is that a good reason?’ This was to prime children to critically evaluate the arguments.

**Training phase.** For the training phase, one child, the naïve (N) child, left the room, playing a game outside with another experimenter. E1 sat by the poor cage and seated the trained child (T) next to the rich cage, saying:

“This is your cage, and this is mine. We will play the game just like before. We say first the name of the animal, where it should go, and why the animal

![Figure 3.5: The setup and materials of Study 2.](image)

The main difference: only one cage on each side with three items in it, and a target circle (yellow) that embodies the rule “one animal per goal”. The animals above were used in warm-up, those below in testing, the two rightmost in each row being fillers.
should go there. That’s the most important part. The animals don’t go on the green outside the cages. Why? Because they would run away. The one who has more animals in their own cage at the end wins.”

The order of the five animals was pseudo-randomized, without the three critical animals appearing in a row. The order was the same in the training and testing phase.

For each animal, E1 first asked T in which cage one might put the animal and then announced having a great idea, pointing out: (a) a property of the animal; (b) an item in a cage; (c) what the animal can do with that item. For the three critical animals, these argument components were as follows:

- Bear: (a) bears hibernate; (b) the alarm clock; (c) it will know when winter is over.
- Lion: (a) lion has fuzzy fur; (b) the brush; (c) it can brush its fur.
- Elephant: (a) elephants have long trunks; (b) the handkerchief; (c) it can sneeze its long trunk.

For the two filler items, unless T gave other arguments, E1 proposed the locations in his own (poor) cage: The piano for the zebra because of its black and white fur matching the piano keys, and the refrigerator for the penguin because of its preference for cold climate.

This training phase was the same and competitive in both conditions. After the last item, E1 asserted T’s victory, and asked T to leave the room and join the others. E1 refilled the dispenser with the five animals for the testing phase, i.e. adult versions of the critical animals and new fillers. Both children were called back in. Importantly, T was asked to sit by the poor cage and N by the rich cage.

**Testing phase.** E1 told both children that the game worked just as before, and said:

“T already knows the zoo, this is our zoo. This is N’s cage, and this is T’s cage now. Let’s play the game just like before. After we see the animal, we say: its name, where it should go, and most importantly, why it should go there.”
Then, E1 gave the same final instructions according to the condition, as in Study 1 (see above). E1 left the room and entered only if the dispenser failed or when the children were done.

**Coding.** Conversations were transcribed verbatim. We coded that a critical argument was reproduced if T (1) mentioned the animal property (e.g., “It has this fuzzy fur”), or (2) mentioned the activity (e.g., “It hibernates”), or (3) acted out the activity, (e.g., brushing the animal’s fur, making a sneezing sound). If T produced the critical argument for an incorrect animal, this was still coded as reproducing the argument because the animal would end up on N’s side. Out of 72 dyads with three trials each, eleven naïve children came up with one trained argument spontaneously, one of them with two. A second rater coded the transcripts of 16 dyads for the reproduction of the critical arguments. The agreement was $\kappa = .72$ (87.5%).

### 3.2 Results

We ran a two-way ANOVA with age and condition as between-subject factors and gender as control factor. The response variable was the number of trained arguments that were reproduced by T (0–3). The analysis revealed two significant main effects of age group and condition. Seven-year-olds reproduced the critical arguments more often than five-year-olds ($F(1,67) = 11.73$, $p = .001$, $\eta^2_p = .15$). Children reproduced the critical arguments more often in the cooperative condition than in the competitive condition ($F(1,67) = 4.44$, $p = .039$, $\eta^2_p = .06$). However, looking at each age group separately, the condition effect seems to be driven by seven-year-olds (Figure 3.6): Seven-year-olds produced significantly more trained arguments in the cooperative condition than in the competitive condition ($F(1,33) = 7.41$, $p = .010$, $\eta^2_p = .18$), but five-year-olds showed no condition difference ($F(1,33) = .25$, $p = .619$).

### 3.3 Discussion

The results of Study 2 suggest that children produced less of the trained arguments, or more often withheld them, in the competitive condition than in the cooperative con-
condition. However, this holds rather for seven-year-olds than for five-year-olds. A possible explanation for the lack of a significant interaction between condition and age group is that most seven-year-olds in the competitive condition withheld only one trained argument. From a strategic standpoint, when the trained child withheld just one critical argument (and potentially secured one critical animal for his/her own cage), that was sufficient for winning the game.

Five-year-olds reproduced fewer critical arguments in both conditions than seven-year-olds. The lack of a condition effect for five-year-olds may have two related explanations. First, five-year-olds overall produce less arguments than seven-year-olds both in Study 1 and 2, and additionally, in Study 2, non-comprehension of the trained arguments might have also limited their production. Study 1 was more intuitive in its set-up, and provided more opportunities to produce arguments than Study 2. Second, five-year-olds may have suffered from limiting their search for evidence for their arguments to their own side of the zoo. Considering that this side only offered unusual items (e.g., piano), five-year-olds may have found producing arguments difficult especially in the competitive

Figure 3.6: Trained argument reproduction. The number of trained arguments reproduced by the trained child in each dyad. Bars represent mean values per group, dots show the outcomes for each dyad. Values indicate means and standard deviations.
condition. Therefore, in this condition, five-year-olds were less strategic than seven-year-olds.

3.4 General Discussion

Overall, our findings suggest that when children aim at a “good” solution, rather than “winning”, they approach the available evidence in a less biased way and produce more arguments. In the present studies, dyads considered more possibilities for placing the toy animals, and did not neglect or wilfully withhold arguments when they were after the best solution for a common goal. Thus, cooperative situations, in which children have joint goal of winning the game together, provide a more motivating context for argument production for both age groups.

Our studies also address the question to what extent a bias in children’s production of arguments can be attributed to strategic behaviour or to cognitive limitations. In Study 1, seven-year-olds in the competitive condition made restricted use of both affirmations and refutations that would favour their peer. Their performance in Study 2 has confirmed that: when producing the critical arguments would not benefit their success in the game, as in the competitive condition, seven-year-olds withheld these arguments. Although five-year-olds produced more self-serving arguments in Study 1, this was based entirely on their use of affirmations. When producing affirmations was harder, as in Study 2, because no trained arguments would help them justify animals next to unusual items, they were not as strategic as seven-year-olds. Therefore, compared to 7-year-olds, five-year-olds relied more on simpler arguments, namely affirmations, which justified a desired proposal/outcome (why an item should be placed somewhere). seven-year-olds, in contrast, additionally relied on refutations, which are cognitively more complex than affirmations because refutations disprove a possible undesired proposal. This shows that seven-year-olds were paying attention to not only what evidence benefits their goals (affirmations) but also what evidence might conflict with their peers’ goals (refutations).

Children’s use of what we termed refutations shows an interesting contrast with what Felton and Kuhn (2001) and Kuhn and Udell (2003) observed with school-age chil-
children. In these studies, children did not frequently make use of advanced argumentative strategies such as counterargument and rebuttal, i.e. steps that anticipate and invalidate argumentative moves that their peer might take. Children rather used such strategies to a greater extent only after explicit training/instructions to produce refutations (Kuhn & Udell, 2003). In Study 1, children of both age groups, five-year-olds to a lesser extent, used refutations. They explained to their peers why an animal should not go to their peer’s side, thereby weakening their peer’s position. Therefore, our studies demonstrate how young children are able to use complex arguments such as refutations at an earlier age than previously shown in the literature.

One limitation of our study is whether the decisions were equally “joint” across conditions. In the strongest sense, joint decisions encompass those cases in which both participants explicitly acknowledge the decision. In the weakest sense, they also encompass cases in which an animal was placed somewhere, without attempts to undo this. In our data, children often reached joint decisions after deliberation but without verbally marking their consent to the decision in both conditions. Thus, we opted to use the weak definition because limiting the joint decisions to verbal acknowledgements would not represent children’s interactions accurately.

To conclude, our results suggest that by age seven, children are able to produce sophisticated arguments such as refutations and adapt them strategically to different discursive goals in their peer interactions. When the context is intuitive and gives children more opportunities to produce arguments, as in Study 1, even five-year-olds could produce refutations. As both studies point out, cooperative situations, in which children have joint goals, provide a more motivating context for them to produce arguments and to mention more aspects of the question at hand.
Chapter Four
Cooperative Reasoning in the Context of Group Competition

1 Introduction

Coordination plays an important role in explaining humans’ evolutionary success (Tomasello, 2014). Coordinated behaviour in novel situations relies on reasoning—flexible communication of intentions, beliefs and the reasons by which these are motivated. Coordination opens up new resources in the face of scarcity, providing a group-level benefit to those that coordinate effectively. Mercier and Sperber (2011, 2017) similarly argue that reasoning as part of communication improves group-level outcomes, putting, however, emphasis on individual motivations that benefit the group on aggregate. On their account, reasoning functions to stabilise communication overall by enabling individuals to counter deception. Regardless of the specific theoretical approach, reasoning as a social phenomenon is hypothesised to confer a benefit on the group level. If reasoning has an adaptive effect for groups, does it follow that it is sensitive to group-level competition? Does competition between groups motivate within-group collaborative reasoning?

Competition between groups has been found to increase motivation for in-group cooperation. Studies with adults have focused on this phenomenon in public goods games (Cárdenas & Mantilla, 2015; M. R. Jordan, Jordan, & Rand, 2017), problem solving tasks (Mulvey & Ribbens, 1999), and physical work (Erev, Bornstein, & Galili, 1993). Cárdenas and Mantilla (2015) review and present evidence that in public goods games individuals contribute more when their own group competes with others. In this type of economic game, individuals in a group are given an amount of money. They need to decide how much of it they contribute to the “common good”, which then is multiplied by a certain factor and distributed equally to participants. If all contribute much, everyone
profits, but if contributions are unequal, some participants take a loss, while others free-
ride. A group competition modification adds that groups are told how high the contribu-
tions in another group are, and that the group that contributes most receives an addi-
tional bonus.

Jordan et al. (2017) narrow down the likely source of a heightened motivation for
the group’s common good to a non-social threshold effect. They show that given a
threshold above which rewards are given—in group competition, this is information
about the other group’s contributions—, group members contribute more. Mulvey and
Ribbens (1999) had adult teams of three build Lego structures, in which participants had
to coordinate individual substructures, and measured their productivity. They compared
the effect of fixed production goals (e.g., the goal to build ten structures) to those of
group competition and to a control condition that had neither of these incentives. Group
competition enhanced group productivity, but similar to findings by Jordan et al. (2017),
so did fixed goals. In a physical task, which involved paid labour on an orange plantation,
Erev et al. (1993) compared reward conditions for small groups picking oranges. Groups
of four participants each worked either under individual payment, under collective pay-
ment, with all group members receiving equal parts of the group’s pay, or under group
competition, with equal pay and rewards for the best group. Relative to individual pay-
ment, participants under collective payment showed freeriding, i.e., lower effort and
greater reliance on their group’s performance. Under group competition, this freeriding
was offset entirely and group productivity increased over time. Overall, out-group com-
petition appears to increase motivation to contribute to the group, and an important
mechanism is the introduction of thresholds towards which group members orient them-
selves.

An important aspect besides the question of how group-directed motivation affects
cooperation is what factors influence group reasoning for joint decision-making. In a
seminal study with adults, Stasser and Titus (1985) operationalised the quality of group
discussions as the amount of unshared information that participants contribute to the
discussion. In this hidden profile paradigm, all participants initially share some informa-
tion for a joint decision pointing to a suboptimal solution. However, each participant also individually holds unshared information that, if everyone shared it, would identify the best solution. This paradigm demonstrates the tenacity with which groups that nominally hold all necessary information can fail to put it together. It has motivated numerous research efforts aiming at better exchange of information. Wittenbaum, Hollingshead, and Botero (2004) review this literature, summarising that the possibilities of influencing discussion quality are modest. For instance, outcomes can be improved somewhat by establishing pre-discussion disagreement or expert roles. Outcomes can also be influenced by how the information is initially distributed among participants. However, external motivation such as emphasising the importance of the task does not improve results. In one study that had a condition with group-level competition, groups’ performance improved (van Hiel & Schittekatte, 1998) if they were mixed gender. However, this result has since been awaiting replication. In sum, adult group discussion is characterised by a robust tendency to attend to information that participants already share, instead of extending it critically with new information.

Group discussions over joint decisions among children have been investigated only rather recently with a focus on relevant aspects for the present study. For instance, a study by Gummerum, Leman, and Hollins (2014) investigates an adaptation of the hidden profile paradigm for 7- and 9-year-old children. Children in these age groups show different levels of skill in being interactive discourse partners. This interactivity is expressed in the extent to which peers engage with each other’s contributions and express different perspectives towards a belief. Surprisingly at first blush, younger children achieved better epistemic outcomes than older ones, i.e., sharing more of their individually held information in the group discussion. This is likely the effect of younger children being less interactively engaged and more focused on recalling their individual information.

Children’s developing competence for dealing with common ground and communicative goals has been investigated in a series of studies. When 5-year-old children give reasons to peers in joint decisions, they adapt the informativeness of these reasons to their common ground. They elaborate their reasons more when knowing that their peers lack
vital knowledge to understand them, or when their reasons support unusual decisions (Köymen et al., 2016, 2014). Also, 5- and 7-year-olds adapt their reason giving to cooperative and competitive contexts. When they cooperate with each other, given the same set of decisions, they produce a greater number and more balanced sample of reasons than when they compete (Domberg et al., 2018).

In this context, an interesting question to investigate is whether group-level competition also has an influence on children’s reasoning. The present study focuses on this question, and on how this impact may look at different developmental stages. There is evidence for an early development of group-related attitudes and preferences. Dunham, Baron, and Carey (2011) show that children aged five are sensitive to the in-group–out-group distinction and show in-group preference on explicit and implicit attitude measures. Among others, children showed greater explicit liking for in-groups and distributed a resource to the advantage of an in-group and disadvantage of an out-group. Children made these distinctions in absence of any information other than group affiliation. Thus, the group context becomes relevant in preschool years, and has a significant motivational effect.

In the present experiment, we gave pairs of 5- and 7-year-old children a cooperative problem solving task, in which they needed to improve a picture together as a group by deciding what elements to add to the picture. In one condition, the “lazy” condition, they were instructed to try and achieve a good result together. The name of this condition relates to our expectation to find less effort and reasoning in the absence of group competition. In the other condition, the competitive condition, they were told that they would win if they reached a better result than a pair of competitors marked as out-group. We hypothesised that the competitive condition would motivate children to spend greater effort on the task, to be more critical about their joint response and discuss it more, but especially to produce more reasons than in the lazy condition. We selected these age groups, on the one hand, because we judged the task to be lighter on incidental demands than the one used by Gummerum et al. (2014). Items were always available and no memorising was required. On the other hand, 5- and 7-year-olds have been observed
to be sensitive to cooperative and competitive goals (Domberg et al., 2018) and to successfully adapt their production of reasons to the context.

2 Method

2.1 Participants

Sixty-four 5-year-olds ($M = 5;9$, $Range = 5;6-5;11$, 32 girls) and 64 7-year-olds ($M = 7;6$, $Range = 7;0-7;11$, 32 girls) in 64 same-age and same-sex dyads participated in this study, carried out in their kindergartens and schools. Each dyad participated in both conditions. Dyads were from the same institution, and from the same or neighbouring groups. All children were native German speakers with various socio-economic backgrounds.

2.2 Materials

The central materials used for each phase of the game (warm-up and two trials), shown in Figure 4.1, were

- a background picture (see Figure B.1 and Figure B.2 in Appendix B),
- a set of tokens each of which had a picture of an item or animal,
- two blinders,
- two pairs of trays,
- a box,
- a still image on a laptop screen of two children facing the camera and the background of the current trial; gender and background were matched to the dyad and trial, with tokens greyed out; the laptop was on the table and closed at the beginning.

The background picture for warm-up was of a picnic basket with five marks, on which to put tokens. For the experimental condition, it was either a zoo map or a dollhouse with seven marks each. In each condition, children received each a set of tokens, which are depicted in Figure 4.2. The “default” token set (see labels in the Figure) was in
place on the image at the start, and each child received one adequate and inadequate set mixed together.

2.3 Procedure

**Warm-up.** The purpose of the warm-up trial was to get children acquainted with the task. Upon entering the testing room, the experimenter (E) sat the children down at a table facing each other. E introduced the topic of having a picnic, encouraging children to name items that are necessary for it. E distributed the picnic-themed tokens (yellow in Figure 4.2) among children, instructing that each had to select five that are most useful at a picnic. The rationale for this prior individual selection was to give children motivation to exchange tokens later in the testing phase, because their own tokens were a selection they had made. When each child was done selecting five and trashing the other tokens (putting them into box D in Figure 4.1), E presented the picnic background with the five “default” tokens already placed and commented that some of them might be okay, but
Method

some would need replacing. In a semi-structured interaction, E enforced the order of asking children (1) which token they would like to replace and why, (2) which of their selected tokens they would like to place instead, and (3) why. E then asked the other child whether the reasons given were okay. This order was kept up until both children signalled that they were satisfied with the result, upon which E praised children for the good work, removed the background and asked children to also remove the remaining tokens.

Testing trials. Each dyad participated in both the competitive condition and the lazy condition. There were two games (material sets): the zoo game and the doll-house game. The procedure was similar to the warm-up game. The order of the games and the condition assigned to each game was counterbalanced between dyads. The following procedure description for the testing trials is based on the zoo game in the lazy condition first, and the doll-house game in the competitive condition second.

E introduced the zoo topic, asking: “What do you know about zoos? What kinds of animals are there?” Children were given 14 tokens each, of which they were to select seven

---

**Figure 4.2: Picnic, zoo and doll-house tokens.**
The tokens as used in this study. In each trial, children received a mixed set of adequate and inadequate items from which to pick seven and trash seven. Each background picture had the items labelled “default” on it at trial start. “Adequate” items were arranged such that they would provoke the problem (and potential reason) of duplicity between children.
to put on their trays (see Figure 4.2); at this point, children had blinds between them (see Figure 4.1) to keep individual selections independent. When children signalled that they were done, E removed the blinds and put the selected tokens on one heap. E presented the zoo background with “default” tokens, commenting that some might be good and some worth replacing. Upon that, E gave the task instructions:

“Put a really nice zoo together and think about how many of these tokens you want to replace. But make sure you always give a good reason for why you want to replace them.”

E opened the laptop, giving condition instructions, which were specific to the condition. In the lazy condition, E said:

“Let me see if you are playing alone or against someone else. No, this round you’ll be playing alone, there’s no-one else.”

E shut the laptop and repeated the task instructions, adding that children should call E back in once they were done, and left the room. Upon re-entering, E removed the materials, without commenting on the result. Then, they moved on to the next condition. The procedure was again parallel, now with the doll-house material set.

E introduced the doll-house topic, asking: “What do you know about doll-houses? What things belong into them?” Children again individually selected seven out of 14 tokens each with blinds between them. When children signalled that they were done, E removed the blinds and put the selected tokens on one heap. E presented the doll-house background with “default” tokens, commenting that some might be good and some worth replacing. Upon that, E gave the task instructions:

“Put a really nice doll-house together and think about how many of these tokens you want to replace. But make sure you always give a good reason for why you want to replace them.”

In the competitive condition, E opened the laptop, giving the condition instructions specific to the competitive condition:
“Let me see if you are playing alone or against someone else. Oh indeed, here’s the red team. They are already done playing the doll-house game. Now back to your task.”

E repeated the task instructions, adding,

“And if you wind up having a nicer doll-house than the red team, then you win against Team Red!”

Upon completion of this task, children were announced winners of the competition.

2.4 Coding

Children’s conversations were video-recorded and transcribed verbatim. Each line of the transcript contains one proposition. We coded for each line whether it constitutes a reason, or expresses the speaker’s attitude or normative belief about acting on a token (see below for utterance categories). If one of these was the case, we coded about which token it is, and if it favours having the token in the picture or out. See Table 4.1 for examples.

**Utterance categories.** We categorised utterances into three main types: reasons, expressions of attitude, and unrelated, including repetitions.

A reason is an utterance that points out a premise entailing that a token should be removed, added, or kept in or out. An example is a token’s adequacy for the zoo or doll-house. This could be expressed by how the depicted item is useful, or that the animal depicted generally belongs into a zoo, using generic language. Since the task was also aesthetic in nature, assertions that token X is prettier (funnier, more interesting/useful, etc.) than token Y counted as reasons. These refer to the implicit premise “select aesthetically superior tokens”.

A second example for a reason, i.e., a premise that entails action on a token, is “We have got this one already”. This utterance points out a fact, which together with an implicit second premise (“avoid duplicate tokens”) entails keeping out the token in question.
Table 4.1: Example arguments.

Examples from the data set of reasons and their role as elaborations or refutations of premises or conclusions. In a), B expands on A’s expression of attitude. In example b), the second utterance refutes child A’s conclusion, while the fourth utterance refers to a premise of B’s implicit conclusion to keep the second lions item out, namely its being identical. c) shows a change in attitude in A.

<table>
<thead>
<tr>
<th>Child A</th>
<th>Child B</th>
<th>Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>perhaps a wardrobe</td>
<td>yes good idea</td>
<td>elaborates conclusion</td>
</tr>
<tr>
<td>vielleicht einen Schrank</td>
<td>Ja gute Idee</td>
<td>add wardrobe</td>
</tr>
<tr>
<td>because you can put things into it</td>
<td>weil man da was reinton kann</td>
<td>reason</td>
</tr>
<tr>
<td></td>
<td></td>
<td>add wardrobe</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Child A</th>
<th>Child B</th>
<th>Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>lions yes we want lions</td>
<td>um we already have lions</td>
<td>refutes conclusion</td>
</tr>
<tr>
<td>Löwen ja wir möchten Löwen</td>
<td>ab wir haben schon Löwen</td>
<td>keep lions out</td>
</tr>
<tr>
<td>but not the same lions</td>
<td>lions twice</td>
<td>refutes premise</td>
</tr>
<tr>
<td>aber nicht die gleichen Löwen</td>
<td>zweimal Löwen</td>
<td>keep lions out</td>
</tr>
<tr>
<td></td>
<td></td>
<td>add lions</td>
</tr>
<tr>
<td>I think the dishes are not so nice</td>
<td>but still</td>
<td>refutes conclusion</td>
</tr>
<tr>
<td>das Geschirr find ich nicht so schön</td>
<td>aber trotzdem</td>
<td>keep dishes out</td>
</tr>
<tr>
<td></td>
<td></td>
<td>add dishes</td>
</tr>
<tr>
<td>c)</td>
<td></td>
<td>changes mind</td>
</tr>
<tr>
<td>if one wants to eat, one needs dishes</td>
<td>wenn man was essen will, da braucht man Geschirr</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>add dishes</td>
</tr>
<tr>
<td>right</td>
<td>stimmt</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>add dishes</td>
</tr>
</tbody>
</table>

As can be seen in Figure 4.2, there were several instances of close-identical tokens for children to point out this reason.

Expressions of attitude, like reasons, point towards the conclusion to put or keep tokens in or out. However, they do not adduce premises that entail this conclusion. This category also applies where utterances rely on normative language, but do not include the respective norm or another relevant premise to support this conclusion. Examples include assertions that a token is “good” or “better than” another one.

For both reasons and expressions of attitude, we coded what token they were about and whether they entailed that the token should ultimately be in or out of the picture. With respect to reasons, we additionally coded whether instead of supporting an action
directly, they targeted a previously used premise (see (b) in Table 4.1 for an example). Reasons that do this differ from those that concern a token-related action directly by introducing a level of embedding or intersubjectivity to the exchange.

Utterances that did not fulfil the criteria for the categories of reason or expression of attitude were not analysed further. Neither were repetitions of reasons for the same token, which in most cases resulted from children’s excitement, instead of any deliberate use of repetition.

**Contextual coding: elaborations and refutations.** So far, the manual coding outlined above identifies and assigns codes for reasons and expressions of attitude. Codes for both categories have as a property the token that they are about and whether the token should end up in or out of the picture. Codes for reasons additionally have the property of whether they focus on a prior premise or are more generally about acting on a token.

We added a context-related property to the codes of reasons by marking whether they were in support of a prior peer’s utterance (elaborations) or against it (refutations, see rightmost column, Table 4.1). For reasons that focused on a prior premise, the target utterance was already coded manually, so the elaboration/refutation code followed straightforwardly from whether both reasons agreed or disagreed. For reasons that did not focus on a prior premise, we automatically searched the preceding dialogue\(^3\) for reasons or expressions of attitude by the peer about the same token. On the basis of whether the prior utterance and the reason in question agreed or disagreed about how to act on the token (i.e., having it end up in vs. outside the picture), this latter reason was coded as elaboration or refutation. In sum, the two essential and independent properties of reasons were, firstly, if they were elaborations or refutations of prior utterances (or neither), and if they focused on premises or conclusions.

A final code was used for those cases in which a refutation led the refuted speaker to change their mind. This was established automatically as well by matching the pattern\(^3\) The search scope was the shorter of ten utterances or 15 seconds before the given reason. If there were more than one peer utterances about the same token in this span, the closest to the given reason was selected.
that speaker A advocates for some action (reason or expression of attitude), B gives a reason against it and after that, A also expresses their attitude or another reason in accordance with B. With these codes attached to each reason, we conducted the quantitative analyses below, which fall into two categories, confirmatory and exploratory. Confirmatory analyses address the hypotheses that we brought to the experiment, concerning the number of reasons that dyads produce across conditions and age groups. We expected older children to be more productive than younger children, and the competitive condition to yield more reasons than the lazy condition. We used exploratory analyses to investigate further relationships among variables.

3 Results

3.1 Manipulation check

To ensure children were aware of whether they were competing as a group against another group in the competitive condition, we checked for the presence of signs that they did perceive this aspect. Children counted as acknowledging the manipulation if they fulfilled one of three conditions. Either they talked about the red team during the trial; or they expressed group motivation when the red team was introduced, e.g., gesturing and displaying a competitive attitude towards the laptop; or they showed interest in the laptop during the trial. Among 5-year-olds, 24 out of 32 dyads did one or more of these, and 28 out of 32 7-year-old dyads did (see also the dot colour coding in Figure 4.3).

3.2 Confirmatory analyses

There were a total of 6,157 utterances. Of these, 2,431 were not classified further, because they did not express attitudes or reasons, or were repetitions of reasons. Among the remaining utterances, we counted a total of 753 reasons.

To test the main hypothesis that children’s production of reasons varies depending on whether they perceive their group in a competition with another, we analysed the number of reasons produced per trial across conditions and age groups. We used a Gener-
alised Linear Mixed Model (GLMM) with a Poisson error distribution using the \texttt{lme4} package in \textit{R} (Bates, Mächler, Bolker, & Walker, 2015). We used age group, condition and their interaction as predictors, with the game material (zoo or doll-house), trial number (first or second) and gender as control variables, with the random effect of dyad and an offset term of number of overall utterances to account for children’s talkativeness.

\textit{full model}: \hspace{2cm} \texttt{reasons \sim age group * condition + material + trial + gender + (1|dyad) + offset(total utterances)}

\textit{null model}: \hspace{2cm} \texttt{reasons \sim material + trial + gender + (1|dyad) + offset(total utterances)}

The full model shows improved fit over the null model ($\chi^2 = 15.62, df = 3, p = .0014$).

To test whether the interaction of age and condition was significant, we compared the full model to a reduced model without the interaction term:

\textit{reduced model}: \hspace{2cm} \texttt{reasons \sim age group + condition + material + trial + gender + (1|dyad) + offset(total utterances)}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure4_3.png}
\caption{Number of overall arguments produced per trial. Each dot represents a trial, so each dyad is represented by two dots, one in the competitive, one in the lazy condition. In the competitive condition, green dyads showed signs of noticing the competitive situation in the manipulation check, yellow ones did not. The numbers above each cell are group means and standard deviations.}
\end{figure}
The interaction term was not significant ($\chi^2 = .01, df = 1, p = .92$), so we dropped the full model from further comparisons. To test the influence of age as a predictor, we compared the reduced model to a model without the predictor of age group:

\[
\text{without age: } \quad \text{reasons} \sim \text{condition} + \text{material} + \text{trial} + \text{gender} + (1|\text{dyad}) + \text{offset(\text{total utterances})}
\]

This comparison suggests a significant main effect of age ($\chi^2 = 10.35, df = 1, p = .0013$). Seven-year-old dyads produced more reasons than 5-year-old dyads. To test the effect of outgroup competition on dyads’ reason-giving, we compared the reduced model to a model without the predictor of condition:

\[
\text{without condition: } \quad \text{reasons} \sim \text{age group} + \text{material} + \text{trial} + \text{gender} + (1|\text{dyad}) + \text{offset(\text{total utterances})}
\]

This comparison suggests a significant main effect of condition ($\chi^2 = 5.14, df = 1, p = .023$). How many reasons dyads produced differed with regard to whether they were in competition with an out-group. Against our hypothesis, inspection of the reduced model and the descriptive data suggests that dyads produced more reasons in the lazy condition (see Figure 4.3).

We considered the possibility that children’s production of reasons was to some extent in response to the instruction to give reasons for every decision. We expected the production of reasons that were not in response to a peer’s utterance to be less influenced by the manipulation. Therefore, we next analysed only those reasons that were in response to some preceding utterance by the partner, i.e., elaborating or refuting a preceding reason or expression of attitude. We term this subset of reasons interactive reasons (see Figure B.3 in the Appendix B for a plot of these), of which there were a total of 315. The comparison of the full and null model based on this subset of reasons proceeded analogously to the full-null comparison of the complete dataset:
Results

full model: interactive reasons ~ age group * condition +
material + trial + gender +
(1|dyad) + offset(total utterances)

null model: interactive reasons ~ material + trial + gender +
(1|dyad) + offset(total utterances)

This comparison suggests that the full model does not significantly improve fit to
the data compared to the null model ($\chi^2 = 4.10$, $df = 3$, $p = .25$). The number of reasons
produced interactively in response to peers’ utterances did not vary significantly with age
or condition.

3.3 Exploratory analyses

We conducted further analyses to explore additional effects that age group or condi-
tion may have had on the nature of children’s production of reasons. Particularly, we
looked at

– children’s tendency to elaborate vs. refute their peers’ contributions depend-
ing on condition and age;
– children’s tendency to respond to their peers’ conclusions vs. to focus on their
premises depending on condition and age;
– children’s tendency to change their mind upon hearing a reason that refutes
them, depending on condition and age;
– the duration of the trial depending on condition and age.

We tested whether children more strongly relied on elaborating their peers’ contrib-
utions or on refuting them depending on age group and condition by fitting a GLMM
with a binomial error distribution. The outcome variable was the response to the ques-
tion of how many interactive reasons produced in each trial were elaborations versus ref-
utations of preceding utterances.\(^4\) By definition, this analysis can only consider interact-
ive reasons.

\(^4\) The outcome variable was realised in R as a `cbind` variable, i.e., two columns, one being the number of elaborations in each trial, the other being the number of refutations.
The comparison between the full model and the null model was non-significant ($\chi^2 = 3.46, df = 3, p = .33$), suggesting no effect of age or condition on this outcome variable. This means that neither age nor out-group competition had a significant influence on whether children rather agreed with and extended each other, or disagreed and criticised each other.

To test whether children focused more on the premises of their peers’ reasons or on their conclusions about how to act on a token, we conducted the same comparison of a full and a null binomial GLMM. In this analysis, the outcome variable was how many of the interactive reasons in each trial were about premises versus conclusions. A significant effect on this variable would mean that age or out-group competition influence how children engage each other’s reasoning; whether they critique more deeply the premises that underlie each other’s reasons, or whether they focus more directly on the outcome.

This comparison was also non-significant ($\chi^2 = .76, df = 3, p = .86$). It suggests that out-group competition and age do not affect this specific aspect of children’s reasoning.

For the analysis of occasions on which a speaker changed their mind, we compared the frequency of this behaviour across age groups and conditions, using a Poisson GLMM and again the same predictors and control variables:
3 Results

**full model:** mind changes ~ age group * condition + material + trial + gender + (1|dyad) + offset(total utterances)

**null model:** mind changes ~ material + trial + gender + (1|dyad) + offset(total utterances)

This model comparison turned out non-significant ($\chi^2 = 1.28$, $df = 3$, $p = .73$), suggesting that age and out-group competition had no significant influence on how likely children were to be swayed by their peers’ reasons.

For the analysis of trial duration depending on age group and condition, we fit a Linear Mixed-Effects Model, with Gaussian error distribution (see Figure B.4 in the Appendix for a plot of these data). The response variable was the logarithm of the trial duration, the predictors were age group, condition and their interaction, and control variables were material, trial number and gender, along with the random variable of dyad:

**full model:** log(duration) ~ age group * condition + material + trial + gender + (1|dyad)

**null model:** log(duration) ~ material + trial + gender + (1|dyad)

This comparison turned out non-significant ($\chi^2 = 2.29$, $df = 3$, $p = .52$), suggesting that trial duration was not significantly influenced by children’s age or presence of out-group competitors.

4 Discussion

The purpose of this study has been to investigate how 5- and 7-year-old children’s collaborative reasoning varies when they do versus do not compete against an out-group. We found a main effect of age, with older children producing more reasons than younger children. We also found a main effect of condition, with children producing fewer reasons when facing out-group competition. This result points in the opposite direction of our initial hypothesis. It suggests the possibility that the introduction of a competing team on the laptop had negative effects on children’s reasoning, rather than increasing motivation. Neither main effect, however, persisted when focusing only on interactive
reasons, which engage a peers’ reasoning and were thought to reflect effects from the manipulation more directly. The results do thus not back our hypothesis that out-group competition leads to an increased production of reasons within a group of cooperators.

There are several possible explanations for the observation that out-group competition did not improve group reasoning. Most superficially, children in the competitive condition may not have noticed that they were competing against another team. Or similarly, children may have gotten disinterested in the laptop and forgotten about the group that they were competing against. The manipulation check does, however, suggest that children were indeed aware of their competitors, commenting about them and showing negative affect towards them during the instructions and/or during the testing phase. Also, if children’s less rich reasoning in the competitive condition was due to distraction by the presence of the out-group on the laptop, this would also speak against ignorance about the manipulation.

Secondly, children may not have recognised that quality-controlling their solution or a longer and more effortful discussion was a good response to competition; or they may have needed quantitative information about the necessary amount of effort. In the adult literature in behavioural economics discussed above, concrete thresholds have been shown to work similarly to, or underlie, increased contributions under group competition (M. R. Jordan et al., 2017; Mulvey & Ribbens, 1999). In the present study, this could have taken the form of informing children that the red team discussed a certain number of possibilities for their good result. On the one hand, this would have been a more concrete cue for children to make a greater effort to improve their decisions. On the other hand, this would mean that part of the task of responding to group competition, namely recognising the need to increase effort, would already be done for children. Concluding together that greater care, e.g., a longer search for solutions, may be necessary, was in fact part of the task.

A third possibility is that there is no straightforward response in collaborative reasoning to looming competition or other external motivators. For comparison, children in Study 1 of Domberg et al. (2018) (Chapter Two) produced significantly different amounts
of reasons between a cooperatively and competitively motivated condition. There were two factors in the competitive condition of that study that could lead to peer-serving reasons not being produced. Children may have ignored their peer’s part of the zoo altogether, or, having a reason in mind, they may have withheld it, recognising that it would decrease their own chances of winning. In the present study, however, there is no obvious point in the peer discussion at which children have an incentive to opt against uttering a reason that they think of. The only intuitive step towards varying the number of reasons produced would be to lengthen the discussion. Children did not do this, and it is not clear that this would even be effective, as an adult group decision study suggests. Investigating the role of importance of a discussion, Larson et al. (1994) conducted a group decision-making study under the hidden-profile paradigm, manipulating the perceived importance of the groups’ decisions. The authors did observe longer discussions on average when participants were told that their decision mattered in further administrative developments in their psychology department. In comparison, discussions were shorter when participants were told that their decision was not important. But importantly, high importance did not lead to a greater exchange of unshared information. One interpretation of this would be that participants knew as a rule or custom how to respond to heightened importance in a group decision—by prolonging their discussion. This knowledge of behavioural expectations towards decision-making groups could be what reasoners acquire over time, whereas the ultimate identification and production of reasons would remain largely unaffected by this awareness.

In summary, we did not find an increase in reasons produced in joint decision-making under competition. This is a null result, which needs to be interpreted with caution. To investigate some of the points discussed above, one could test a more salient realisation of the competitive situation. This could involve a reward as a more straightforward motive, and a video or even live presentation of competitors to signal the manipulation more strongly. Alternatively, children may need a tangible, even quantitative, picture of how much effort they have invested, and how much is necessary to beat the other team. This would be along the lines of investigating the non-social threshold hypothesis of
group-competitive motivation brought up in adult economic games (M. R. Jordan et al., 2017; Mulvey & Ribbens, 1999). If, however, these manipulations do bring about behavioural variation, the key question remains whether the change is better reasoning, or, as with adults in Larson et al. (1994), only prolonged discussions.
Chapter Five
Summary and general discussion

Using reasons to persuade interlocutors is the main function of reasoning—the use that it evolved for and for which it is best suited (Mercier & Sperber, 2011, 2017; Norman, 2016; Tomasello, 2014). I have reviewed evidence that supports an early competence for reasoning under this social account, and reported experiments that extend this view. For each of these experiments, I summarise briefly what significance its results have for the study of reasoning and its development, and discuss some of the issues that it raises, including limitations and possible follow-ups. Upon that, I discuss some of the central predictions of the two social theories of reasoning in some more detail, and close with a general outlook.

1 Identification of reasonable partners
(Chapter Two)

1.1 Summary of this study

Chapter Two reports an experiment that required children to embed their judgement of reasons into a partner choice task. Five- and 7-year-old children were presented with two candidates with one of whom they would solve a cooperative problem together. In the experimental condition, one candidate acceded to a good reason, the other to a poor reason. In the control condition, each agreed to a different good reason. Crucially, in both conditions, neither could be construed as knowing better. The results suggest that 7-year-olds, and 5-year-olds to a lesser degree, preferred the partner who followed the good reason in the experimental condition, but they showed no preference in the control condition. Thus, children prefer partners who respond to good reasons, independently of partners’ incidental accuracy in the model situation.
1.2 Issues raised

**Judging reasons as a flexible skill.** The present study shows that around school entry, children use their judgement of reasons beyond the direct evaluation of informants. Children are able to use this evaluation to ascribe to potential partners a trait such as reasonableness, and to prefer those who have it. There is room for further exploring the development of mastery in judging reasons.

On the one hand, the decision for which the judgement of reasons is used could vary. For instance, children could apply their judgement in a task in which they do not choose a partner, but a group. One group would display relative frugality in justifying a given decision, i.e., accept a poor reason, whereas another group would show a more critical attitude to the same decision. Children would thus judge the candidate groups’ epistemic standards and what standards they rather want to accept and share.

On the other hand, an interesting parameter is how the two reasons under consideration vary. In the present study, the contrast between both reasons is quite strong, with the poor reason being utterly inappropriate in several respects. Future experiments could replace this stark contrast with more subtle problems. For instance, instead of appeal to one’s favourite colour, the poor reason could be deficient by resting on a condition that has been previously shown not to hold: Out of two tools that an actor has available to clean up the floor, they pick the vacuum cleaner instead of the broom. However, it has been established beforehand that there is no power outlet near. Candidates to partner choice now endorse the proponents of each option. More generally, this paradigm offers the possibility to test children’s evaluation of reasons and explanations in a whole range of contexts without asking them explicitly about the quality of reasons. Not asking children such unusual questions to test their reasoning is certainly advantageous.

**Argument strength and informant accuracy.** In the present study, children’s choice of partners obtains meaning through the fact that the strength of the reasons is separated from the possibility that any of the candidates is objectively right. Children’s preference for a partner honours his judgement, not his better knowledge. Without this
measure, it would not be clear what ultimately motivated their choice—valuing partners who do not commit reasoning errors or concluding that one of them is better informed. As argued in Chapter Two, this is an improvement over the study by Koenig (2012), but also a point to observe in all studies to come that involve judging how characters deal with reasons. Future studies that test children’s trust based on the evaluation of reasons would need to make sure they exclude the possibility that children over-ascribe knowledge.

1.3 Limitations

In this study, 26% of 5-year-olds were not included in the analysis because they failed in the prior false belief probes. That is, at some point during the narration, these children ascribed to one of the characters knowledge about the lost brush that he could not have. This has an impact on how the result should be interpreted. Although a significant majority of 5-year-olds who passed these probes also preferred the partner who followed good reasons, sweeping conclusions about all 5-year-olds’ ability to evaluate someone else’s reasoning are not warranted. To generalise to the whole sample including the children who failed in the false belief probes, the answer is more graded: Some children would have chosen the wrong partner, a relative majority the right partner, and 26% of children would have made a choice based on the belief that one candidate is correct.

2 Strategic use of reasons (Chapter Three)

2.1 Summary of this study

The study reported in Chapter Three gave 5- and 7-year-old children the possibility to determine the outcome of a joint decision by providing or withholding reasons. After a training, in which one child learned three convincing reasons, they could exert influence in joint decisions with a naïve peer by deciding how much information they disclose. The manipulation between conditions, by introducing a cooperative or competitive rule set, gave trained children varying motivation to share their trained reasons. Seven-year-olds
responded measurably to this variation, reproducing reasons learned in the training phase more often when cooperating with their peer, but less so when they were in competition. While the evidence from the foregoing study could not exclusively be explained by older children strategically withholding reasons under competition, the present experiment showed this behaviour much more directly.

2.2 Issues raised

**Motivation and reasoning.** A question often discussed in the context of bias in using evidence is whether reasoners are aware of it. In the literature about confirmation bias (e.g., Nickerson, 1998), this is an important distinction made theoretically, but known to be hard to investigate empirically. Confirmation bias can be a spontaneous preference to look for evidence supporting a belief under consideration. Or it can be deliberate case-building, including awareness of evidence against a preferred belief, and acting to discount or withhold such evidence.

In the prior study, in which children used entirely their own reasons, we cannot pin down their bias as spontaneous or deliberate. There is always residual uncertainty about children’s degree of awareness of evidence against their motives. Also, conclusions are always hard to make from not observing a behaviour, such as producing reasons. But where children explicitly learned some critical reasons, their bias results much more likely from deliberate case-building. Only 7-year-olds showed this type of strategic behaviour. Observing this distinction in development has been a key motivation of this study, and as a future direction, the role of motivation in reasoning should receive more attention (see Section Fehler: Verweis nicht gefunden below).

An analysis by Kunda (1990) of the role of motivation in reasoning points in this direction. When pursuing their goals in reasoning, whether it is to find the best justified result or the best justification for a pre-established result, reasoners are constrained by the available evidence. The very essence of reasoning is, after all, that it is not as arbitrary as making and imposing unsubstantiated claims. Motivated reasoners consequently tend to
direct their attention and strategies to have an influence on the availability of evidence to them and others, in line with their given motivation.

An interesting question is thus, whether children also apply strategies that work towards the goals that they pursue in reasoning. For instance, in an experiment that involves reasoning for joint decision-making, children could be provided, as in the present one, with evidence for a set of reasons for use in their reasoning exchange. As the key modification in this experiment, children would be given the possibility to manipulate the visibility of evidence. The research question would then be at what age children apply strategic behaviour so far as to “curate” the evidence that underlies reasoning for joint decisions.

2.3 Limitations

**Productiveness of reasons in younger children.** Trained children in this study learned three critical reasons so that they could choose to reproduce or withhold them depending on the manipulation. While older children in the cooperative condition on average reproduced 2.56 out of 3 reasons, younger children only reproduced 1.56 on average. The absence of an effect in younger children should therefore be put into perspective. It is not clear whether the younger children would have also lacked a condition effect if their overall production of reasons in the cooperative condition had been stronger. Testing 5-year-olds would therefore certainly profit from a more sensitive measure. Alternatively or additionally, independent measures such as working memory or intelligence could help ascertain to what extent they were challenged by incidental task difficulty. As a whole, however, this age group did not produce an effect that older children did.

**Significance beyond the dyad.** The unit of analysis in this study has been the dyad. Also, the primary interest of this study was in conversation, not in the decisions (the animal placements) ultimately made by children. For researchers interested in group decision processes, this means measuring a secondary variable in a special case scenario. What can this study tell us about larger groups and other influences? As soon as the unit of observation becomes larger than two, complexity increases manifold. If only two en-
gage in social reasoning, either one drops a position that she has been shown violates a
shared commitment, or she stands her ground and motivates a lasting schism. But more
members mean more options for proceeding upon a disagreement, to do with making
and breaking alliances in the group. In the study mentioned earlier by Moshman and
Geil (1998), group dynamics were presumably not too turbulent because the shared-goal
condition held: All should have agreed about the task, and that no extra benefit was ex-
pected for the individual author of the group’s solution. What little loyalty to their own
prior position those participants may have had who did not find the solution of the prob-
lem individually, they likely dropped immediately upon hearing out the one peer with
the principled—and thus convincing—solution.

But with the appearance of goal divergences among members, subgroups arise.
These subgroups will have non-overlapping goals. If their goals differ, their shared com-
mitments do as well. And if that is the case, individual exchanges can easily end in dis-
agreement and changing subgroup loyalties, followed by more group-dynamic phenom-
ena such as groupthink and polarisation. Note that the decisive factor is not the number
of individuals, but the motivational alignment among them. If someone presents reasons
for a decision to the wider public, these may work precisely as in a face-to-face dialogue
with those hearers who feel aligned with the speaker. Those from the other side of the
aisle, however, will promptly identify what beliefs that underlie the speaker’s reasons they
do not share.

3 Influence of an inter-group context
(Chapter Four)

3.1 Summary of this study

In the last study, children received a collaborative reasoning task and the manipula-
tion concerned whether they were competing against another group. Their behaviour in-
dicated that they did notice whether they were in competition against an out-group or

5 An alternative not mentioned in the theoretical Introduction above: Arguers may also ignore the dis-
agreement to the extent that it does not disturb further coexistence.
not. Some children showed a proactive attitude about this, others expressed pessimism about their success, and most simply acknowledged the situation without further ado. However, the results do not suggest that competition from an out-group raises children’s motivation to put effort into their joint decisions. Indeed, also taking into account reasons that were likely rather in response to the instruction to give reasons than to convince their peer, the result spoke against the main hypothesis. Children gave more reasons in absence of out-group competition.

3.2 Issues raised and limitations

The absence of an effect in line with the prior hypothesis can be due to several things, as discussed above. The instrument may have been inadequate, or children did not understand the competitive situation, or this aspect is not relevant to reasoning. We used children’s expressions of affect towards their competitors to indicate if they were aware of their presence and understood that they were competing. But more proven and tested procedures to test attitudes than this observational measure would also have been available to use alongside the reasoning task:

- Based on minimal group assignments, 5-year-old children show greater liking for in-group members explicitly and implicitly, allocate more resources to them and ascribe more pro-social dispositions to them (Dunham et al., 2011).
- They recall more positive things about in-group members from a story, and more negative things about out-group members (ibid.).
- Six-year-olds subject out-group members to harsher costly third party punishment (J. J. Jordan, McAuliffe, & Warneken, 2014).

Administering measures like these before or after the competitive trial would likely have given greater certainty whether children understood the antagonistic aspect of their situation. If we know with reasonable certainty that children did understand their situation, and observe that they do not vary their reasoning across contexts, it becomes more plausible to conclude that the context does not matter for their reasoning. Alternatively, an experiment with the two conditions of presence versus absence of group competition
that compares two tasks, reasoning and a common goods game, may ascertain the effectiveness of the competition manipulation.

A second issue has already been raised in the Discussion of this study. Namely, children do not have an incentive to produce fewer reasons, as they do for instance in the studies in Chapter Three. As a future direction, one approach could be to add a cost of information search to the existing task, e.g. by making information available incrementally over time or as a result of work. A challenge might then be to get children to understand such a complex game. It would contain the joint reasoning task, a form of external motivation, and an optimisation problem between the cost of additional information and the quality of the solution. But in exchange, it could offer another way of quantifying children’s motivation to expend effort on a joint reasoning task depending on external motivators.

4 Conclusion and Outlook

The Social Turn in reasoning has put an important focus on the question of the function of reasoning. The short answer is communication. This has been an influential new impetus in the research into this capacity, regardless towards which of the above reviewed theoretical proposals one leans. In this thesis, I have reviewed research that documents how the social skill of reasoning emerges in childhood as part of communicative and social development. I have extended this picture by reporting experiments that make two main contributions.

Firstly, children can deal flexibly with others’ reasons, beyond evaluating them. For instance, they can assess the cogency of reasons presented to them and judge if other individuals respond to them appropriately. Even children in their sixth year already show to some extent a dispreference towards someone who accepts poorly justified proposals. The relevant study in Chapter 2 thus extends what we can say about children’s understanding of reasons, namely, that it persists when applied to further decisions. Also, this study makes a contribution from a methodological point of view. It shows that children’s judgement of reasons should not be conflated with their ascription of knowledge to the produ-
cers of reasons. And it offers a framework in which to investigate without this conflation how children assess reasons that can vary in many other diverse and more subtle ways.

Secondly, children are skilled at putting their own reasons to use in a systematic way to pursue their goals. In situations of joint decision-making, they can identify in what respects their motivations do or do not align with their peer’s, such as in the prototypical example of cooperation versus competition in an otherwise identical task. Seven-year-old children’s reason-giving reflects this recognition measurably: They adapt how they select reasons for production, even withholding them when strategically preferable. The experiment in Chapter 3, along with its predecessor, is first to show this sophisticated form of pragmatic ability in development.

In addition, these data lend themselves to further analysis in a first approach to relate reasoning, objectivity and shared commitments to each other. This is the theoretical step proposed in the comparison of the ATR and IAM that would begin to distinguish experimentally between the theories. These data allow investigating children’s reactions to reasons when they are committed to a joint goal or to diverging goals. Recall that in these two experiments, children made a series of joint decisions in either a cooperatively or competitively motivated game. A suitable approach is thus to analyse what happens when one child has given a counterargument against her peer’s goal to place an animal on their own side. An expectation following from the considerations about joint commitments would be that in the competitive condition, children more often act against their peers’ reasons. That is, we should observe that child A proposes to place an animal on her own side, child B gives a reason against that, but child A does not comply. A further key step in a new experiment would be to inquire children’s commitment to a belief after an argumentative exchange in which they dismissed this very belief.

Finally, the last study makes a contribution by pointing out further avenues for future research. Using the idea outlined in the Introduction that the “objective” nature that the ATR ascribes to reason evaluation is deeply tied to joint commitments, out-group competition may be shown to have a noticeable influence after all. Namely, if it means that these commitments are at stake. Partners engaged in cooperative reasoning may face
the option of using—and thereby acknowledging—evidence that would help solve a co-operative task, but at the same time would question the team’s superiority over its competitors. Asking out-groups for help might fall in this direction. Cooperative reasoning may thus still turn out to be influenced by out-group competition under some circumstances. Future studies on reasoning can profit from framing their research question along the lines of whether a given task puts participants before such a balancing decision.

Overall, the issues raised about the role of motivation and joint commitments in reasoning prove to be exciting future directions. The basic idea is that reasons develop their force of objectivity based on prior commitments that are implicit, but can become subject to discussion in cases of disagreement. If reasoners subscribe to these commitments, they provide a basis on which they can resolve incompatible beliefs and desires co-operatively by reasoning.

In sum, children show proficient reasoning starting before school age. They evaluate and produce reasons skilfully, while connecting their communicative strategies aptly to their overall motives. Research in how children’s reasoning develops can give deep and extremely important insight into how the daily considerations of pros and cons proceed in individuals, in and between groups, and in society at large.
Bibliography


Wittenbaum, G. M., Hollingshead, A. B., & Botero, I. C. (2004). From cooperative to motivated information sharing in groups: moving beyond the hidden profile
https://doi.org/10.1080/0363452042000299894
Appendix A: Narration Materials

for Chapter Two

Materials used for the narration in Chapter Two

Page 1. Here is a little story. At the end you can pick a new partner. One of these two: Blue Boris [point at BB] or Green Gustav [point at GG]. You need to pay attention, because one of them is a real good detective, just like you, and the other one is not. So listen well to what these two say.

Page 2. There are two other children, Mützen-Max [point at MM] and Blonde Ben [point at BB] and they met yesterday to paint together.

Page 3. They grabbed the brush and painted everything they could get their hands on.

Page 4. Mützen-Max and Blonde Ben tidied up the brush into the box and then went out the door.

Page 5. But at night, the thief came. He stole the brush from the box and ran off.

Page 6. [E unfolds the extra leaflet that from now on remains visible to the side of the booklet.] The thief now has the brush in his bag. And the two boys, Mützen-Max and Blonde Ben, didn’t notice it.

Page 7. And these two guys here, remember them?

   Free response

Blue Boris and Green Gustav met this morning and they also wanted to paint. And what’s missing?

   Response: the brush

Right. And no one knows where it is. So the two pondered: Where can we find the brush?
Page 8. But they were lucky. Mützen-Max and Blonde Ben appeared again and said: Guys, we will help you find the brush.

False belief probe (1): Although, do they know where the brush really is?

Negative response, otherwise correction.

Page 9. Mützen-Max said:

Good reason (both conditions):
Check the red box, because that is where we put the brush yesterday, and that’s also where I last saw it.

False belief probe (2): But does Mützen-Max know where the brush really is?

Negative response, otherwise drop.

Page 10. And Blonde Ben said:

Poor reason (experimental condition):
Check the yellow box, because yellow is my favorite color, and I like that color the most.

Other good reason (control condition):
Check the yellow box, because that’s where we always put the brush, and that’s where it also belongs.

False belief probe (3): But does Blonde Ben know where the brush really is?

Negative response, otherwise drop.

Page 11. Next, Blue Boris said:
I think the idea with the red box is good!

Memory question: [name], do you remember what the reason for the red box was?

Correct response: reproduce reason; give cues if hesitant, correct if false.

False belief probe (4): But does Blue Boris know where the brush really is?
Page 12. And Green Gustav said:
I think the idea with the yellow box is good!

Memory question: [name], do you remember what the reason for the yellow box was?

Correct response: reproduce reason; give cues if hesitant, correct if false.

False belief probe (5): But does Green Gustav know where the brush really is?

Negative response, otherwise drop.

Page 13. And now I have a question for you.

False belief probe (6): Which of these guys is the only one who knows where the brush really is?

Correct response: the thief; drop if false.

Page 14. So Mützen-Max thinks it’s in the red box because he last saw it there, and Blue Boris thinks that’s a good clue. And Blonde Ben thinks it’s in the yellow box because [he likes yellow/it belongs there], and Green Gustav thinks that’s a good clue.

Page 15. If you think hard about what these two said, which one is the better detective for our game?
Appendix B: Additional material pictures and plots for Chapter Four

Figure B.1: Background picture of the zoo material.

Figure B.2: Background picture of the doll-house material.
Figure B.3: Number of interactive arguments produced per trial.
Each dot represents a trial, so each dyad is represented by two dots, one in the competitive, one in the lazy condition. In the competitive condition, green dyads showed signs of noticing the competitive situation in the manipulation check, yellow ones did not. The numbers above each cell are group mean values and standard deviations.

Figure B.4: Trial durations across conditions and age groups.
Each dot represents a trial, so each dyad is represented by two dots; one in the competitive, one in the lazy condition. Green dyads showed signs of noticing the competitive situation in the manipulation check, yellow ones did not. The numbers above each cell are group mean values and standard deviations.