

# **Selective social belief revision in preschoolers**

Dissertation

zur Erlangung des mathematisch-naturwissenschaftlichen Doktorgrades

„Doctor rerum naturalium“

der Georg-August-Universität Göttingen

im Promotionsstudiengang Behavior and Cognition (BeCog)

der Georg-August University School of Science (GAUSS)

vorgelegt von

Nadja Miosga

aus Halle (Saale)

Göttingen, 2019

### Betreuungsausschuss

Prof. Dr. Hannes Rakoczy, Abteilung für kognitive Entwicklungspsychologie, Georg-Elias-Müller-Institut für Psychologie, Universität Göttingen

Prof. Dr. Stefan Schulz-Hardt, Abteilung für Wirtschafts- und Sozialpsychologie, Georg-Elias-Müller-Institut für Psychologie, Universität Göttingen

Prof. Dr. Julia Fischer, Abteilung für kognitive Ethologie, Deutsches Primatenzentrum, Universität Göttingen

### Mitglieder der Prüfungskommission

Referent: Prof. Dr. Hannes Rakoczy, Abteilung für kognitive Entwicklungspsychologie, Georg-Elias-Müller-Institut für Psychologie, Universität Göttingen

Korreferent: Prof. Dr. Stefan Schulz-Hardt, Abteilung für Wirtschafts- und Sozialpsychologie, Georg-Elias-Müller-Institut für Psychologie, Universität Göttingen

### Weitere Mitglieder der Prüfungskommission

Dr. Tanya Behne, Abteilung für kognitive Entwicklungspsychologie, Georg-Elias-Müller-Institut für Psychologie, Universität Göttingen

Prof. Dr. Annekathrin Schacht, Abteilung für Affektive Neurowissenschaften und Psychophysiologie, Georg-Elias-Müller-Institut für Psychologie, Universität Göttingen

Prof. Dr. Nivedita Mani, Abteilung für Psychologie der Sprache, Georg-Elias-Müller-Institut für Psychologie, Universität Göttingen

Tag der mündlichen Prüfung: 27. März 2019

## Preliminary Note

This dissertation is a monographic thesis. Study 1 was write up in article format, however, was so far not published. My thesis supervisors Prof. Dr. Hannes Rakoczy, Prof. Dr. Stefan Schulz-Hardt, and Dr. Thomas Schultze-Gerlach supported me with helpful advice and discussions throughout the process.

I hereby declare that all parts of this dissertation were written by myself, assistance of third parties was only accepted if scientifically justifiable and acceptable with regard to the examination regulations, and all sources have been noted.

## Many thanks to...

... Hannes Rakoczy for great supervision, friendly feedback, guidance and discussion of research ideas, and for always having an open door.

... Stefan Schulz-Hardt and Thomas Schultze-Gerlach for valuable, and helpful feedback in the process of article writing.

... Julia Fischer for constructive feedback and discussion in the thesis committee meetings.

... Julien Hambuckers and Thomas Schultze-Gerlach for support in statistical analysis.

... Rebecca Jürgens for great organization of the Research Training Group and empathic encouraging during the last 3 years.

... Regina, Marlen, Konstanze, and Lisa for coordinating the lab, organizing studies, and inviting participants.

... all my colleagues from the Kindsköpfe lab, Tanya, Anika, Ben, Britta, Franzi, Lisa, Jonas, Marina, Nese for interesting discussions and pleasant conversations, encouraging, supportive atmosphere and a joyful every day.

... all the students and research assistance who helped conducting these studies, Ariane, Marika, Christin, Anna, Josephine, Sari, Isabel, Daniel, Lydia, Johanna.

... **last but not least, to all the children who participated in my studies and their parents.**

---

**Table of contents**

<b>0. Abstract .....</b>	<b>1</b>
<b>1. General introduction .....</b>	<b>3</b>
<b>2. Social learning.....</b>	<b>5</b>
2.1. The role of social learning in humans .....	5
2.2. Epistemic vigilance.....	7
2.3. Selective social learning.....	9
<b>3. Social belief revision.....</b>	<b>13</b>
3.1. Social psychology research.....	13
3.1.1. The judge-advisor-system .....	13
3.1.2. Functionality of advice-taking behavior .....	15
3.1.3. Egocentric advice discounting .....	16
3.2. Developmental psychology research.....	17
<b>4. Argumentation.....</b>	<b>20</b>
4.1. Advice-taking and argumentation .....	20
4.2. Reasoning and argumentation .....	22
4.3. Informal fallacies .....	23
4.4. Children's understanding and evaluation of arguments .....	25
<b>5. Collective decision-making .....</b>	<b>27</b>
5.1. Collective decisions of adults .....	27
5.2. Cognitive foundations of collective decision-making.....	29
5.3. Collective decisions of children .....	32
<b>6. Aim of this dissertation.....</b>	<b>34</b>

---

<b>7. Study 1: Selective social belief revision</b> .....	<b>36</b>
7.1. Introduction .....	36
7.2. Experiment 1.....	37
7.2.1. Method .....	37
7.2.2. Results .....	40
7.2.3. Discussion.....	44
7.3. Experiment 2.....	45
7.3.1. Method .....	46
7.3.2. Results .....	47
7.3.3. Discussion.....	48
7.4. Experiment 3.....	48
7.4.1. Method .....	48
7.4.2. Results .....	49
7.4.3. Discussion.....	49
<b>8. Study 2: Argument evaluation in preschoolers</b> .....	<b>50</b>
8.1. Introduction .....	50
8.2. Experiment 1.....	51
8.2.1. Method .....	51
8.2.2. Results .....	53
8.2.3. Discussion.....	54
8.3. Experiment 2.....	55
8.3.1. Method .....	55
8.3.2. Results .....	56
8.3.3. Discussion.....	58
8.4. Experiment 3.....	58

---

8.4.1. Method .....	59
8.4.2. Results .....	64
8.4.3. Discussion.....	69
<b>9. Study 3: Collective decision-making .....</b>	<b>72</b>
9.1. Introduction .....	72
9.2. Method.....	73
9.3. Results.....	76
9.4. Discussion.....	91
<b>10. General discussion .....</b>	<b>95</b>
<b>11. Selective social belief revision.....</b>	<b>97</b>
11.1. Belief revision as a function of epistemic status .....	97
11.2. Methodical considerations and limiting performance factors .....	99
11.3. Developmental trajectories of selective advice-taking .....	100
<b>12. Argument evaluation in preschoolers .....</b>	<b>102</b>
12.1. Advice-taking as a function of argument quality .....	103
12.2. Methodical considerations and limiting performance factors .....	106
12.3. Argumentative skills as cue for selective learning and future credibility .....	108
12.4. Discounting of poor argumentators .....	109
<b>13. Collective decision-making .....</b>	<b>111</b>
13.1. Collective decision-making in children and adults .....	111
13.2. Successful collective decisions in children.....	113
13.3. Structure of dialogs leading to collective decisions .....	115
13.4. Success as a function of meta-talk and argumentative strategies .....	117
<b>14. References.....</b>	<b>120</b>
<b>Appendix.....</b>	<b>135</b>

## 0. Abstract

Recent research has shown that from the early beginning of development, children selectively form new beliefs by monitoring the competence and reliability of social sources of information (e.g., Harris, 2012). Successful learning strategies, however, do not just rely on testimony of others but rather need to be selective both as a function of the quality of supplied information and one's own informational quality. So far, little is known about the development of selective revision of existing beliefs in response to socially conveyed information. In contrast, such selective social belief revision has been extensively studied by social psychologists in the context of advice-taking. Findings of this research show that adults revise their judgments taking into account advice in a selective and systematic fashion; and thus benefit from these revisions by increasing the accuracy of their judgments (e.g., Bonaccio & Dalal, 2006). From a developmental point of view, the fundamental question is how such advice-taking or social belief revision evolves. As part of this dissertation, three studies have been conducted to address children's willingness and competence to revise beliefs as a function of their own and the advisor's state of knowledge, their sensitivity toward advice justification, and their ability to solve decision problems collectively.

In Study 1 4- to 6-year-old children and adults solved a perceptual judgment task, received advice, and subsequently made final decisions. The informational access (perceptual quality) of participants and advisor were experimentally manipulated. Adults revised their judgments systematically as a function of both their own and the advisor's informational access while children based their adjustments only on their own informational access. Two follow-up experiments suggest, however, that this pattern of results in children reflected performance rather than competence limitations: In suitably modified tasks, children did consider both their own information and that of the advisor in their selective social belief revision.

Study 2 was designed to investigate whether preschoolers are sensitive to advice justified by arguments differing in quality, and the effect of this justification on subsequent advice utilization. To this end advisors supported their judgment by giving a reason for why they want the final judgment to be consistent with the advice. Results revealed the following: Firstly, children weigh information as a function of argument quality, however, only if cues of argument quality are unambiguous. And secondly, arguments phrased similarly were endorsed differently as a function of context. This supports children's ability to judge arguments based on the epistemic relationship between information and argument.



In Study 3 4- to 6-year-old children and adults worked in dyads and jointly solved the perceptual judgment task. In a first step participants judged a stimulus individually, then discussed their individual judgments and agreed on a joint decision. Dyads mainly trusted the information supported by the strongest evidence. However, children differed in the successfulness of their joint decisions as a function of meta-talk strategies. Results suggest that preschoolers can reason with one another appropriately, and in particular reflect on individual informational access. However, more advanced meta-talk strategies identifying causal relationships between the informational access and the quality of an individual judgment may develop in an explicit form not before school age.

Taken together, the findings of this dissertation indicate that 4- to 6-year-old children are willing and able to revise existing beliefs as a function of the quality of both their own and socially supplied information. This competence was shown in different forms of social interaction: firstly, when information is presented as mere judgment of an advisor, secondly, when advice is supported by arguments and thirdly, when information is reviewed in the process of collective decision-making.

## 1. General introduction

Children are being born into a complex world. In order to successfully navigate through this world children have to accumulate enormous amounts of knowledge and skills during their first years of life and subsequent development. Much of what children as well as adults know, they learn from testimony of others. People rely on reports of others for scientific discoveries, historical facts, names of objects, and even their own date of birth. This information is varying in quality and, thus not worth considering to the same extent. Testimony can disguise reality, whether due to ignorance, verbal inaccuracy, or bad intentions. Research on developing selective trust has documented remarkable early capacities of preschool children to perceive variation among informants in knowledge and reliability (e.g., Harris, 2012). This line of research has focused on the acquisition of new information, however, in many relevant real-life situations, we do not act fully ignorant in search of completely novel information. Rather, we start out with pre-existing beliefs and judgments and the fundamental challenge is to flexibly and selectively update and revise beliefs in the light of new social information.

Such social revision of beliefs has been the focus of advice-taking behavior in adults and extensively been studied in social psychology. Findings of this research show that adults do heed advice in a sensible, yet not perfectly rational fashion. Adults' advice-taking falls short of perfect rationality since, for example, adults weight advice insufficiently and overweigh their own initial judgments. On the other hand it is considered functional because adults revise their judgments taking into account advice in a selective and systematic way; and thus benefit from these revisions by increasing the accuracy of their judgments (for reviews, see Bonaccio & Dalal, 2006; Rader, Soll, & Larrick, 2017). For developmental psychologists the question is: when does such abilities arise and what trajectories do they follow? To date, however, research on advice-taking in developmental contexts applying to children is still scarce. In particular, little is known about whether the possibility to adjust judgments in fine-grained continuous ways influences children's willingness to revise beliefs in the first place. In addition, it is unclear whether children's ability to select informants as a function of competence – as documented in the selective trust research – extends to their selection of advisors; and how much the level of children's own knowledge influences their decision-making.

Also so far it has not been investigated whether children's advice-taking is sensitive to argumentation, in particular, the presentation of arguments differing in quality, and the effect of such justification on subsequent advice utilization. The pure aggregation of individual judgments, such as studied in social psychology, ignores relevant characteristics of advice-

taking as form of social interaction between two or more people. In real-life advice is most often provided with the intent of helping someone to improve individual decisions. Advisors presented their opinions generated as piece of information suitable for the specific judge in a specific situation. Advice is then most often: firstly presented justified by arguments and secondly, discussed in form of a social dialog between judge and advisor. Thus, it represents a situation in which judges individually aggregate two judgments to transform into a joint decision. This happens as a result of discussions when evidence for individual judgment is collectively weighted for it's reliability.

The aim of this dissertation is to acquire a more detailed understanding of how social belief revision develops. In particular, the focus lies on how children develop the ability to weigh advice as a function of their own and the advisor's state of knowledge, whether advice justification affects advice-taking, and how children reach joint decisions when allowed to discuss evidence supporting individual judgments. In the following chapters I will present theoretical background on social learning, belief revision, argumentation, and collective decision-making and provide an overview of current research findings in children and adults.

## 2. Social learning

### 2.1. The role of social learning in humans

The acquisition of socially supplied information is essential for humans to successfully navigate throughout the complex world that they live in. Even small task in our daily life would become unmanageable if being restricted to self-acquired information alone. Assuming serving an exotic dinner for friends including food that one never had prepared before. The availability of geographical and cultural information supplied by other people in the first place makes the knowledge about exotic food and appropriate recipes possible. Deprived of testimony of others our spatial and temporal horizon would alone be restricted to facts we are able to perceive within our own direct environment. Consequently, we would be unaware of the existence of exotic fruits and vegetables as well as the wide variety of countries worldwide and historical events happening generations in the past. Being born into a specific culture and surrounded by adults who all speak a specific language makes effective learning important from early on in life. In order to become part of their family, community, and society children have to acquire cultural knowledge and learn the complicated structure of an abstract language. Human social organization is argued to be unique in the animal kingdom (Tomasello, Kruger, & Ratner, 1993). Although there are many social species, human cultures with their material artifacts, social institutions, behavioral traditions, and languages are incomparable. Powerful forms of social cognition and learning are necessary to prevent information loss (ratchet) and thus, to pool knowledge and skills across generations (Tomasello et al., 1993). The learning mechanism allowing such cumulative culture to develop and persist is called imitation. In order to be considered as imitation, all three sources of information (goal, action, and result of the model) have to be copied by the imitator (Call & Carpenter, 2002), thus, requiring the understanding of the intentional state of the model concerning its behavior. This sophisticated form of social learning has so far not been proven convincingly in non-human species.

From early on in life children observe their social surroundings and acquire behaviors by imitating what others demonstrate. Meltzoff and colleagues claim that already infants at an age between 12 and 21 days are able to imitate both facial expressions and manual gestures and thus, adjust their own unseen behaviors by what they have observed themselves in others (Meltzoff & Moore, 1977; see Meltzoff & Moore, 1983 for a methodological review and Oostenbroek et al, 2016 for a critical position). Infants have been described as imitative generalists, meaning that they copy actions on objects as well as basic body movements. The ability to defer imitation is crucial in situations, in which infants are not allowed to act on an

object immediately after the adult's demonstration has occurred. This ability is challenging since infants have to capture, remember and reproduce actions on a sufficient level. The capacity of deferred imitation across a 24-hour-delay or a one-week-delay was demonstrated in infants youngest at the age of nine months (Meltzoff, 1988b) or at the age of 14 months, respectively (Meltzoff, 1988a). Children's copying fidelity increases with age leading into cases of 3- to 5-year-old children who even overimitate models (Horner & Whiten, 2005; McGuigan, Whiten, Flynn, & Horner, 2007). They copy action sequences that are causally irrelevant to achieve desired outcomes and employ imitation in situations, in which alternative social learning strategies may be more efficient (Nagell, Olguin, & Tomasello, 1993). Overimitation has been shown to be a universal human trait and reflects an evolutionary adaptation that is fundamental to the development and transmission of human culture. One important aspect of cultural knowledge is *the way* things are done, not *what* gets done (Nielsen & Tomaselli, 2010). Relating back to the idea of sharing an exotic dinner with friends and knowing that Indians, for example, eat chicken provides only limited information. More importantly, one needs to know how the chicken has to be cooked and what spices need to be added in order to make it an Indian dish.

Imitation is a form of observational learning but not all knowledge can be learned by observation. Abstract facts about the world, opposed to actions, are often not visible. The same is valid for events that happened in the past and need to be passed on via verbal or written documentation. Language acquisition begins early in life. Children start speaking words around their first birthday but know the meaning of several common words already at the age of six months (Bergelson & Swingley, 2012). The ability to communicate about what is not perceptually present is cognitively more demanding, since new information needs to be incorporated into own representations of absent objects and was shown by infants at 14 months of age (Ganea, 2005). This ability allows humans to learn from testimony about a near infinite number of topics and to overcome the boundaries of their own spatial and temporal perceptions. In fact, before the end of their second year of age infants make use of this achievement and update representations of absent objects based on testimony of others (Ganea, Shutts, Spelke, & DeLoache, 2007).

## 2.2. Epistemic vigilance

Humans depend heavily on learning from others. But not all information presented is of the same quality and worth considering to the same extent. Testimony can transfer imprecise or even wrong information. Thus, social learning, besides all its benefits, bears the risk of becoming accidentally or intentionally misinforming. For communication to stabilize and allow the exchange of information in the first place costs and benefits of both communication partners need to be balanced. On one side, the communicator invests effort by performing the communicative act and in return hopes to produce the intended effect in the addressee. On the other side, the addressee needs to process the offered information mentally and benefits by acquiring information that is true and relevant. While strength of interest and degree of honesty often overlap for partners communicating in pleasant ways they rarely coincide exactly. Mechanisms of epistemic vigilance allow to control for balanced costs and benefits of communication partners and thus, are indispensable if communication itself is to remain advantageous and stabilize continuously.

Humans are not just unique in the animal kingdom because of their communication. In addition they hold an exceptional richness and strength of cognitive abilities. The skills of communication and cognition are linked and most likely evolved in parallel phylogenetically and ontogenetically. Communication advanced enough to transfer complex knowledge builds upon advanced cognitive abilities, in particular language and mindreading. These complex abilities in return can not be elaborated individually but rather rely on cognitive skills, conceptual tools, and background knowledge acquired from others. Promising models to acquire knowledge arise from uniting good intentions and high competences. If epistemic vigilance developed to diminish the risk of misinformation, as argued by Sperber and colleagues (2010), it requires the capacity to select models as a function of their epistemic status and intentions. This includes false belief attribution and the ability to identify the intention to induce false belief in addressees. Such attributions require recipients to be sophisticated mindreaders: understanding that A is trying to deceive B - as opposed to unintentionally misinforming B - involves attributing to A's intention to cause B to form a false belief. The liar's intention being metarepresentational, the attribution of such an intention is itself a second-order metarepresentation (Sperber, 2000). False belief tasks are classically used to evaluate mindreading capacities and indicate developmental trajectories. Research in both epistemic vigilance and false belief understanding point towards a striking developmental stage at around four years of age. Firstly, at this age the ability to represent the relationship between two or

more persons' epistemic stages emerges and becomes established (Wimmer & Perner, 1983) and secondly, epistemic vigilance towards dishonesty (Mascaro & Sperber, 2009) and incompetence increased tremendously (Call & Tomasello, 1999).

Children at the age of four years start to acknowledge differences in epistemic quality and selectively accept information as a function of quality. Two aspects of presented information might differ in quality: firstly, the content of the testimony – what is being said, thus, the information itself - and secondly, the source of testimony – who is providing the information. Both aspects do not have to coincide. Characteristically, A can be more or less convincing independent on characteristic B. Some information communicated is intrinsically believable or absurd independent on the apparent reliability of the informant itself. Imagine a drunk person at a party, representing an unconvincing source of information generally, tells someone that German is her/his first language speaking grammatically correct German without any accent. In this situation one can rely on the truth of this information because it was evidenced sufficiently by the act of communication itself, even though it was given by an unreliable source. Other testimony might be intrinsically unbelievable because it has been proven wrong by the own immediate perception. For example, a person at the height of approximately 1,60 m claims by telling to be 1,90 m instead. In particular, when new information can be acquired by perception it is quite sound to rely on own perceptions more than on testimony of others. Perception has evolved evolutionarily to provide individuals with reliable information about their living environment. One mechanism of epistemic vigilance checks whether testimony of others is consistent with immediate situational knowledge. In the case of inconsistency, one should trust in own perceptions, since they are, presuming everything else being equal, more likely to be reliable (Mercier, 2017). However, in many cases it is more demanding to review information provided by others concerning its quality. Believability of information can be determined by a multitude of factors like factual knowledge, past experiences, or domain specific competences. Even balancing of visual information might be complex when visual access is provided with some uncertainty.

Trustworthiness of the source of information is another important aspect of judgement when gathering knowledge from others. Those judgments need to consider the competence and benevolence of informants and may result in trusting an informant generally or in calibrating trust context or domain specifically. Such precise calibrations of trust are cognitively demanding leading people to rely on general impressions of the informant's trustworthiness. However, general impression formation of trustworthiness bears the risk of misjudging

someone as generally trustworthy or not trustworthy. In this respect a phenomenon known as committing a *fundamental attribution error* was described by Ross (1977) as follows: people base their predictions of someone's future behavior on an experience with this person in a specific situation in the past, commonly overestimating psychological dispositions, and underestimating situational factors. However, the fact that social learning plays a tremendous role in human life and allows enormous accumulation of knowledge suggests that humans are able to calibrate their trust well enough to make it advantageous phylo- and ontogenetically. The next section will discuss how and under which circumstances skills of selective social learning develop ontogenetically.

### **2.3. Selective social learning**

Young children have long been described as trusting disciples with blind credulity (e.g., Reid, 2000). Indeed, several studies support strong claims on indiscriminate credulity in children. When informed about simple matters of fact they follow social leads, even if proven false repeatedly. A strong bias to trust testimony (via pointing or words) about the location of an object was shown in a study with 3-year-old children who continued searching for a sticker in the wrong location when being encouraged by deceptive testimony (Jaswal, Croft, Setia, & Cole, 2010). Moreover, the abilities to understand the falsity of lies presented by an informant exposed as a liar or to recognize lies of an informant introduced as intending to lie do not demonstrate before the age of four years whereas younger children conform to these kinds of misinformation (Mascaro & Sperber, 2009). Children of two to three years of age trust testimony that conflicts with their own previous experience, even, if they had an incentive to ignore this testimony (Jaswel, 2010). Young children also endorse from demonstrations they have no possibility to check for themselves and treat new practices as generalized norms instead of context-specific rules (Rakoczy, Warneken, & Tomasello, 2008). They incorporate information about hidden properties of the world, for example, scientific or religious matters, into their own judgments and verifications (Harris & Koenig, 2006).

Besides all the risk of misinformation discussed earlier credulity represents an effective strategy to acquire new knowledge for young children. They rarely encounter informants who aim to misinform them about matters of the world or human life in general. Children are most often provided with true information due to the fact that parents and caregivers purposely provide knowledge in most cases with the good intention to educate children in an honest



fashion. The acquisition of language, for example, was debated as case of primitive trust where children have to accept word-object links instinctively to be able to join the linguistic communities (Burge, 1993). Furthermore, Dawkins (2006) emphasised the advantages of credulity from an evolutionary point of view, where during the process of natural selection selective pressures act on individual fitness: “Theoretically, children could learn from personal experience, for example, by not approaching a very near cliff edge, not eating untried berries, or not swimming in crocodile-infested waters. However, there will be a selective advantage to the child’s brain that possess the rule of thumb: believe without questioning whatever adults tell you.” Nonetheless, blind trust in everything being told would not be a successful trait (Richerson & Boyd, 2005) neither evolutionarily nor ontogenetically. In fact, what emerges from current research on selective social learning is a more elaborated vision of children’s attitude toward testimony. Even at a very young age, children do not indiscriminately accept information coming from testimony.

Research on children’s selective social learning, also referred to as selective trust, traditionally has adopted variations of a basic two-informant paradigm (e.g., Koenig, Clément, & Harris, 2004): A familiarization phase is used to establish in children a differing profile of epistemic history for two informants. In four trials children usually are introduced to two unfamiliar informants (adults, children, or puppets) who consistently provide accurate or inaccurate information in a context revisable for children. Most often these two informants name common objects, for example, a ball is placed on a table and the reliable informant calls it a “ball“, whereas the unreliable informant calls it a “shoe“. The two informants offer conflicting testimony on a matter unfamiliar to the children, for example, differing names for an unknown object. Subsequently, in test trials children are being examined about their marked preference in terms of informant’s testimony endorsement.

Research on selective trust indicate that preschoolers do not learn inflexibly from any source alike, but instead review a variety of different characteristics of a potential source (for reviews see Harris, 2007; 2012; Harris & Corriveau, 2011; Robinson & Einav, 2014): When learning novel words, children at the age of four years monitor variables of competence and prefer knowledgeable over ignorant (Sabbagh & Baldwin, 2001), accurate over inaccurate (e.g., König & Harris, 2005), and confident over unconfident informants (e.g., Jaswal & Malone, 2007; Birch, Akmal, & Frampton, 2010). While often tested with word learning task this selective trust is not restricted to the verbal domain only but also shown for factual information (Clement, Koenig, & Harris, 2004). Sensitivity extends to more subtle differences of informants reliability,

an attenuated contrast of 75 % correct versus 75 % wrong or non-committal remarks will be monitored and trusted, respectively (Pasquini, Corriveau, Koenig, & Harris, 2007). Children rather prefer to trust competent models even one week after exposure to competence information (Harris & Corriveau, 2011). Children capitalize on cues of benevolence to select informants, in this respect they prefer nice versus mean models (Landrum, Mills, & Johnston, 2013), models of high versus low reputation (Fusaro & Harris, 2008), honest versus dishonest models (e.g., Li, Heyman, Xu & Lee, 2014) and even select for attractiveness (Bascandziev & Harris, 2014). Moreover, children are affected in their model choice by age, familiarity, and group membership. When presented with unfamiliar models 4-year-old children expect that an adult but not necessarily a child knows the meaning of specific words (Taylor, Cartwright, & Bowden, 1991). In respect of familiarity a study showed that children are more receptive to information offered by a familiar compared to an unfamiliar caregiver (Corriveau & Harris, 2009). A preference for familiarity was also found at a group level: children demonstrate selective trust in native- accented speakers over non-native speakers independently on semantic content. (Kinzler, Corriveau & Harris, 2011).

Furthermore, preschoolers do not just review cues for reliability independently, if accessible they weigh various cues against each other and rely on the ones most relevant in a specific learning context. Children at the age of four years prefer to learn from adults rather than from peers by default. However, they show the reverse learning pattern when the topic is likely to be more familiar to peers (e.g., toys) or when the adult has previously proven to be less reliable than the peer (Jaswal & Neely, 2006). Four- and 5-year-old children prefer to believe information provided by a confident rather than hesitant individual, however, when confidence conflicted with accuracy, preschoolers increasingly favored information from the previously accurate but hesitant individual (Brosseau-Liard, Cassels, & Birch, 2014). Finally, it has been shown that infants select informants by age as a function of whether they presented novel or familiar actions (Zmyj, Daum, Prinz, Nielsen, & Aschersleben, 2012).

Studies applying a single-informant paradigm to investigate selective trust provide evidence that children are generally willing to learn from unreliable informants when no other competing testimony existed (Koenig & Woodward, 2010; Kim, Paulus, & Kalish, 2017). For example, a direct comparison of both paradigms, two-informants versus single-informant, revealed that 3- and 4-year-old children would endorse a previously inaccurate informant for an unfamiliar object's label as long as there was no other informant who proposed an alternative (Vanderbilt, Heyman, & Liu, 2014). These results correspond to other studies showing a

general bias to initially accepted social information prior to examining it and possibly rejecting it thereafter (Gilbert, Krull, & Malone, 1990).

Research studies on developing selective trust report about remarkable early capacities of selective trust. The capacities under investigation in this research, however, are restricted to the acquisition of new information. In typical test situations used in this research, the child is faced with a question being ignorant about (e.g., “What is this novel object called?”, “what is this novel tool used for?”, etc.). It is then provided with competing answers by different models to choose from. This research, thus, investigates on the development of selective social belief *formation* rather than selective social belief *revision*. However, in many relevant real-life situations humans do not act completely ignorant in search of completely novel information. Instead, pre-existing beliefs and judgments affect this process. The fundamental challenge consists of flexibly and selectively updating and revising beliefs according to new social information. In the next section approaches used in Social- and Developmental psychology to investigate belief revision in adults and children will be specified and main findings of these two research areas be reported on.

### 3. Social belief revision

Testimony offered by informants differs in quality. Similarly, own knowledge depends on the specific situation. Sound own knowledge often is involved in decision-making processes or at least such feeling exists. The quality of someone's knowledge is topic dependent, which is especially true for children. They are already experts in certain domains, know something about others and never heard of numerous domains. Nevertheless, it would be disadvantageous to constantly disregard own knowledge and blindly trust others's testimony when being presented with conflicting claims. Instead of that, the ability to correctly discriminate the degree of credulity as a function of own knowledge in relation to the quality of the social information seems most promising. Such social revision of beliefs has been extensively studied in social psychology under the rubric of *advice-taking*. When faced with important decision makings in life people often consult colleagues, parents, and specific experts for advice. Major motivations for seeking advice include the hope to improve decision accuracy (Yaniv, 2004), enhanced justifiability of someone's final decision (Kennedy, Kleinmuntz, & Peecher, 1997), or shared responsibility with others in terms of the outcome of a decision (Harvey & Fischer, 1997). The focus of this research regarding advice-taking lies on modeling such decision structures under various conditions. From a developmental point of view, the fundamental question is how such advice-taking or social belief revision develops. However, until today, other than in adult social psychology, systematic research on belief revision in developmental context is still scarce. The following section will give an overview about paradigms, main questions, and findings of this research.

#### 3.1. Social psychology research

##### 3.1.1. The judge-advisor-system

The standard method used in advice-taking research is the „judge-advisor-system“ (JAS; Sniezek & Buckley, 1995). This method is constructed perfectly to investigate the interplay of various judgments. The paradigm allows to calculate the weight of an advice given by a participant in relation to an own initial judgment (Harvey & Fischer, 1997). In a JAS two participants are assigned to the role of the “judge” (the decision-maker) or the “advisor” and both are informed about the decision task. Firstly, the judge is asked to make an initial judgment under some uncertainty (e.g. “How big is Ecuador?”), the judge then receives advice from another informant (the advisor), and lastly is asked to make a final judgment. Participants

may also be asked to express a level of confidence regarding the accuracy of their judgments. It is up to the judge whether to take the advice into consideration at all and, if so, how much weight to attribute to the advice. The judge's final decision is then often evaluated in terms of functionality, explaining to what extent the consideration of the advice increases the accuracy of the judgment.

The crucial measure is the degree of advice-taking (AT), thus, how much the initial advice has been adjusted by the judge in light of the advice (e.g., Yaniv & Kleinberger, 2000), defined by Harvey and Fischer (1997) as  $[(\text{final judgment} - \text{initial judgment}) / (\text{advice} - \text{initial judgment})] \times 100\%$ . In case the judge fully endorses the advice, AT would be 100 %. When judgment and advice are equally weighted, AT would count for 50 %, whereas, without any consideration of the advice AT would equal 0 %. Other AT scores between 0 % and 100 % represent partial shifts from the initial judgment towards the advice. These percentages correspond with revision strategies thematized in the advice-taking literature (e.g., Larrick & Soll, 2006; Soll & Mannes, 2011).

People tend to use two basic strategies: the choosing strategy in which judges choose between own initial judgment and the advice and the averaging strategy that aggregates both judgments by averaging them. It was shown that some strategies are more prevalent among judges than others (Soll & Larrick, 2009). More precisely, regarding the choosing strategy in approximately 40 % of their decision making cases judges retain their initial estimate whereas in approximately 10 % of all cases judges switch to the advisor's estimate. The averaging strategy is the least preferred strategy used by judges in only 20 % of all cases. In the remaining 30 % of decision-makings judges combine both estimates with varying weights of each but often tend to place more weight on their own estimate. However, people's preference for the choosing strategy seems not to be the most favorable since the averaging strategy has been proven to be more effective across a wide range of commonly encountered situations. However, the averaging strategy can only be applied in a JAS using continuous response formats compared to decision tasks that rely on a selection of disjoint alternatives. The distinction of categorical and continuous response formats was introduced by Hastie (1986) and defined accordingly as decisions versus judgments. Decision tasks require selecting one of two or more options whereas judgment tasks require locating a target along a continuum (Stasser & Dietz-Uhler, 2001). For example, selecting a specific flavor of ice cream out of numerous other kinds would be a decision task, while estimating the distance between Berlin and Paris would represent a judgment task. When given a decision task judges decide by an all-or-nothing-heuristic, thus, either accepting the advice to 100 % or ignoring the advice

completely. In contrast, judgment tasks allow a more fine-grained investigation of judge's adjustments of their initial estimates towards the advice. Accordingly, most studies investigating advice-taking in adults use continuous response formats.

### **3.1.2. Functionality of advice-taking behavior**

One central finding of studies investigating advice-taking in various tasks and contexts is that adults heed advice in a functional way. In general, using advice has been proven to increase decision accuracy (e.g., Sniezek, Schrah & Dalal, 2004). Combining estimates reduces random errors tied to individuals estimates and this leads to more accurate final judgments (Yaniv, 2004). Accuracy should increase even if advice were slightly to be inaccurate because heeding advice decreases decision processes in their overall complexity (Brehmer & Hagafors, 1986). In particular, the accuracy of final decisions is related to the amount of information available to the advisor, the average advice accuracy, and the advice-taking itself, thus, the weight a judge allocates to the advice (e.g., Hollenbeck, Ilgen, Sego, Hedlund, Major & Phillips, 1995). Furthermore, task-relevant experience that improves the accuracy of the advisor's recommendations and feedback have been linked to increased decision accuracy. Various types of feedback have been shown to be beneficial, for example, feedback on accuracy of the judge's initial and final judgment (Fischer & Harvey, 1999).

Adults revise their judgments considering advice in selective and systematical ways (for reviews, see Bonaccio & Dalal, 2006; Rader, Soll, & Larrick, 2017). They rely on cues about advice accuracy and prefer advisors with a promising epistemic history (Feng & Mac George, 2006), better access to information (e.g., Birnbaum & Stegner, 1979; Sniezek & Buckley, 1995), or more experienced (Harvey & Fischer, 1997) and accomplished ones (Yaniv, 2004). Moreover, adults use advice more likely when their expertise is limited (Yaniv, 2004) or the task is difficult (Gino & Moore, 2007). In generally, judges are motivated to evaluate their trust in advisors and, therefore, prefer trusted advisors (Sniezek & Van Swol, 2001).

These findings are robust even though that feelings and emotions unrelated to the decision task may distort trust in a way that gratitude enhances and anger diminishes trust (Gino & Schweitzer, 2008). Similarly, confidence has been found to moderate advice-taking behavior. A judge's pre-advice confidence and the advisor's confidence integrate into the judge's decision-making process, resulting in the judge's post-advice confidence. The *confidence heuristic* postulates that judges use the advisor's confidence as an indicator for expertise, knowledge or accuracy (Price & Stone, 2004). Accordingly, it has been found that

judges weigh advices more given by confident advisors versus less confident advisors (e.g., Van Swol & Sniezek, 2005). Judges' own confidence in the form of pre- and post-advice confidence has also been studied. Judges take less advice the more confident they are (Gino & Moore, 2007). However, confidence is affected by emotions. In this respect confidence enhances by feeling powerful (e.g., Tost, Gino, & Larrick, 2012) and diminishes by anxiety (Gino, Brooks, & Schweitzer, 2012). Post-advice confidence increases as advisors become more accurate (Budescu, Rantilla, Yu, & Karelitz, 2003), have better access to information, and when recommendations of multiple advisors overlap greatly (Budescu & Rantilla, 2000). Finally, judges increase their confidence during the advice-taking process and end up with higher post- than pre-advice confidence (Savadori, Van Swol & Sniezek, 2001).

### **3.1.3. Egocentric advice discounting**

Results of research on advice-taking show that adults heed advice in a sensible but not perfectly rational way. Adults' advice-taking lacks perfect rationality due to the fact that adults, for example, weigh advice insufficiently and overweigh their own initial judgments, a phenomenon called *egocentric advice discounting* (EAD; Yaniv & Kleinberger, 2000; Yaniv, 2004). Principally, there are no standard criteria for rationally weighing strategies of advices. The concept of EAD relies on a plausibility model predicting optimal decisions when judge's and advisor's estimates are each weighted 50 %, however, the model assumes both partners being equally competent. Accordingly, advice given by advisors being more competent than the judge need to be weighted exceeding 50 % whereas, advice provided by less competent advisors need to be weighted less than 50 %. Thus, in cases of EAD an equally competent advisor is weighted less than 50 % and a more competent advisor is weighted 50 % or less. Empirical studies on advice-taking showed that judges do not follow recommendations enough to conform to such assumptions of plausibility. Competence levels being equal, they shift towards the advice only by 20 – 30 % (e.g., Soll & Larrick, 2009). Adults discount advice even when it is based on perfect knowledge (Gardner & Berry, 1995). These findings coincide with studies investigating the judges decision-making as a function of feedback. Despite the fact that advice was more accurate than own judgments adults did not increase their advice-taking (Yaniv & Kleinberger, 2000).

Several attempts to explain these findings exist. Firstly, one approach explains EAD in terms of metacognitive processes. The underlying basic claim emphasizes that adults hold on to their own judgment due to greater insight into their own reasons and reasoning compared to the advisor's (Privacy of thoughts; Yaniv & Kleinberger, 2000). Secondly, the anchoring and

adjustment strategy (Tversky & Kahneman, 1974) proposes that judge's initial judgment may serve as an anchor that gets (insufficiently) adjusted to the advice (Lim & O'Connor, 1995). Thirdly, Krueger (2003) argued that EAD may occur because of an egocentric bias. People rely more on their own judgments and opinions because they value them stronger than those of others, for example, an advisor. Krueger supports this theory referring to judges showing an egocentric bias even when they are unable to rely on their own supporting evidence for a position (Cadinu & Rothbart, 1996) or in the absence of an initial decision to serve as an "anchor" (Clement & Krueger, 2000). Finally, one approach proposes the *mere ownership effect* (Beggan, 1992) as a mechanism underlying EAD. According to this, adults see opinions as intellectual property, which increases the subjective value of one's own ideas (Abelson & Prentice, 1989). Advice might thus be perceived as devaluing one's own ideas and be discounted consequently. So far, JAS researchers disagree on one specific explanation that can be uncritically supported empirically and theoretically.

### **3.2. Developmental psychology research**

Up to today, no systematic research as in adult Social psychology exists in Developmental psychology. Some research approaches, however, investigated precursors of, or phenomena similar to advice-taking.

At first sight, studies on conformity in children may show social belief revision. Conformity was described as tendency of individuals to forgo their own knowledge or preferences in order to endorse judgments and behavioral strategies of a majority (Haun, Van Leeuwen, & Edelson, 2013). Adults show conformity across a wide set of cultural contexts with its frequency and strength varying due to situational factors (Bond & Smith, 1996; Bond, 2005). Most studies investigating conformity in children are modeled on the classical Asch paradigm (Asch, 1956) and showed that children as young as four years of age conform to a majority of same-age peers (Walker & Andrade, 1996). In such experiments subjects are asked to make a judgment about an obvious matter (e.g. which of several sticks is the longest, when it is visually very clear what the correct answer is, e.g. "stick A") in light of a majority of other children who give an incorrect answer (e.g. "stick B"). Under these conditions, many children (just like adults) go along with the wrong group judgment ("stick B"). Empirical data on conformity in children is still limited, however, the general assumption exists that this phenomenon seems to follow the same structure found in studies including adults. Furthermore, control tasks clearly revealed that in public children do not necessarily express a judgment or belief, but just pay lip-service



(Corriveau, 2010; Haun & Tomasello, 2011). In other words, these studies merely tap normative rather than informational social influence (Deutsch & Gerard, 1955), also termed compliance versus conversion (Jaswal, Lima, & Small, 2009). Normative conformity is a social strategy, individuals adopt a majority position to gain social benefits by conforming to others (Haun & Over, 2015). In contrast, informational conformity is a strategy to optimize individual decision outcomes. Individuals adopt the majority position because they believe it to be more reliable than own judgments and accordingly follow an epistemic motive - to increase final judgments in accuracy.

Other research has shown that children are sensitive to the opinion of others when their own perceptual evidence is ambiguous (Jaswal & Markman, 2007; Jaswal, 2004). In these studies children were asked, for example, to judge what was depicted in the picture of a hybrid animal (an animal looking slightly more like a cat than a dog). Children who solely based their categorisation on visual characteristics (e.g. "cat") and children who heard an adult categorise this animal differently (e.g., "dog") followed this assertion. Jaswal and colleagues demonstrated that children as young as 24 months of age are receptive to category labels even if testimony conflicted with their own expectations and used these new labels as basis for inferences. In terms of belief revision, however, these studies are difficult to interpret since children did not announce an initial judgment. This makes it difficult to quantify the degree of belief revision or to prove that there actually existed a belief that could have been revised.

Other researchers attempted to study belief revision more directly and used already a JAS-like paradigm – children started with an initial judgment, received advice and formed a final judgment. One series of studies used tasks in which subjects were asked to perceptually identify objects based on two established procedures: the *picture game* (based on O'Neill et al., 1992) and the *tunnel game* (Whitecomb & Robinson, 2000). In the picture game the task was to identify an object portrayed on a picture. While one player (child or experimenter) saw only an uninformative excerpt of a picture, the other player saw the whole picture. For example, in the role of the less informed judge one only pictures a yellow spot, whereas in the role of the better informed judge the whole lemon was seen. In the tunnel game a cardboard tunnel with a window cut in one side for seeing through and end openings for feeling is used. An object is secretly moved into the tunnel and judges are asked to specify the characteristic "color" (e.g., red versus blue) or the characteristic "softness" (e.g., hard versus soft) while one judge was allowed to look through the window side of the tunnel and the other to feel into the opening of the tunnel. Accordingly, children in the role of the less informed judge had to base their initial judgment on insufficient perceptual access, such as looking at an uninformative excerpt of a

picture or tactile information when asked about an object's color. Results showed that children only changed their own judgment when the contradicting informant was better informed.

While all described studies so far investigated categorical belief revision, shifting entirely from one belief to the other (e.g., banana versus lemon, or red versus green), Rakoczy and colleagues (Rakoczy, Ehrling, Harris, & Schultze, 2015) investigated fine-grained belief revision and allowed participants to partially adapt their initial judgments. Their procedure allowed participants to select one out of two advice-taking strategies, averaging both judgments, or choosing between the two judgments (Soll & Larrick, 2009). Advisors were introduced as differing in competence ranging from ignorant (not knowing the name of an animal) over knowledgeable at basic level (knowing the category name of an animal, e.g., "fish") to knowledgeable at expert level (knowing the full name of the specific species, e.g., "humphead wrasse"). Participants revised their judgment systematically as a function of their advisor's expertise, taking more advice from informants that were introduced as more knowledgeable. However, children generally placed high weight on advice. They weighed advice more than their own judgments even when the advisor was clearly ignorant and thus, likely to be less competent in the feeding game than the child. This result indicates quantitative differences in advice-taking behavior between children and adults since adults overweight own judgment even though the advisor is introduced as being more competent at the judgment task (Bonaccio & Dalal, 2006). In terms of advice-taking strategies this study showed that children are both similar to and different from adults. Similar to adults children preferred the choosing strategy over the averaging strategy but crucially and different to adults children chose the advice over their own judgment while in contrast adults are known to overweight own judgments.

One limitation of these studies, however, is the uncertainty whether children's initial assumptions were only uninformed guesses. In test situations, children did not see the whole picture, they did not have the right perceptual access, or did not know the animal in question. Hence, it remains unclear whether children express initial judgments rather than just guesses; and correspondingly, whether these studies really investigate belief revision rather than guess revision. In summary, so far it is known that: children selectively acquire new beliefs on the basis of social information; children at an early age show susceptibility to normative social influence; and there is some indirect evidence for precursors of selective social belief revision. However, so far the proper selective social belief revision or advice-taking in young children remains unknown.

## 4. Argumentation

### 4.1. Advice-taking and argumentation

Within the last 30 years advice-taking research focused increasingly on empirical studies about decision-making. As a result certain understanding about why people utilize specific advice messages differently in specific situations was revealed. Research findings showed that several independent variables might determine the weight judges allocate to provided advice. People use advice differently depending on *what about it was given*, for example, topics might include science, business, or sport (Bonaccio & Dalal, 2006; Yaniv & Foster, 1997). Furthermore, it has been shown that advice utilization depends on *by whom* the advice is given, for example, friends versus strangers or experts versus novices (Bonaccio & Dalal, 2006; Feng & MacGeorge, 2006), *to whom* it is given, for example, experts versus novices or men versus women (Harvey & Fischer, 1997; Feng & MacGeorge, 2006), and the *content of the advice*, for example, the presentation of alternatives (Bonaccio & Dalal, 2006).

However, so far it has mainly been ignored whether advice can be supported by arguments. In particular, the question remains as to whether the advice may be justified by arguments differing in quality and the effect of the justification on subsequent advice utilization. This lack of information might be a result of the focus set and the method used in advice-taking research. Social psychologists study how individuals aggregate opinions, often their own (judges) and that of others (advisors). From the perspective of socially interacting with other people the majority of these studies do not deal with advice in the sense of its daily use because it is presented merely as the judgment of another person rather than as opinion generated as piece of advice suitable for the specific judge in a specific situation. However, in real life advice is most often provided with the intent of helping someone make an improved decision or judgment. Advice is than: Firstly, presented justified by arguments and secondly, discussed in combination with a potential final decision in form of a social dialog between judge and advisor.

The most robust finding in advice research is the “egocentric discounting effect” (EAD; Yaniv & Kleinberger, 2000). Judges overweigh their own opinions relative to that of their advisors (Harvey & Fischer, 1997). Within the epistemic vigilance framework, a mechanism to overcome such conservatism is argumentation. Advisors can offer arguments in defense of their views. If judges considered arguments sufficiently strong enough, they were able to overcome their tendency to discount advisors’ opinions. Yaniv (2004) argued that EAD occurs

because judges have access to their own underlying justifications and due to the strength of the arguments that support their decisions. In contrast, they do not have access to their advisors' thoughts, experiences, and qualifications, consequently, they have no access to evidence advisors base their opinions on. This could imply that providing arguments to support advice would benefit advice-taking behavior because a decision maker will have more sufficient insight into the underlying justification for the given advice. In fact, when adding a reason to a request it increases its compliance even if the reason conveyed no information (Langer, Blank, & Chanowitz, 1978). However, intuitively one would assume that an argument presented to support advice is processed for accuracy by the advisee, just like other cues in the advice-taking context are processed to assess the quality of the advice (Yaniv & Kleinberger, 2000). Depending on the judge's evaluation of the advisor's argumentation, one might expect a different effect on the degree of advice-taking.

In social psychology research there is only one study known that investigated advice-taking (the percentage rate a judge shifts own initial judgment toward the judgment of the advisor) using a classical JAS paradigm (Tzioti, Wierenga, & van Osselaer, 2014). This study paid attention to how justification of advice, in particular, advice justified by intuition ("my gut tells me so") or by analysis (sales figures of the last years), affects the weight judges allocate to the advice. Participants took on the role of a junior product manager in a company. They were asked to decide on the percentage (0 – 100 %) to which they would recommend the top management to invest in a specific product. After defining their initial judgment, judges received advice either given by a junior or senior manager and were then subsequently asked for a final judgment. Findings differed as a function of condition in terms of justification type and seniority status: Firstly, presented with an advisor at the same level of seniority as the decision-maker advice justified by intuition was discounted completely. Moreover, it was even followed significantly less than in the absence of advice justification (control group). In contrast, analytic justifications caused judges to significantly shift their opinion toward the direction of the advice. Secondly, when the advice was given by a senior advisor no significant difference in advice utilization existed between an intuitively versus an analytically justified advice. Accordingly, participants judged varying types of justifications in a way they best fit the situation in which the advice was given: Thus, a similar argument increased or decreased advice-taking depending on the specific test condition. Participants seemed to assess the quality of the advice as a function of the quality of arguments used for advice justification.

## 4.2. Reasoning and argumentation

The question arises as to how people assess the quality of arguments. Argumentation - simple or complex, implicit or explicit - results from reasoning. The term *reasoning* refers to different meanings in psychological research. Two main concepts in the literature are known: firstly, reasoning in a broader sense as a synonym of *inference* (in particular in developmental and comparative psychology). *Inference* is understood as a cognitive process that takes a representation as input, processes it, and delivers a modified (ideally enriched) representation as output. These inferences are drawn intuitively without their inputs or mechanisms being represented as reasons or arguments. For example, someone enters a cafeteria at lunchtime and sees a crowd of people there. This might imply that all these people appeared at the cafeteria to have lunch. Such conclusion is drawn spontaneously without realizing that the judgment formed was based on specific information related to the present situation (cafeteria, lunchtime) and to general knowledge about people (hungry at lunchtime). Although being sophisticated, people are most likely not aware of the reasons supporting such inferences.

The second concept contrarily argues that *reasoning* can refer to a form of inference which attends to these reasons and reflects on them before accepting a subsequent conclusion. Regarding the cafeteria scenario, reflective inferences may go further by thinking about a group of older men being dressed in workwear and standing among the people in the cafeteria. Questions arise whether these men take a lunch break from work (done somewhere in the university) or whether the same men currently do some repair work in the cafeteria building. All this may support both hypotheses, thus the given piece of evidence in this case is inconclusive. Further observations on the group's subsequent behavior might provide more sufficient insight: another man arrives at the group and the group moves toward the buffet to stand in line waiting for food servings at the lunch counter. Now the conclusion can easily be drawn and supported by the given evidence. Reflective inferences are concerned with such evidential or logical relationships among representations as to how far a given representation (initial judgment) is a good reason for accepting another representation (final judgment).

The *Argumentative Theory of Reasoning*, proposed by Mercier (Mercier, 2011; Mercier & Sperber, 2011; Mercier, 2016), claims that exactly this form of reflectively inferring is the true meaning of reasoning and that the main function of reasoning is argumentation. This entails the ability of producing arguments in order to convince others and to evaluate others' arguments appropriately. Consequently, reasoning becomes a tool for epistemic vigilance and vigilant communication (Sperber et al., 2010). Thus, this theory implies that people are interested in

and able to judge arguments regarding their sufficient quality in order to avoid being deceived by poor arguments. This is especially true when own beliefs are asked to be revised. However, people need to be objective enough to accept verified information, even if the arguments challenge their views or come from untrustworthy sources. Skilled argumentation benefits joint decision-making as well as simple cases of testimony. In case someone does not like biking this person generally will not be convinced by a colleague to ride the bike to a business meeting held today afternoon. In contrast, the same person may change plans to take the bus when being confronted with the argument that public transportation is on strike today. From an evolutionary perspective communication is difficult to maintain: senders usually benefit from deceiving and manipulating receivers. If receivers are not motivated to communicate, they stop receiving, thereby communication collapses. Argumentation can stabilize communication allowing messages to be transmitted even in the absence of sufficient trust. Accordingly, arguing is beneficial: those who produce arguments are more likely to get their messages across than if they would rely only on trust and those who receive arguments can accept verified information they might otherwise have rejected.

Classically, reasoning has been discussed as having emerged for individual cognition. In this framework, reasoning was proposed as a mechanism that allows humans to go beyond spontaneous inferences and to make more advanced judgments and decisions (Kahneman, 2003). However, Mercier and Sperber argued that reasoning is not an effective tool for individual cognition because it fails in solving logical problems (Evans, 2002), leads toward bad decision-makings (Shafir, Simonson, & Tversky, 1993), and causes poor epistemic outcomes (Kunda, 1990). In combination with its high operating costs it seems rather unlikely that reasoning developed as a tool for individual cognition. Reasoning used in solitary thinking and belief revision is rather an adapted application of reasoning skills incorporated into the concept of epistemic vigilance towards communicated information. Reasoners, in fact, image to discuss potential solutions with others (themselves) and alternate among different judgments as a function of evidence and good arguments supporting specific judgments.

### **4.3. Informal fallacies**

Research has shown that adults are motivated and qualified to evaluate arguments (e.g., Hahn & Oaksford, 2007). In particular, they differentiate between strong and weak arguments (e.g.,

Rips, 2002). Arguments perceived strong enough can lead people to revise their beliefs even without the possibility to discuss evidence with the source of information (Tzioti et al., 2014).

An interesting fact in argumentation research is the study of informal reasoning fallacies (Van Eemeren, Grootendorst, & Henkemans, 1996), i.e., arguments that are “psychologically persuasive but logically incorrect; that do as a matter of fact persuade but, given certain argumentative standards, shouldn’t” (Copi & Burgess-Jackson, 1996, p. 97). Formal fallacies can be detected by examining the form of the argument, for example, in domains such as mathematics, logic, or statistics. In the case of informal fallacies, contrarily, arguments need to be evaluated either regarding their context (e.g., the specific situation in which the argument is presented) or their content (e.g., the meaning or vagueness of words).

Three classical forms of informal reasoning fallacies include: Firstly, the argument from ignorance (*argumentum ad ignorantium*), meaning a conclusion is assumed to be true because it is not known to be false (e.g., Walton, 1999). Secondly, circularity (*petitio principii* or *begging the question*) whereby the conclusion is already contained in the premises (Rips, 2002), for example, “tomorrow it will snow because it will snow”. And thirdly, slippery slope arguments that recommend a course of action on basis of their perceived consequences. However, these specific consequences will result only from a series of intervening future steps currently not proven to occur (Corner, Hahn, & Oaksford, 2006). Several studies showed that adults detect and most often reject poor arguments. Argumentative fallacies, however, are perceived as differentially strong depending on their content (Hahn & Oaksford, 2007). Circular arguments, for example, are not always poor arguments but in some cases they rather may be the best arguments among all. As suggested by Bayesian analysis acceptance is determined by the probability of alternative explanations. Many textbook examples of these arguments are not fallacious due to their structure but rather, because they occupy the extreme weak end of the argument strength spectrum given the probabilistic quantities involved. Furthermore important, besides content, is people’s sensitivity to the context in which fallacies are presented. Rationally, arguments should be evaluated based on their soundness and strength, however, participants’ ability to identify informal fallacies is influenced by several factors such as the truth value of the argument’s premises (Neuman, Glasner, & Weinstock, 2004), familiarity with norms of argumentation (Weinstock, Neuman, & Tabak, 2004), the degree of involvement in the argument (Johnson & Eagly, 1990), or whether being presented in a reasoned dialog or not (Neuman, Weinstock, & Glasner, 2006).

#### **4.4. Children's understanding and evaluation of arguments**

So far, no studies have investigated whether children are sensitive to arguments in classical advice-taking scenarios used in social psychology (for one study in adults see Tzioti et al., 2014; also described above). Developmental work has studied, firstly the understanding and evaluation of arguments in observational and experimental designs and secondly, very recently, children's use of given arguments to determine informant's credibility. The following paragraphs will give an overview of developmental studies on the acceptance of arguments.

Observational studies on parenting style have shown that parents use reasons to convince their children and that this giving of reasons can be successful (Hoffman, 1970). Parents providing arguments to convince children are more successful in promoting, for example, altruism or moral reasoning (Grusec & Goodnow, 1994). Moreover, children provide reasons of their own (e.g., Perlman & Ross, 2005) and exchange arguments with each other to learn about their interests and engage themselves in solving problems (Ram & Ross, 2001). These observational studies suggest that children do not blindly accept arguments but may evaluate them regarding accepting sound arguments only and rejecting weak arguments.

Experimental studies have investigated children's argumentative skills more stringently. The influence of different arguments has mostly been studied in terms of moral behavior. It was observed that children do not indiscriminately accept any argument: they prefer empathic over normative arguments (Eisenberg-Berg & Geosheker, 1979) and are sensitive to the description of feelings (Kuczynski, 1982). Interestingly, arguments targeting emotions of others were more effective than telling children what they themselves would feel when following recommendations. This suggests that children can assess their own future emotions and realize that some predictions may not completely be true (e.g., „If you share your icecream, you will be very happy.“).

Studies mostly used circularity as a tool to investigate children's ability to weigh arguments. The results demonstrate that children prefer noncircular explanations starting at the age of five years and being robust in their preference by the age of 10 years (Baum, Danovitch, & Keil, 2008). Mercier and colleagues focused on even younger children (Mercier, Bernard, & Clément, 2014). They presented 3-, 4-, and 5-year-old children with a series of vignettes in which two speakers offered contradictory arguments. Subsequently, children decided where to look for a lost pet. In Experiment 1 children were presented with a strong versus a weak argument (an argument supported by perceptual evidence versus a circular argument) whereas



in Experiment 2 they were asked to choose between the weak argument and no argument. Although all age groups endorsed the strong argument, only 4- and 5-year-old children endorsed the weak argument over no argument. In a study with similar design 3- to 5-year-old children were asked to look for a toy hidden in one out of two boxes (Bernard, Mercier, & Clement, 2012). Information on the location of the toy was presented by two speakers, one giving the argument using the causal connective term *because*, whereas the second speaker used the phatic term *well*. Four- and 5-year-old children searched in the location identified by the speaker using causal connectives.

Finally, two other studies investigated whether children use explanation quality as cue to selectively trust informants or to revise a perceptual judgment. Corriveau and Kurkul (2014) showed that preschoolers extend selective trust formed based on whether informants offer strong or weak arguments from one domain to another. Five- and 3-year-old children preferred noncircular over circular explanations and extended their evaluation to their inferences about the informants' future credibility. Subsequently, all children learnt novel explanations from the informant with higher credibility. For 5-year-old children, this generalization of trust extended to word learning. Finally, one study investigated whether 2-year-old children are already sensitive to information presented in arguments (Castelain, Bernard, & Mercier, 2018): In a first phase, children were asked to name a hybrid picture of animals (e.g., fish versus bird), then received a contradicting judgment supported by either a strong argument, a circular argument, or no argument, and in the end they were asked for a final judgment. In the second phase with a similar procedure the informant never provided any argument for the own position. Finally, in the third phase, children repeated what they thought the correct names of the hybrids are in the absence of the informant. Results revealed that already at the age of two years children endorse labels supported by a strong argument and moreover, generalized their trust granted in the more credible informant. In conclusion, these studies suggest that preschoolers are able to evaluate explanatory coherence and base their decisions on the perceived quality of explanations.

## 5. Collective decision-making

### 5.1. Collective decisions of adults

Many important decisions are not made by one person acting alone but rather by two or more people in form of joint decisions. Social psychology research has a long tradition investigating the question of how people aggregate information and whether they do so in beneficial ways. Hereby, the focus was set on two aspects of collaborative decision-making: Firstly, the functionality of joint decisions, whether groups or dyads of decision-makers can outperform individuals by providing more accurate final judgments. Secondly, processes and abilities which are necessary to come up with joint decisions based on individuals sharing information with each other and weighing it optimally. To this end several studies focused on communication and exchange of reliability ratings regarding collective decision-making.

Studies that compared the accuracy of individual and group decisions have yielded somewhat inconsistent results. Experiments implementing a wide range of tasks have indicated that cooperative groups performed better than independent individuals in terms of decision-making (for review see e.g., Hill, 1982; Hastie, 1986; Kerr & Tindale, 2004). Some of the earliest evidence for that was found in the study of abstract problems (Shaw, 1932). Group and individual performance of solving the, so called, *Tartaglia problem* was compared. In this task three Cannibals and three Missionaries needed to be moved from one river side to the other by one boat that can carry only two people at a time. All Missionaries and one Cannibal were able to row. Cannibals were not allowed to outnumber Missionaries except when there were no Missionaries present. Results implied that groups, compared to individuals, make their first error later within the problem-solving process and produce more correct solutions. Group members rejected each other's incorrect suggestions more often than the individual who proposed the suggestion noticed the inaccuracy.

Further studies used crossword puzzles as problem tasks and showed group productivity being superior to individual productivity as a function of task complexity (Shaw & Ashton, 1976). Working on easy puzzles the number of groups and individuals who successfully solved the problem was proportional. When the puzzle was difficult to be solved, however, groups were more successful than individuals, implying that groups pooled their resources when no member alone was able to solve the problem. Latest research (e.g., Laughlin, Bonner, & Miner, 2002; Laughlin, Zander, Knievel, & Tan, 2003) presented adults with letters-to-numbers problems. For this task letters (e.g., A, B, C, D, E, F) were randomly assigned to numbers (e.g., 0, 1, 2, 3, 4, 5) prior to the experiment. Participants were required

to identify this mapping in as few trials as possible. In each trial the problem solver(s) proposed an equation in letters (e.g.,  $A + D = D$ ) and received feedback on the correctness of this proposal. Groups of three, four, and five people proposed more complex equations and needed fewer trials for solving the task compared to the best of an equivalent number of individuals (Laughlin, Hatch, Silver, & Boh, 2006).

However, numerous studies have documented instances where group performance was worse than that of the best member. In “social loafing” (Latané, Williams, & Harkins, 1979) individuals work differently alone than in groups; they make less effort in the presence of group members leading to a reduced overall performance. This phenomenon was explained by individual’s decreased responsibility for possible failures because no specific member within a group of people could be singled out and held directly responsible (Karau & Williams, 1993). A further reason for groups to fail was the interdependence of individual decisions (Raafat, Chater, & Frith, 2009). “Groupthink” (Turner & Pratkanis, 1998) happens in situations in which people do not start making up their own minds prior to group discussions and subsequently fail to develop and voice disagreeing opinions. Finally, another case of collective failure was described as “hidden profile” paradigm (e.g., Stasser & Titus, 1985). Group members often fail to effectively pool all information because discussions concentrate on information shared by everybody rather than on information that most relevantly provides optimal solutions. Thus, groups are biased away from “hidden profiles” leading them to under-exploit available information.

In order to make optimal joint decisions people need to share information held by individuals and weigh this information according to its reliability. Subsequently, another aspect of social psychology research concentrates on communication and its interplay with metacognitive abilities. People communicate the confidence in their individual judgments, however, this process might bear at least two risks: People might be poorly calibrated in terms of their metacognitive awareness and confidence might diverge from accuracy or might be reduced down to various egocentric biases of communication. When senders communicate information they often overestimate the clarity of their messages (Gilovich, Savitsky, & Medvec, 1998). This is an illusion of transparency or the tendency of people to believe their internal states were easily recognized by the addressee. For example, it has been shown that liars overestimated the detectability of their lies or people believed that their feelings of disgust were more apparent than they were actually in reality. Similarly, listeners are affected by egocentric biases. They often interpret the meaning of messages from their own (rather than the speaker’s) perspective (Keysar, Lin, & Barr, 2003).

Previous research gives no explanation about why egocentric biases keep up despite the recurrence of collective failure. Assumptions were made that people feel obligated to treat others equally to oneself even if less qualified or to contribute to the group despite being less competent. Following this debate Bahrami and colleagues (Bahrami, Olsen, Bang, Roepstorff, Rees, & Frith, 2012) investigated whether providing participants with a non-verbal system to share reliability ratings would remove or at least reduce these egocentric biases and in turn improve collective decision-making. Participants were asked to solve a visual perception task. Firstly, they did so individually and in a second step they combined their individual judgments. Depending on the condition participants were allowed to use verbal or non-verbal communication to discuss their potential final decision. Participant's competence was manipulated by introducing more or less noise to their individual visual perception. Scholars identified a wide gap in competence among decision partners as cause for failure of communication: Asymmetric visual sensitivity led partners that could only share their confidence non-verbally to perform significantly better than those who had talked to each other directly. In contrast, when group members had similar sensitivity and based their collective judgment on individual judgment of similar reliability near optimal performance was achieved with direct verbal communication. Based on these results the authors concluded that shared metacognitive awareness holds at least two aspects of a sensory experience, firstly, it's strength (e.g., how well did I see a mark on a stick) and secondly, it's reliability (e.g., how confident am I about the precise location of the mark on the stick). Both aspects of a sensory experience do not need to be consistent since an important role of metacognitive awareness may be to substitute missing outcomes in situations where outcome is necessary but impossible to establish.

## **5.2. Cognitive foundations of collective decision-making**

From a developmental point of view it is interesting to look at the abilities necessary to allow collective decision-making and the trajectory of these abilities. In order to jointly decide about a matter people need to be willing and able to collaborate. Basically, a collective decision is a collaborative interaction in which participants share psychological states with one another. A phenomenon that refers to such interactions on a cognitive level by functioning as an umbrella for relevant socio-cognitive skills is called *shared intentionality* (e.g., Searle, 1995). One of these socio-cognitive skills is *joint attention*, which was proposed as origin of collaboration during development (Kuhn, 2015). Joint attention goes beyond knowing what others see or follow their gaze, it is two people that experience the same thing at the time

both knowing that they do so (Tomasello, 1995). This ability is essential to allow collaborative activities and cooperative communication.

Already prior to their first birthday infants are able to recognize that they can share an object of attention with another person (Tomasello, & Carpenter, 2007) and are motivated to do so (Bakeman & Adamson, 1984). The next important prerequisite for optimizing collective decisions is the ability to discuss different pieces of evidence that may arise from different perspectives of decision partners on the same task. Increasing skills of joint attention allow to build up a joint attentional frame that functions as common ground between both decision partners to cooperatively exchange information about potential solutions. For example, there are two people participating in a visual perceptual task that requires to transfer a mark on a stick (20cm in height) as precisely as possible into a picture of the stick. Firstly, individual judgments are made and subsequently both participants compare their individual judgments and are asked to make a joint decision agreeing on a precise location of the mark on the stick. In the moment of a joint decision it is fully understandable for participant A if participant B just pointed onto the individual judgment because it is performed in the context of this specific decision task that both participants were introduced to and agreed to solve collectively. Continuously, participant B might not convince participant A with the argument "we should decide for my judgment, it is the better one" when both participants had the same visual access. In contrast, in a situation in which one participant had better visual access than the other participant and both are aware of this difference in informational access no further explanation is required because of the already established common ground of this specific judgment task.

However, optimal joint decision-making requires more than coordinating two people's attention. Decision partners have to consider reasoning together which is assumed to be more complex than joint attention. Literature reviews on developmental aspects so far did not clearly indicate how and under what circumstances social interactions affect reasoning. A person A sharing experience and knowledge with person B most likely exposes person B to new ways of thinking (Levine, Resnick, & Higgins, 1993). It remains unknown, however, to what extent and in what manner, this new knowledge is incorporated into person B's thoughts. Person B is likely to become aware that alternative solutions might exist and if convinced of it's quality person B may revise own individual beliefs or judgments to achieve a more optimal collective decision. This implies changes at an epistemic level, thus, changes in which person B adopts person A's beliefs as superior to achieve a good collective

performance. These are changes at the meta-level of what an individual knows about what oneself compared to others know.

Metacognition encompasses both *metacognitive knowledge* and *regulation*. The former can be seen as part of theory of mind, which refers to the ascription of mental states (e.g. intentions, beliefs, and desires) to oneself (metacognitive knowledge) and to other people. Metacognitive regulation describes the conscious regulation of one's cognitions and can thus be seen as part of executive functioning which refers to the conscious regulation of one's cognitions, actions, and emotions. Some aspects of *metacognitive regulation* show developmental improvements from childhood to adolescence (Zelazo, Frye, & Rapus, 1996). Moreover, children's understanding of their own epistemic states has been reported to fully develop in their late childhood; in a reflective/explicit form not even until before school age (for review see: Kloo & Rhower, 2012). Specifically, preschoolers are able to assess what they know, but have fundamental problems to correctly distinguish own uncertainty or ignorance. They have been shown to take a 'sense of knowing' for perfect knowledge. While the same feeling is described as tip-of-the-tongue phenomenon in adults, they do not commit to such judgment as being the correct answer unless supported by evidence. In contrast, young children form *relevant guesses* that seem to subjectively turn into knowledge limiting children's ability to accurately assess their own ignorance.

A further challenge for preschoolers is the argumentative discourse itself that leads to a joint decision, in which people produce and evaluate arguments to optimize their decision outcome. Giving and evaluating reasons was described as reasoning in the narrow sense by Mercier and colleagues (Mercier & Sperber, 2011) that developed as a social skill to function in interpersonal communication. Their "Argumentative Theory of Reasoning" was motivated by the concept of "Epistemic Vigilance" (Sperber et al., 2010). Listeners avoided being deceived by mistrusting speakers. In order to overcome such mistrust speakers presented good reasons for their point of view and acceptance was based on logic and evidence instead of trust. On the other hand, it was emphasized that reasoning is also important within cooperative contexts (Tomasello, 2014). Decision partners evaluated individual judgments and arguments because they shared the mutual interest to reach the best decision based on logic and evidence. In the end the argument quality itself was more important for partners than the fact who's argument they followed.

In order to give reliable reasons for individual judgments advanced metacognitive skills are required because speakers need to firstly, refer to the source of their information (where or from whom they received it) and secondly, express how reliable this information

source is (Mahr & Csibra, 2017). Toddlers as young as 18 months produced arguments to justify their violations or to argue with their parents (e.g., Perlman & Ross, 2005). By the age of three years, children chose between positive and negative reasons as a function of desired courses of action or specific beliefs (Dunn & Munn, 1987). However, the question then arises to what extent this early production of arguments is appropriate and successful. One example for successful production of arguments comes from research on moral norms: Children use different arguments depending on the kind of violation (conventional or moral) that has been committed (Nucci, 1985). Subsequently, more complex argumentation increases the likelihood of successful conflict resolution in children (e.g., Ram & Ross, 2001). However, the ability to use meta-talk in order to assess the quality of arguments and/or evidence directly has so far been shown convincingly only in older children at around 11 or 12 years of age (Kuhn, Zillmer, Crowell, & Zavala, 2013).

### **5.3. Collective decisions of children**

Studies investigating collective decision-making in preschoolers are scarce. During the last five years Köymen and Tomasello studied children's argumentation and meta-talk in dyadic interactions. The following paragraph will give an overview about the procedures and results of these first studies.

Firstly, preschoolers are able to jointly reason with peers based on appropriate common ground assumptions or warrants (Köymen, Rosenbaum, & Tomasello, 2014). Dyads of 3- and 5-year-old children built a toy zoo together, in particular they were asked to place toy items onto a cloth with marked walkways, sand-floor, lake areas etc. The toy items were either conventional (items regularly found in a zoo, e.g., different kinds of animals, cages, visitors) or unconventional (items not seen regularly in a zoo, e.g., washing machine or piano). In order to justify where to place specific items children relied on implicit warrants. For example, when finding a home for polar bears they only stated the fact "This is ice", but did not share information on habitat conditions of ice bears. On the other hand, placing unconventional items children were not able to assume common ground assumptions between each other, thus, they more often articulated the warrant explicitly. Children at the age of five versus three years more often made explicit warrants, produced justifications, and reached mutual agreement. Furthermore, preschoolers also adapted the content of their arguments to their partner's assumed level of informedness (Köymen, Mammen, & Tomasello, 2016). Under the experimental condition in which peers learnt about unique

characteristics of a novel animal (e.g., eating rocks) separately preschoolers referred more often to these unusual characteristics to justify recommendations in subsequent joint decisions than when both partners together were introduced to the new animal.

Secondly, preschooler's argumentative strategies were sensitive to the situational degree of competitiveness (Domberg, Köymen, & Tomasello, 2018). Dyads of 5- and 7-year-old children were asked to decide jointly where to place specific animals in a zoo scenario. They were required to decide on one of two halves of a playing field owned by either child. Dyads played two versions of the same game: a cooperative version in which both players were rewarded for finding the nicest home for each animal species and a competitive version in which only the child gathering most animals on its own half of the zoo was rewarded. Children produced arguments at a higher frequency and also more 'two-sided' arguments in the cooperative compared to the competitive version of the game. In the later version children's arguments showed a bias that was the result of withholding known arguments. Thus, cooperative decision-making seemed to provide a more motivating context for children to produce arguments than in a competitive situation.

Thirdly, children from around five years of age were able to jointly reflect on evidence to optimize collective decisions (Köymen & Tomasello, 2018). Peer dyads played a feeding game in which they were asked to collectively choose the correct kind of food for a new animal they met in a familiarization phase prior to the test phase. In the Experiment 1 dyads of 5- and 7-year-old children were individually introduced to the animal's feeding habits, however, were presented with conflicting information (e.g., feeding on rocks vs. feeding on sand). Moreover, the sources of information differed in reliability between both partners of one dyad: Child A received first-hand explanations given by a cartoon version of the new animal describing its own food preferences. In contrast, child B received indirect evidence presented by a cartoon girl who gave information she had heard about this new animal. Dyads decided for the food option supported by the most reliable evidence. Moreover children, especially at the age of seven years, engaged in various kinds of meta-talk about the evidence and its validity. Experiment 2 investigated even younger children which were presented with a modified procedure: 3- and 5-year-old children interacted with a puppet who tried to convince children to change their minds by producing meta-talk. Five-year-old children receiving unreliable information were more likely to change their minds, whereas three year old children did not. These results suggest that the ability to reflect on individual judgments based on evidence with differing quality emerges in children earliest at the age of five years.



## 6. Aim of this dissertation

The aim of this dissertation is to acquire a more detailed understanding of children's selective revision of existing beliefs in response to socially conveyed information. In particular, the focus is set on how children develop the ability to weigh advice as a function of their own and the advisor's state of knowledge, whether advice justification affects advice-taking, and how children reach mutual agreements based on collective reasoning about individual judgments differing in quality. To this end this dissertation integrates two lines of research: social psychological research on advice-taking in adults and developmental research on selective social learning in children.

The first study of this dissertation aimed at analysing children's (and adults') selective advice-taking as a function of both their own informational status (will they selectively revise beliefs more sufficiently under poor epistemic conditions?) and the informational status of informants (will they selectively revise beliefs more in response to informants under better epistemic conditions?). Based on classical methods used in social psychological research a basic JAS set-up in terms of being simple and ecologically valid was implemented. Importantly, this task provided children with actual sensory evidence to base their judgments on differentiating it from mere guesses, and manipulated both the subject's own informational access as well as that of the advisor. Using different variation of the basic task design Study 1 investigated children's competence of selective belief revision and performance limitations of this competence due to specific task demands.

In the second study argumentation as a further aspect of advice-taking in social settings is investigated. In daily life advice is most often not received purely as alternative judgment of an other individual but rather as proposal justified by relevant facts. Study 2 explored this aspect of decision-making, in particular, whether preschoolers are sensitive to advice justified by arguments differing in quality, and the effect of this justification on subsequent advice utilization. Importantly, in the current study two arguments were contrasted both containing epistemic information. This required children to weigh arguments based on different degrees of epistemic values, opposed to, for example, circular arguments. This design allows to investigate children's skills of argument evaluation and its effect on social belief revision in a more naturalistic version of argumentation, in which differences in argument quality are more detailed and context dependent.

Finally, the third study focused on social belief revision in the process of collective decision-making. Therefore, in dyadic interactions decision partners reasoned about

individual judgments and the evidence they are based on to reach optimal joint decisions. The rationale of Study 3 firstly investigated children's ability and willingness to reach mutual agreements based on individual judgments differing in quality and secondly identified meta-argumentative strategies and elements of meta-talk that correlate with the functionality of children's joint decisions.

## 7. Study 1: Selective social belief revision

### 7.1. Introduction

Recent research has shown that from early development children selectively acquire new beliefs on the basis of social information. Evidence exists about children's early susceptibility to normative social influence and indirectly about precursors of selective social belief revision. But, so far, young children's proper selective social belief revision or advice-taking remains unknown. The goal of the present study was, thus, to investigate the early development of selective social belief revision more directly, systematically and stringently. In particular, children's (and adults') selective advice-taking was analysed as a function of both: their own informational status (will they selectively revise beliefs increasingly when they themselves are in poor epistemic conditions?) and the informational status of informants (will they selectively revise beliefs increasingly in response to informants who are in better epistemic conditions?).

In Experiment 1, dyads of participants with three contrasting levels of visibility were tested: (1) Judge and advisor had both poor visibility (*poor-poor*). (2) The judge had high and the advisor poor visibility (*high-poor*), and (3) the judge had poor and the advisor high visibility (*poor-high*)<sup>1</sup>. Past research has shown that children prefer knowledgeable informants when learning new information (König & Harries, 2005) and are willing to revise guesses following an informant that had better visual access (Whitecomb & Robinson, 2000). Therefore, it was anticipated in this experiment that children, just as adults, would accept advice more likely from an advisor with better visibility and respectively more sufficient information. Furthermore, it was assumed that children would increase accepting advice when they themselves are rather poorly than better informed. However, results in this research showed that children (unlike adults) adjusted their degree of advice-taking as a function of their own informational quality entirely. Experiments 2 and 3, therefore, investigated potential explanations for this lack of advice differentiation and suggested that under suitable circumstances children are able to adjust their social belief revision selectively as a function of both the own and the advisor's informational access.

---

<sup>1</sup> In the present design, it was conceptually impossible to implement a high-high condition because the advice consisted of random judgments to ensure comparable differences between initial estimates and advice in all conditions. While such differences are highly plausible whenever at least one of the two judgments is based on poor visibility, the same is not true when both judge and advisor have good visibility of the stimulus. Here, participants would expect the advisor to make a similar judgment. Frequent violations of this expectation could result in participants losing trust in the advisors' perceptual reliability or the credibility of the experimental set-up.

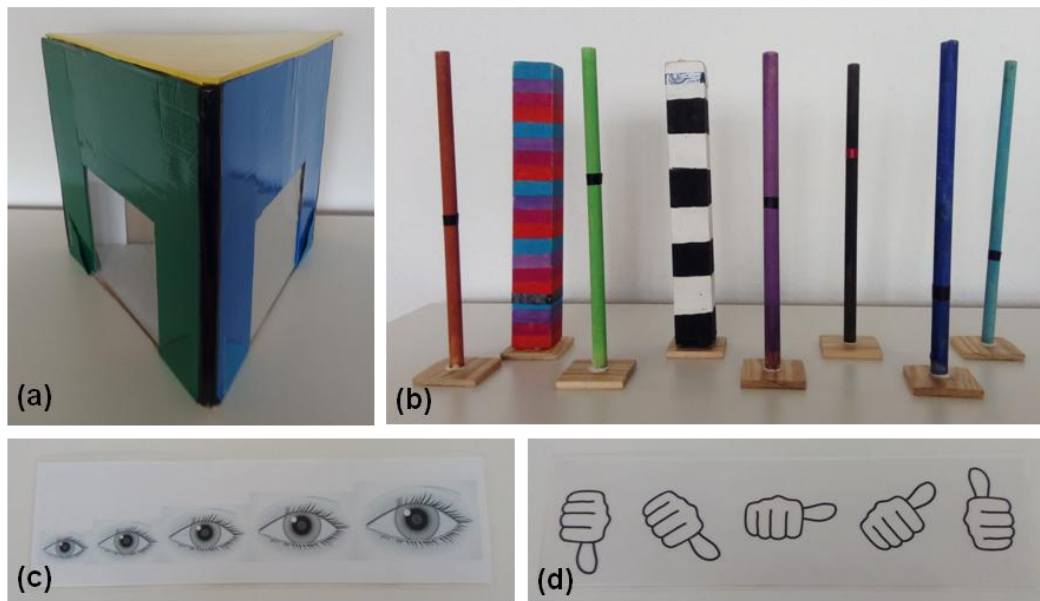
## 7.2. Experiment 1

### 7.2.1. Method

**7.2.1.1. Participants.** Thirty-five adults (age range: 18 - 44 years,  $M = 24.7$  years, 26 women) and forty 4- to 6-year-old children (age range: 48 - 71 months,  $M = 59.5$  months, 23 girls) were included in the final sample. Subjects were native German speakers with mixed socioeconomic backgrounds. Adults had answered our study announcement on public notice-boards. Children were recruited from a database of families who had previously given consent to their participation. Six additional children were tested but excluded from the final sample due to experimenter error ( $n = 1$ ), comprehension problems ( $n = 2$ ) or because they were uncooperative ( $n = 3$ ). Participants were tested individually in the lab.

**7.2.1.2. Material.** Participants were presented with a perceptual judgment task in which they were to estimate the location of a mark on a stick. Wooden sticks of 20 cm length were used. Sticks were placed in a triangular box (edge length of triangle: 57 cm, height: 44 cm). The box consisted of a clear window side, a blurred window side (window measurements: 27.5 cm  $\times$  23.5 cm), and an opaque backside (see Fig.1). At the clear window, the black mark was easily visible while at the blurred window it was hardly visible. Participants made their judgment by indicating on an analogue scale (a picture of the stick of the same length) where the mark was. Advice consisted in randomized judgments, ranging from one to 200 mm, produced by a random generator and secretly marked on the advisor's scale prior to the test sessions.

For visibility ratings (how well participants could see the mark), children were presented with an eye scale (8 cm  $\times$  29 cm) consisting of five eyes of different size. For confidence ratings (how confident participants were of their judgments), we presented them with a 5-point scale consisting of 5 thumbs ranging from one pointing straight down to one pointing straight up (see Fig.1). Adults rated visibility and confidence using a standard 5-point Likert scale.



**Figure 1.** Three-sided box with clear and blurred window (a). Colored sticks (b), rating scales for visibility (c) and confidence (d).

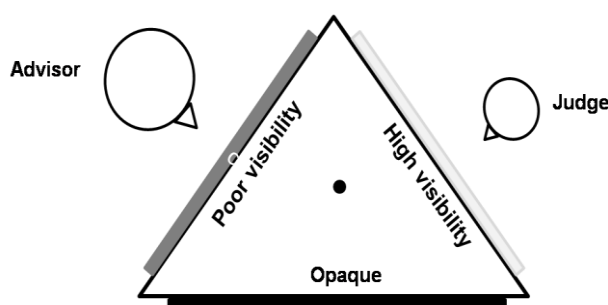
**7.2.1.3. Design and procedure.** The task was presented in JAS format with participants taking the role of judges. Advisors were confederates of the experimenter who presented the judges with the randomized advice. We used a 2 (age groups)  $\times$  3 (conditions) mixed factor design, with age groups (children and adults) as between factor and conditions (*poor-poor*, *high-poor*, and *poor-high*) as within-subjects factor. All subjects received two trials of each condition. Order of conditions was counterbalanced across subjects.

Every session started with an introduction phase prior to the test phase. In the introduction phase, participants were, in a first step, familiarized with the set-up material to make sure that they understood the design of the three-sided box, the concept of transferring a black mark on a colored stick onto a picture of the stick, and the subjective rating of visibility and confidence.

**7.2.1.3.1. Introduction phase.** The introduction phase consisted of two steps. In the first step, children were introduced to the confidence rating scale. Examples were used to illustrate the different points of the confidence scale. The thumb up symbol was introduced by presenting a toy frog with the sentence “This is a frog. I am absolutely sure about this.”. Point three of the scale - a horizontal thumb - was introduced by presenting a picture of a toy hybrid animal consisting of a pig and a fish with the sentences “This is a pig. But it could also be a fish. Now, I am not so sure.”. For the thumb down symbol, an undefined soft toy was used and presented with the sentence “I do not know what this is. Now, I am not sure at all.”. Subsequently, children were asked to show the “absolutely sure – thumb”, the “not so sure –

thumb” and the “not sure at all – thumb”. When they answered incorrectly the introduction of the confidence scale and comprehension question was repeated two times. When participants failed the comprehension question three times the experimenter moved on. The visibility rating scale was introduced in similar ways. The biggest eye was introduced by pointing at a toy duck located in a corner of the test room and saying “Now, you see this duck really clearly.”. To introduce point three of the scale, a medium size eye, children were allowed to put on sun glasses and the experimenter said “Now, you can see the duck not so clearly.”. Point one of the scale, the smallest eye, was introduced by asking children to close their eye and emphasizing “Now, you do not see the duck at all.”. Children were then asked to show the symbol for “seeing clearly”, “seeing not so clearly” and “do not see at all”. If they failed to answer these questions correctly, the introduction of the visibility scale was repeated two times before the experimenter moved on. For the adult participants (who used a Likert scale), a similar familiarization with the confidence and visibility ratings was not necessary. In the second step, participants judged one stick through the clear and one stick through the blurred window (order of windows counter-balanced). Visibility and confidence ratings were requested for each judgment. This step served as manipulation check to ensure the validity of our set-up.

7.2.1.3.2. *Test phase.* Judge and advisor were each placed at one side of the box (for illustration, see Fig. 2). Then, one stimulus was put into the box (in such a way that subjects could not observe the placement), and both judge and advisor were asked to make their initial judgments. Then, judges were seated at the opaque side of the box, received the scale of the advisor in addition to the scale with their own initial judgment, and were asked to make their final judgment on a third scale. Judges rated visibility and confidence both for their initial and for their final judgment.



**Figure 2.** *High-poor* condition with child judge.

The dependent measures were as follows:

- *Visibility rating*: Children accessed their own visibility for their initial and final judgment on a 5-point scale (5 eyes of different size). Adults were asked to rate their visibility ranging from 1 (low) to 5 (high).
- *Confidence rating*: Children accessed their own confidence for their initial and final judgment on a 5-point scale (5 thumbs ranging from thumb down to thumb up). Adults were asked to rate their confidence ranging from 1 to 5.
- *Deviation from the true value*: The deviation from the true value is the distance of the mark drawn by the judge from the true location of the mark on the stick (measured in cm).
- *Advice-taking (AT)*: The advice-taking measure is the weight the judge gives the advice, defined by Harvey and Fischer (1997) as  $AT = \frac{\text{final estimate} - \text{initial estimate}}{\text{advice} - \text{initial estimate}} \times 100\%$ . If the judge fully endorses the advice, AT would be 100 %, if the judge equally weighs her judgment and the advice, AT would be 50 %, and if she does not consider the advice at all, AT would be 0 %. Other AT scores between 0 % - 100 % represent partial shifts from the initial judgment towards the advice. In order to control for extreme cases and outliers, AT scores are usually truncated. Accordingly, we limited them to 100 % and -100 % (see Schultze, Mojzisch, & Schulz-Hardt, 2017, for a similar approach).

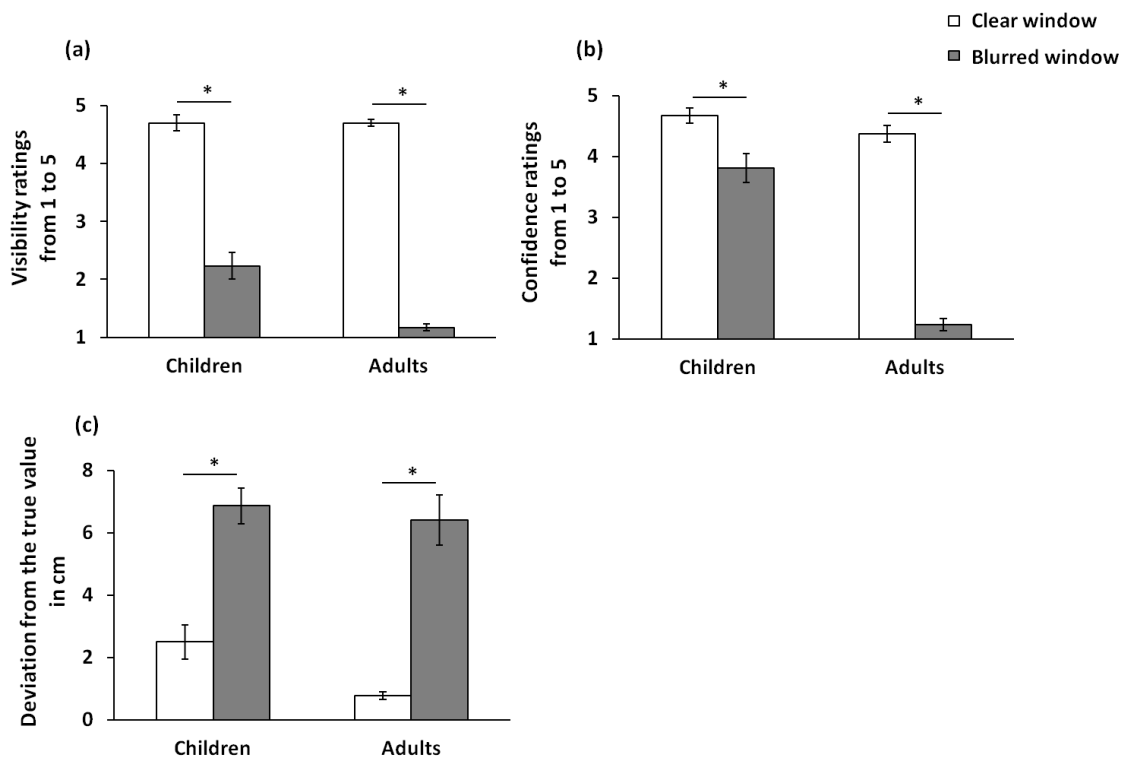
## 7.2.2. Results

**7.2.2.1. Preliminary analyses.** Analysis of the manipulation check showed that the design of the set-up was valid. The participants' visibility and confidence ratings as well as accuracy (indicated by the deviation from the true value) were successfully manipulated by watching either through the clear or blurred window. An ANOVA was calculated for repeated measures with window (clear or blurred) as within-subjects variable and age group (children or adults) as between-subjects factor to analyze whether there were significant differences in ratings and accuracy between the two windows. For post-hoc analyses independent and paired sample *t*-tests were used. Participants rated their visibility and confidence in their judgment significantly higher behind the clear than behind the blurred window and in fact made more accurate judgment behind the clear window (see Fig. 3)

For visibility and confidence results, the ANOVAs revealed significant main effects of quality of the window (visibility:  $F(1, 76) = 461.39, p < .001, \eta_p^2 = 0.89$ ; confidence:  $F(1, 76) = 222.87, p < .001, \eta_p^2 = 0.75$ ) and age group (visibility:  $F(1, 76) = 7.65, p = .007, \eta_p^2 = 0.09$ ; confidence:  $F(1, 76) = 55.82, p < .001, \eta_p^2 = 0.42$ ) and a significant interaction of both factors

(visibility:  $F(1, 76) = 18.17, p < .001, \eta_p^2 = 0.19$ ; confidence:  $F(1, 76) = 72.44, p < .001, \eta_p^2 = 0.49$ ). Post-hoc tests revealed that ratings of both children and adults were higher behind the clear than behind the blurred window (visibility: children:  $t(42) = 9.87, p < .001, d_z = 1.52$ ; adults:  $t(34) = 46.29, p < .001, d_z = 7.94$ ; confidence: children:  $t(42) = 4.16, p < .001, d_z = 0.64$ , adults:  $t(34) = 20.38, p < .001, d_z = 3.49$ ). Furthermore, children and adults differed in their ratings when judging sticks behind the blurred window (visibility:  $t(48.65) = 4.48, p < .001, d = 1.28$ ; confidence:  $t(56.5) = 10.07, p < .001, d = 2.68$ ) but not behind the clear window (visibility:  $t(56.62) = -1.05, p = .298, d = -0.28$ ; confidence:  $t(76) = 1.65, p = .104, d = 0.38$ ).

For the deviation from the true value, the ANOVA revealed a significant main effect of quality of the window,  $F(1, 72) = 86.96, p < .001, \eta_p^2 = 0.55$ . Post-hoc tests revealed more accurate judgments behind the clear window (children:  $t(39) = -5.88, p < .001, d_z = -0.94$ ; adults:  $t(33) = -7.43, p < .001, d_z = -1.29$ ) with adults being more precise than children behind the clear window ( $t(43.17) = 3.06, p < .004, d = 0.93$ ) and no difference between children and adults behind the blurred window ( $t(62.31) = 0.47, p = .638, d = 0.12$ ). In sum, these results confirm the success of our experimental manipulations.

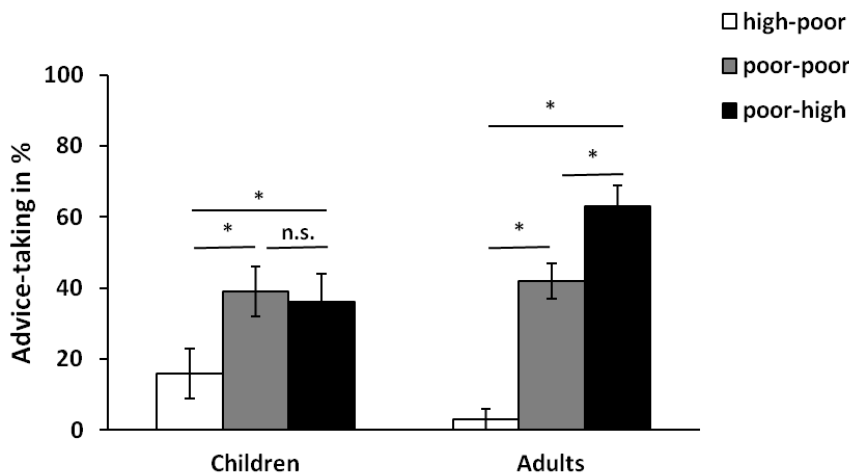


**Figure 3.** Children's and adults' subjective visibility (a), confidence (b) and deviation from the true value (c) behind the clear and blurred window. Error bars indicate standard error of the means.



**7.2.2.2. Main analyses.** Performance in advice-taking as a function of age groups and condition is depicted in Fig. 4. A 3 (condition)  $\times$  2 (age groups) ANOVA showed a significant main effect of condition ( $F(2, 73) = 21.58, p < .001, \eta_p^2 = 0.23$ ), no main effect of age group ( $F(1, 73) = 1.04, p = .311, \eta_p^2 = 0.01$ ) and a significant interaction of both factors ( $F(2, 73) = 5.10, p < .001, \eta_p^2 = 0.06$ ).

Post-hoc tests revealed that while adults differed ( $hp < pp < ph$ ) in their advice-taking between all three conditions (*high-poor* vs. *poor-poor*:  $t(34) = 5.52, p < .001, d_z = 0.95$ , *poor-poor* vs. *poor-high*:  $t(34) = -2.84, p = .008, d_z = -0.49$ , and *high-poor* vs. *poor-high*:  $t(34) = -8.91, p < .001, d_z = -1.53$ ) children's advice-taking differed ( $hp < pp = ph$ ) only between the *high-poor* compared to the other two conditions (*high-poor* vs. *poor-high*:  $t(39) = -2.06, p = .046, d_z = -0.33$ , *high-poor* vs. *poor-poor*:  $t(39) = 2.29, p = .027, d_z = 0.37$ , and *poor-poor* vs. *poor-high*:  $t(39) = 0.21, p = .838, d_z = 0.03$ ). Contrasting adults' and children's performance, the main difference in advice-taking was found in the *poor-high* condition ( $t(69.70) = -2.72, p = .008, d = -0.65$ ). In contrast, children and adults did not significantly differ in the *poor-poor* ( $t(69.28) = -0.37, p = .710, d = -0.09$ ) and the *high-poor* conditions ( $t(57.01) = 1.67, p = .100, d = 0.44$ ).



**Figure 4.** Advice-taking as a function of condition. Error bars indicate standard error of the means.

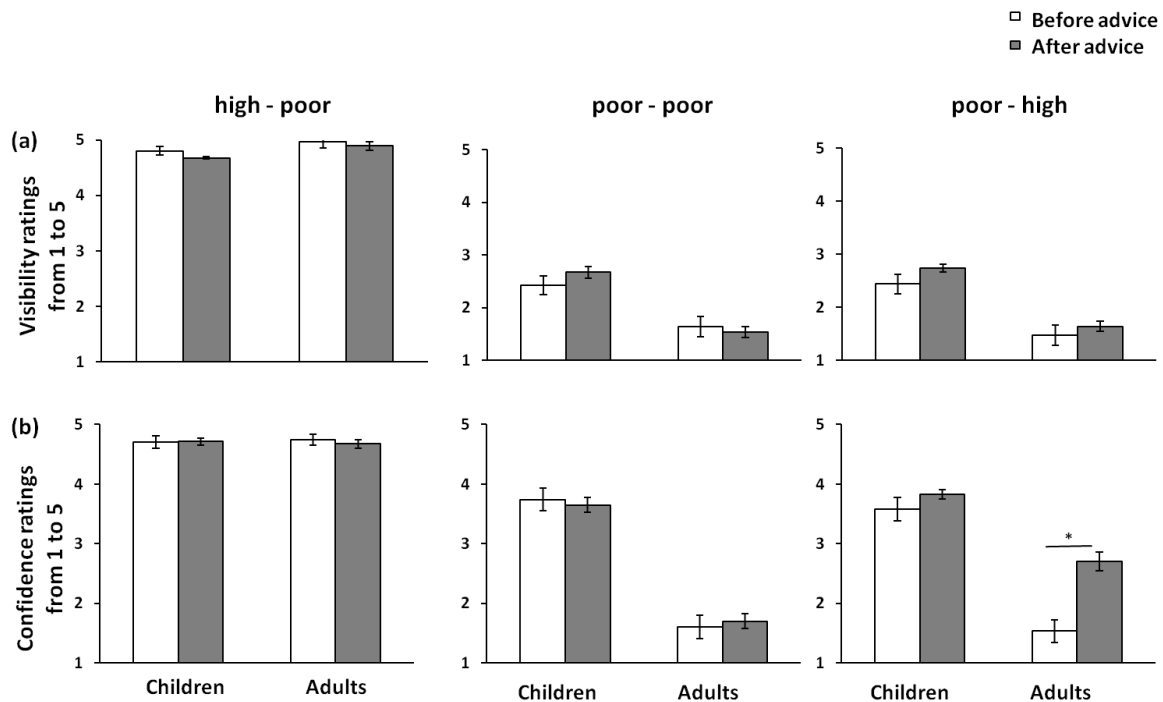
**7.2.2.2.1. Visibility and confidence rating.** Participants' visibility (a) and confidence (b) ratings as a function of age groups and condition are depicted in Fig. 5. For both visibility and confidence ratings, the target analysis was a 2 (age group)  $\times$  3 (condition)  $\times$  2 (judgment: initial vs. final) ANOVA.

*Visibility rating.* The  $2 \times 3 \times 2$  ANOVA revealed main effects of condition ( $F(2, 72) = 306.06, p < .001, \eta_p^2 = 0.81$ ) and age group ( $F(1, 72) = 13.20, p = .001, \eta_p^2 = 0.15$ ) as well as a significant interaction of condition and age group ( $F(2, 72) = 15.17, p < .001, \eta_p^2 = 0.17$ ). Post-hoc tests revealed that while in the *high-poor* condition children and adults did not differ in their initial ( $t(42.4) = -1.77, p = .083, d = -0.54$ ) or final judgments ( $t(52.5) = -1.57, p = .123, d = -0.43$ ), children rated their visibility for both judgments higher than adults in the *poor-poor* condition (before advice:  $t(73) = 3.11, p = .002, d = 0.73$ ; after advice:  $t(72) = 3.92, p < .001, d = 0.92$ ) and the *poor-high* condition (before advice:  $t(73) = 3.79, p < .001, d = 0.89$ ; after advice:  $t(72) = 4.16, p < .001, d = 0.98$ ).

There was no significant effect of judgment ( $F(1, 72) = 1.61, p = .209, \eta_p^2 = 0.02$ ). Additionally, no interactions of judgment and age group ( $F(1, 72) = 1.85, p = .178, \eta_p^2 = 0.03$ ), condition and judgment ( $F(2, 72) = 2.86, p = .061, \eta_p^2 = 0.04$ ), or condition, judgment, and age group ( $F(2, 72) = 1.20, p = .304, \eta_p^2 = 0.02$ ) were found.

*Confidence rating.* The  $2 \times 3 \times 2$  ANOVA revealed main effects of condition ( $F(2, 73) = 164.31, p < .001, \eta_p^2 = 0.69$ ), judgment ( $F(1, 73) = 11.85, p = .001, \eta_p^2 = 0.14$ ), and age group ( $F(1, 73) = 39.35, p < .001, \eta_p^2 = 0.35$ ), interactions of condition and age group ( $F(2, 73) = 37.04, p < .001, \eta_p^2 = 0.34$ ), judgment and age group ( $F(1, 73) = 6.63, p = .012, \eta_p^2 = 0.08$ ), as well as of condition, judgment, and age group ( $F(2, 73) = 6.83, p < .001, \eta_p^2 = 0.09$ ).

To follow up on this 3-way interaction, separate  $2$  (age group)  $\times$   $2$  (judgment: initial vs. final) ANOVAs were conducted for each condition. In the *high-poor* condition, this  $2 \times 2$  ANOVA revealed no effects of judgment ( $F(1, 73) = 0.15, p = .700, \eta_p^2 = 0.001$ ) or age group ( $F(1, 73) = 0.02, p = .885, \eta_p^2 = .001$ ), and no interaction between both factors ( $F(1, 73) = 0.36, p = .548, \eta_p^2 = .001$ ). In the *poor-poor* condition, the ANOVA showed a main effect of age group ( $F(1, 73) = 53.78, p < .001, \eta_p^2 = 0.42$ ), but no effect of judgment ( $F(1, 73) = 0.001, p = .955, \eta_p^2 = 0.001$ ), and no interaction of both factors ( $F(1, 73) = 0.72, p = .398, \eta_p^2 = 0.01$ ). In the *poor-high* condition, the ANOVA showed main effects of judgment ( $F(1, 73) = 28.99, p < .001, \eta_p^2 = 0.28$ ) and age group ( $F(1, 73) = 33.65, p < .001, \eta_p^2 = 0.31$ ) as well as a significant interaction of both factors ( $F(1, 73) = 12.18, p < .001, \eta_p^2 = 0.14$ ). Post-hoc tests revealed that this interaction was due to the fact that only adults got significantly more confident after getting advice from the clear window (adults:  $t(34) = 5.43, p < .001, d_z = 0.93$ ; children:  $t(39) = 1.57, p = .124, d_z = 0.25$ ).



**Figure 5.** Visibility and confidence ratings before and after advice. Error bars indicate standard error of the means.

### 7.2.3. Discussion

The main findings of Experiment 1 included the following: In a judge-advisor experimental task designed to be suitable across age groups, adults revised their judgments functionally in regard to their own and the advisor's epistemic status. Furthermore, adults became more confident after receiving advice from the clear window when they themselves saw poorly. In contrast, children's advice-taking considered their own epistemic status (the better the child's visual access, the less advice it accepted) but not the advisor's epistemic status (children did not differentiate between well and poorly informed advisors). Children's confidence ratings of their judgments were generally higher than that of adults and, more importantly, were not influenced by the quality of the advice. In particular, unlike adults, children did not increase their confidence based on the availability of high quality advice under poor epistemic circumstances. Thereby, the confidence findings largely mirror those of advice-taking.

One interpretation of the present results regarding children's selective social belief revision is that 4- to 6-year-old children lack the conceptual competence to revise beliefs as a function of advice quality. This interpretation seems to contradict the outcomes of previous research on advice-taking in children reporting that children consider their advisor's expertise in a functional manner (Rakoczy et al., 2015). Alternatively, the present findings may not reflect competence but rather performance limitations due to extraneous demands of the

present experimental task. In fact, three task characteristics can be determined that may have increased task difficulty relative to the study of Rakoczy et al. (2015). Firstly, advice quality varied within the advisor instead of between several advisors. Thus, children needed to keep track of both their own and the advisor's perceptual access to determine how valuable the advice was in a given trial. Secondly, in the presented study continuous judgments were applied rather than having children select one out of a small set of options. Unlike in the case of disjunct options with one correct solution, two continuous judgments could be different from each other but equally inaccurate, which might be too challenging for children to understand. Finally, the response format allowed for a greater number of revision strategies. In the case of a limited set of answers, judges need to choose one, whereas in the continuous case, they can choose one judgment, average their initial judgment and the advice, or choose from an unlimited number of weighted averages.

All of these task features might have made it difficult for children to consider advice quality in their final judgments, even if they were principally able to selectively revise their beliefs in the light of differently accurate advisors. In order to investigate whether such performance limitations might potentially explain why children did not consider their advisor's visual access to the stimulus in Experiment 1, two simplifications to the original task were introduced: Firstly, a structurally analogous selective social learning task was implemented that did not involve any belief revision (Experiment 2). Secondly, the revision task was simplified by presenting a categorical instead of a continuous scale to the children (Experiment 3).

### **7.3. Experiment 2**

In Experiment 2, children were presented with similar material as in Study 1 except that they were not asked to revise existing beliefs, but instead expected to form new ones in response to information supplied by other agents with varying epistemic status. This task was, thus, structurally analogous to classical selective trust tasks widely used in recent research on social-cognitive development (see Harris, 2012), and is similar to the panel-of-experts paradigm used in research on adult judgment and decision-making (Mannes, 2009). Experiment 2 allows to test children's ability to infer the quality of advice when this quality varies within advisors.

### 7.3.1. Method

**7.3.1.1. Participants.** Forty 4- to 6-year-old children (age range: 49 - 72 months,  $M = 60.1$  months, 21 girls) were included in the final sample. Subjects were native German speakers with mixed socioeconomic backgrounds. Children were recruited from a database of families who had previously given consent to their participation. One additional child lost interest in the study from early on and could not be included in the final sample ( $n = 1$ ). Participants were tested individually in the lab.

**7.3.1.2. Material.** The same stimuli (box and the colored wooden sticks) as in Experiment 1 were used (see, Fig. 1, p. 38).

**7.3.1.3. Design and procedure.** The child was presented with two informants simultaneously: one with clear visual access and one with poor visual access. While the child was sitting at the opaque side of the box, it was first asked “Which of the two informants knows better where the line on the stick is? Who would you ask?” (Ask measure). In the second part of each trial, both informants handed over their judgment (the scale with the mark), the child compared both judgments and was asked to make a final decision about the right height of the black mark on a third unused scale (Endorse measure).

In Experiment 2, children were familiarized with stimuli and box and subsequently participated in eight test trials in total. Trials were of two kinds (four trials each). In half of the trials, the child was sitting at the opaque side of the box. In the other half of the trials, the child was initially sitting behind the blurred window together with the poor informant and was asked to closely look at the stick in the box and to identify the color of the stick. For the rest of these trials, the child moved to the opaque side. We conducted four trials of each kind, and the order of condition was counter-balanced. To avoid preference for one side, informants switched position every two trials. Data for Experiment 2 and 3 were collected within one test session lasting 30 min. All test sessions were videotaped. Children participated first in Experiment 3 and afterwards in Experiment 2.

The dependent measures were as follows:

- *Ask-measure:* Children were presented with a forced-choice question whom of the two informants they would ask about where to draw the black mark. We calculated the percentage of trials in which each child chose the respective informant.
- *Endorse-measure:* The mean number of trials (out of eight) for each of three categories (high quality informant, low quality informant, and between) was calculated. Children’s plotting within 2 cm around the judgment of the high or low

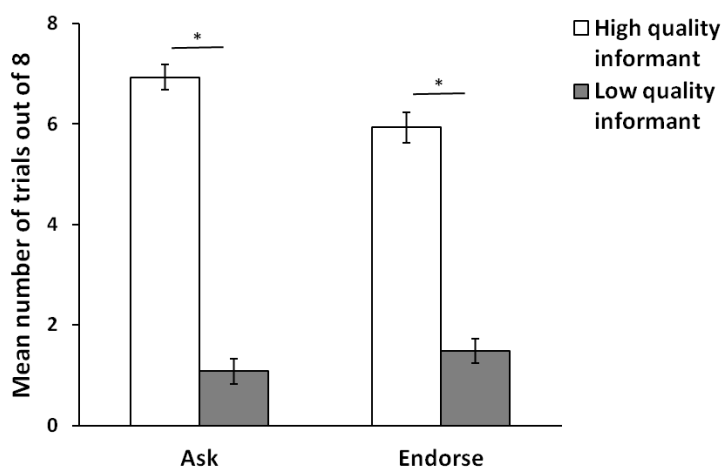
quality informant was coded as choice of this informant, respectively. Judgments in the interval between both radii were coded as “between”.

### 7.3.2. Results

Performance in choice between the two informants as a function of informant quality is depicted in Fig. 6.

0.6 % of the trials fall in the “between” category and were not included in our further analyses. We calculated a  $2 \times 2 \times 2$  ANOVA with position of the child (blurred window, backside), measure (ask, endorse) and quality of informant (poor, high) as within factors. We found main effects of the two factors measure ( $F(1, 38) = 7.91, p = .008, \eta_p^2 = 0.17$ ) and quality of informant ( $F(1, 38) = 140.20, p < .001, \eta_p^2 = 0.79$ ) and a significant interaction of both factors ( $F(1, 38) = 9.84, p = .003, \eta_p^2 = 0.21$ ). In contrast, we did not find a main effect of the position of the child ( $F(1, 38) = 0.35, p = .555, \eta_p^2 = 0.01$ ) and no interaction of position of the child with measure ( $F(1, 38) = 0.35, p = .555, \eta_p^2 = 0.01$ ) or with quality of the informant ( $F(1, 38) = 0.03, p = .863, \eta_p^2 = 0.001$ ).

Post-hoc tests revealed that children chose to ask the informant sitting behind the clear window significantly more often ( $t(38) = 11.62, p < .001, d_z = 1.86$ ) and also preferred to transfer the black mark of the high quality informant onto their own scale ( $t(38) = 8.88, p < .001, d_z = 1.42$ ). Furthermore, children chose to ask the informants sitting behind the clear window significantly more often than they did endorse their judgment ( $t(38) = -3.73, p = .001, d_z = -0.60$ ) but showed no significant difference between the ask and endorse measurement for the low quality informant ( $t(38) = 1.55, p = .129, d_z = 0.25$ ).



**Figure 6.** Ask and endorse measurement for children’s choice of the high or low quality informant. Both kinds of trials (child at the back side and child behind the blurred window) are summerized.

### 7.3.3. Discussion

In a selective trust version of the current task children, thus, did form new beliefs selectively as a function of information quality. This confirms their ability to infer the quality of advice not only in situations in which stable differences in expertise between two or more advisors exist, but also when advice quality varies within advisors as a situational factor. Thus, children's neglect of the advisor's visual access when accepting advice in Experiment 1 was most likely not based on some principal difficulty in distinguishing between perceptual information of varying quality in the present kind of task. Experiment 3 will investigate whether the continuous response format or the high number of possible revision strategies inherent to continuous belief revision (e.g., choosing one judgment, forming the statistical mean, or selecting from an infinite number of weighted averages) made the task in Experiment 1 overly difficult.

## 7.4. Experiment 3

In Experiment 3, the decision process was simplified by using categorical rather than continuous belief revision. The aim was to investigate whether 4- to 6-year-old children are able to revise existing beliefs as a function of their own and the advisor's epistemic state when task demands are reduced in terms of belief revision (compared to Experiment 1).

### 7.4.1. Method

**7.4.1.1. Participants.** The same forty 4- to 6-year-old children (age range: 49 - 72 months,  $M = 60.1$  months, 21 girls) as in Experiment 2 were included in the final sample.

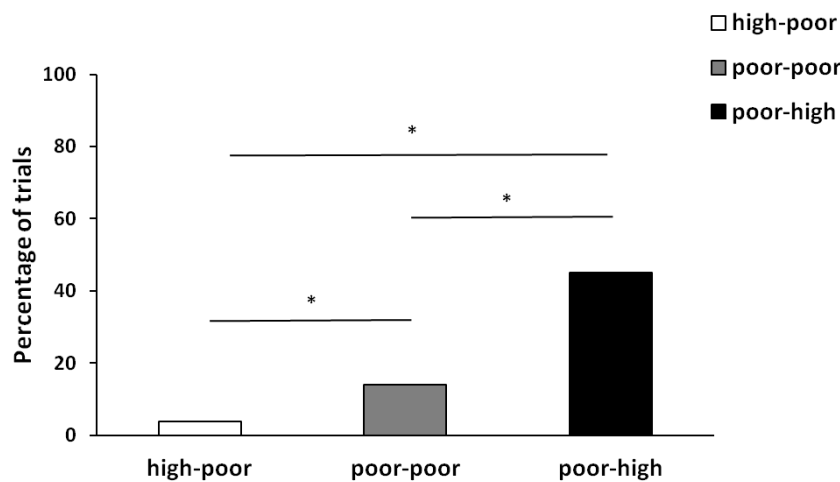
**7.4.1.2. Material.** In Experiment 3, we used the same box as in Experiment 1 and 2, but with new stimuli. Stimuli were five toy animals affixed on a cardboard strip in the order fox, pig, rabbit, sheep, and dog. Pictures of the stimuli were used as scales. During test trials, one of the five toy animals was marked with a black belt. Children were asked to identify this animal on a scale. Advice was randomized and produced by a random generator and secretly marked on the advisors' scale prior to the test sessions.

**7.4.1.3. Design and procedure.** Design and procedure were identical to Experiment 1 with the following exceptions: Firstly, only children were tested. Secondly, the judgments required were categorical (which animal has a mark?) rather than continuous (where on the stick is the mark?). Thirdly, children did not rate visibility and confidence.

The dependent measure was children's categorical advice-taking, that is, whether they adopted the judgment of the advisor or retained their initial judgment.

### 7.4.2. Results

The proportion of trials in which children engaged in advice-taking as a function of condition is depicted in Fig. 7. A Friedman test revealed differences in advice-taking between conditions ( $\chi^2(2) = 34.36, p < .001$ ). Post-hoc Wilcoxon rank tests revealed that children took more advice in the *poor-high* condition than in the other two conditions (*poor-poor* vs. *poor-high*:  $z = -2.12, p = .034, r = -0.34$ ; *poor-high* vs. *high-poor*:  $z = -4.3, p < .001, r = -0.68$ ), and more advice in the *poor-poor* condition than in the *high-poor* condition ( $z = -3.96, p < .001, r = -0.63$ ).



**Figure 7.** 100 % Advice-taking as a function of condition.

### 7.4.3. Discussion

The main finding of Experiment 3 includes that in a judge-advisor task with categorical (rather than continuous) belief revision, children did revise their judgments selectively as a function of their own and the advisor's epistemic status. This finding is consistent with the idea that children's apparent inability to consider the advisor's visual access in Experiment 1, might be caused by extraneous demands of the task rather than by a limited competence for belief revision. The requirement to make judgments on a continuous scale and then integrate two such continuous judgments rather than choosing one answer could cognitively overburden children in a way that they no longer consider the quality of advice despite their principal ability to do so.



## 8. Study 2: Argument evaluation in preschoolers

### 8.1. Introduction

Advice-taking research demonstrated judges' sensitivity to numerous characteristics of the advisor, for example, competence, age, and familiarity. In the classical paradigm used in this research, the JAS, judges aggregate their own judgment and an advice. Study 1 showed that 4- to 6-year-old children, just like adults, are sensitive to their advisor's state of knowledge and are able to weigh that knowledge against their own epistemic status. However, in daily life advice is most often given within a social context and supported by arguments. Study 2 was designed to investigate this aspect of decision-making, in particular, whether preschoolers are sensitive to advice justified by arguments differing in quality. Furthermore, the effect of the justification on subsequent advice utilization was determined. To this end children participated in three experiments.

In Experiment 1 and 2 of Study 2 the same visual perception task as in Study 1 was used. However, advisors supported their judgment reasonably regarding the question why they want the final judgment to be consistent with the advice. Children from around age two have been shown to be sensitive to argument quality (e.g., Castelai et al., 2018; Mercier, 2014). However, these studies concentrate on acquisition of new knowledge without the child having access to information on its own. The method used in this research is a classical selective trust paradigm in which two informants present conflicting information supported by different arguments. The effect of own knowledge on the evaluation of arguments has so far not been investigated. Furthermore, previous experimental studies mainly used circularity as a tool to investigate children's ability to weigh arguments. A circular argument fails to provide any new information about preferences in choosing a specific judgment. This makes it impossible to entangle whether children are able to weigh arguments based on different degrees of epistemic value, because the epistemic value of a circular argument is zero. Moreover, epistemic value of contrasted arguments was impaired by different lengths, verbal complexity or semantic structures. The current experiments go beyond previous research by investigating argument evaluation in the process of belief revision and contrast arguments based on epistemic content. Particular focus was placed upon children's understanding that the same argument in different contexts can be good, neutral, or even bad.

Experiment 3 furthermore investigated whether preschoolers infer a speaker's future credibility based on argumentative skills and whether information by insufficient argumentators might even be rejected. Therefore, children were presented with two formats

of a selective trust paradigm: a two-informant and a single-informant version. A single-informant paradigm allows to contrast two accounts: firstly, whether children just prefer information from qualified argumentators but judge differently skilled argumentators as generally trustworthy, or secondly, whether children might even actively reject information from insufficient argumentators and rather prefer an alternative option.

## 8.2. Experiment 1

### 8.2.1. Method

**8.2.1.1. Participants.** Thirty-nine 4- to 6-year-old children (age range: 48 - 71 months,  $M = 59.5$  months, 22 girls) were included in the final sample. Subjects were native German speakers with mixed socioeconomic backgrounds. Children were recruited from a database of families who had previously given consent to their participation. Participants were tested individually in the lab.

**8.2.1.2. Material.** The following material as in Study 1 was used: a three-sided box with one clear and one blurred window, colored wooden sticks, and analogue scales (see Fig. 1, p. 38). Advice, again, ranged from 1 to 200 mm. Advice was produced by a random generator and secretly marked on the advisor's scale prior to test sessions.

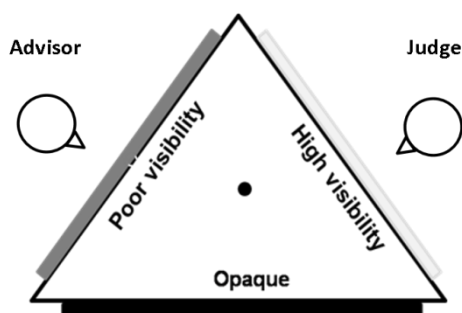
**8.2.1.3. Design.** The task was presented in JAS format with participants taking the role of judges. Confederates of the experimenter were playing a female puppet in the role of the advisor. Advice was presented with or without the argument, "*I think it is here because I looked through here.*" (German: "Ich denke es ist hier, weil ich hier durchgeschaut habe."). Phrasing of the argument stayed consistent across conditions but quality of the argument varied as a function of condition (high versus poor visual access of the advisor). We used a 2 (conditions)  $\times$  2 (argumentation) design with conditions (*high-poor* or *poor-high*) and argumentation (*with* or *without*) as within-subjects factors. All subjects received two trials of each condition with or without argument, respectively. Order of conditions was counterbalanced across subjects. However, the four trials without argument were always presented first.

**8.2.1.4. Procedure.** Every session started with an introduction phase prior to the test phase. In the introduction phase, participants were, in a first step, familiarized with the set-up material to make sure that they understood the design of the three-sided box and the concept of transferring a black mark on a colored stick onto a picture of the stick.

Additionally, data of the introduction phase were used as manipulation check to ensure that children's visibility was successfully manipulated by watching either through the clear or the blurred window.

*8.2.1.4.1. Introduction phase.* After entering the test room participants were invited to explore the three-sided box. The two windows were indicated and their perceptual quality assessed. To familiarize participants with the stimuli, they looked at one stick placed on a table and were invited to transfer the black mark as precisely as possible onto a picture of the stick. Participants' understanding of the scale as an identical image of the stimuli was ensured. Subsequently, participants judged one stimuli first behind the blurred and second behind the clear window. It was illustrated that the two judgments differ in quality.

*8.2.1.4.2. Test phase.* Judge and advisor were each placed on a different side of the box (for illustration, see Fig. 8). Then, one stick was placed in the box (in such a way that subjects could not observe the placement), and both judge and advisor were asked to make their initial judgments. After that, judges were seated at the opaque side of the box, received the scale of the advisor in addition to their own scale, and were asked to mark their final judgment on a third scale. Every two trials judge and advisor switched their starting positions (initial judgment behind the clear or blurred window). In the first four trials, only the advice was presented while in the second four trials the advice was supported by an argument. All test sessions were video recorded and lasted approximately 25 min.



**Figure 8.** Starting positions in the *high-poor* condition. The child (judge) is sitting behind the clear window and the advisor (puppet) is sitting behind the blurred window.

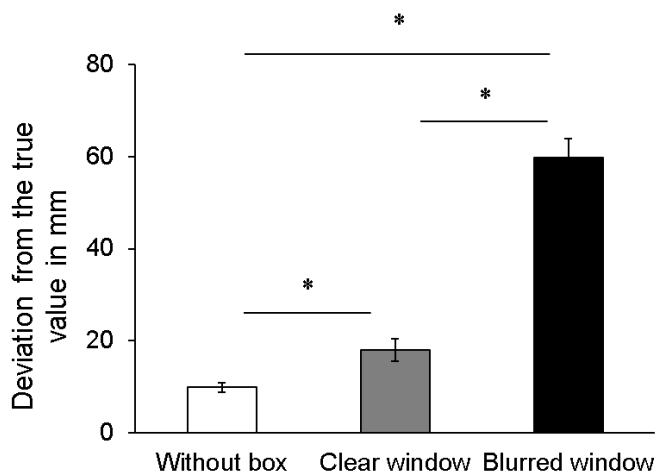
The dependent measures were as follows:

- *Deviation from the true value:* The same measure as in Study 1 (Experiment 1) was used. For definition of the measure see p. 40.
- *Advice-taking (AT):* The same advice-taking measure as in Study 1 (Experiment 1) was used. For definition of the measure see p. 40.

## 8.2.2. Results

**8.2.2.1. Preliminary analyses.** Analysis of the manipulation check showed that the design of the set-up was valid. The participants' visibility was successfully manipulated by presenting them with a clear and a blurred window. Children's deviation from the true value when sitting at the table, behind the clear, and behind the blurred window are depicted in Fig. 9.

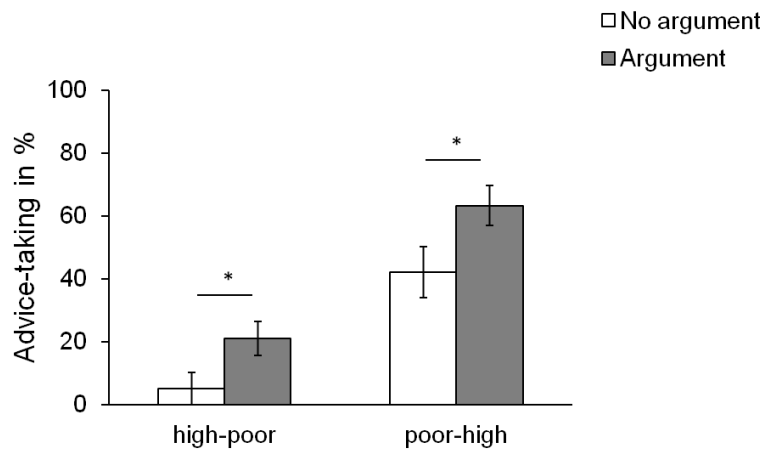
An ANOVA for repeated measures with form of visual access (without box, clear or blurred window) as within-subjects variable was used to analyze whether there was a significant difference in the deviation from the true value between the two windows. The deviation from the true value when sitting at the table was included for reasons of completeness. The ANOVA revealed a significant main effect of variation of visual access ( $F(2, 76) = 81,12, p < .001, \eta_p^2 = 0.68$ ). Post hoc paired sample  $t$ -tests revealed significant differences between all three variations of visual access (without box vs. clear window:  $t(38) = -3.46, p = .001, d_z = -0.56$ ; without box vs. blurred window:  $t(38) = -11,16, p < .001, d_z = -1.80$ ; clear vs. blurred window:  $t(38) = 7.96, p < .001, d_z = 1.29$ ).



**Figure 9.** Deviation from the true value as a function of form of visual access (without box, clear window and blurred window). Error bars indicate standard error of the means.

**8.2.2.2. Main analyses.** Performance in advice-taking as a function of condition and argumentation is depicted in Fig. 10. A  $2 \times 2$  ANOVA for repeated measures with condition (*high-poor* or *poor-high*) and argumentation (*with* or *without argument*) as within-subjects variables was calculated. The ANOVA showed significant main effects of condition ( $F(1, 37) = 38.35, p < .001, \eta_p^2 = 0.51$ ) and argumentation ( $F(1, 37) = 14.67, p < .001, \eta_p^2 = 0.28$ ) but no interaction of both factors ( $F(1, 37) = 0.17, p = .681, \eta_p^2 = 0.01$ ). Post hoc tests revealed a significant increase of advice-taking in both conditions when advice was accompanied by

agumentation (*high-poor* condition:  $t(37) = -2.32$ ,  $p = .026$ ,  $d_z = -0.38$ ; *poor-high* condition:  $t(38) = -2.71$ ,  $p = .010$ ,  $d_z = -0.44$ ).



**Figure 10.** Advice-taking as a function of condition and argumentation. Error bars indicate standard error of the means.

### 8.2.3. Discussion

The main findings of Experiment 1 were firstly, that children weigh advice, given by a puppet without any argument as a function of their own and the advisor's state of knowledge: They accepted advice more likely when the advisor was better informed than themselves and less likely when they themselves were better informed. This result confirms findings of Study 1. Secondly, supporting advice with an argument increased children's advice-taking significantly compared to situation in which advice was given without justification. However, this increase in advice-taking did not vary as a function of argument quality. The advisor who justified the initial judgment increased advice-taking similarly regardless of the own good or poor visual access. This result is surprising within the available literature because it shows preschooler's competence in argument evaluation (e.g., Mercier et al., 2014). Children's competence might have been diminished by two factors, the perceived quality of the argument and the consistency of the advisor.

Argument quality was manipulated by the informational access of the advisor, while phrasing of the argument stayed constant across conditions. Thus, children were able to access argument quality only as a function of context (condition). More closely considered, phrasing of the argument might have been too neutral in order to be perceived as differing in quality between the two conditions: high quality when the advisor had good visual access and low quality when the advisor had poor visual access. Children might have interpreted the content of the argument "*I think it is here because I looked through here*" differently between

conditions (which was the idea of the experiment), however, meaningfully in both situations. In the *poor-high* condition children might have been reminded that the advisor had clear visibility in a way such as *“This is a good advice because I saw it clearly”*. In the *high-poor* condition they might have interpreted the same argument as sign of good intention in the sense of *“This is a good advice, because I tried the best I can”*.

An additional aspect of the experimental design might have made it artificially difficult for children to increase advice-taking as a function of argument quality: advice was consistently given by the same person (puppet). Results in literature so far have not shown the ability of preschoolers to evaluate argumentative skills detached from the speaker. Children might just form an overall impression of one’s argumentative skills without adjusting to the informant’s argumentative competence context specifically.

In order to investigate whether these aspects of the experimental design limited children’s ability to evaluate argument quality in Experiment 1, the original task was simplified in two ways: Firstly, advisors supported their judgment presenting an argument of increased strength and secondly, two different puppets were used under the *high-poor* and *poor-high* condition.

### **8.3. Experiment 2**

In this Experiment, therefore, the same perceptual judgment task was used as in Experiment 1, however, with two modifications. The argument was phrased more precisely by increasing the speaker’s commitment of advisors: *“I really want us to take this, I am absolutely sure because I looked through here”*. Under the *poor-high* condition the increased argument strength supported the advice and potentially resulted in increased advice-taking. In contrast, under the *high-poor* condition increased argument strength weakened the advice by revealing the advisor’s own poor competence and consequently, making an increase in advice-taking irrational. Furthermore, the consistency of the advice source was manipulated in that two puppets were introduced and each puppet presented the advice either under the *high-poor* or *poor-high* condition.

#### **8.3.1. Method**

**8.3.1.1. Participants.** Thirty-seven 4- to 6-year-old children (age range: 48 - 71 months,  $M = 58.8$  months, 20 girls) were included in the final sample. Subjects were native German speakers with mixed socioeconomic backgrounds. Children were recruited from a

database of families who had previously given consent to their participation. Participants were tested individually in the lab.

**8.3.1.2. Material.** The same stimuli (box and the colored wooden sticks) as in Experiment 1 were used with the exception that every participant judged only four wooden sticks. One half of the participants judged set A (brown, multi-colored, green and black-white) and the other half judged set B (purple, black, blue and turquoise) of the wooden sticks.

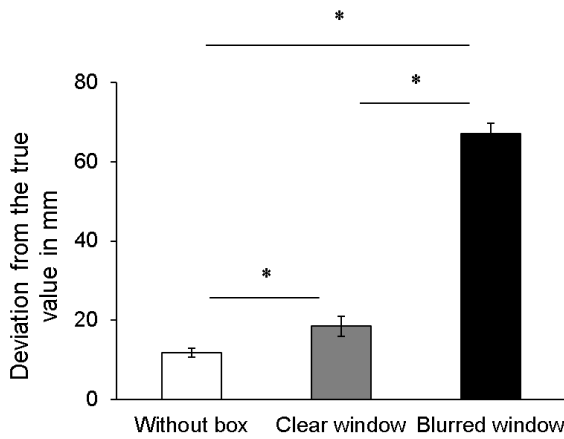
**8.3.1.3. Design.** The task was presented in JAS format with participants taking the role of judges. Advisors were confederates of the experimenter playing two puppets (A and B) who presented the judges with the randomized advice. Advice was presented with the argument, "*I really want us to take this there. I am absolutely sure because I looked through here.*". Phrasing of the argument stayed consistent across conditions, but quality of the argument varied as a function of condition (high versus poor visual access of the advisor). Advice was given by puppet A in one condition (*high-poor*) and by puppet B in the other condition (*poor-high*). All subjects received two trials of each condition with argument. Order of conditions was counterbalanced across subjects.

Data of the *high-poor* and *poor-high* conditions without argument collected in Experiment 1 were reanalyzed using a 2 (conditions) × 2 (argumentation) design with conditions (*high-poor* or *poor-high*) and argumentation (*with* or *without*) as within-subjects factors.

**8.3.1.4. Procedure.** The procedure was identical to Experiment 1 with the following exceptions during the test phase: (1) children participated in four trials with argument. (2) The argument was phrased, "*I really want us to take this there. I am absolutely sure because I looked through here.*". (3) The advice accompanied by the argument was presented by different puppets in the two conditions.

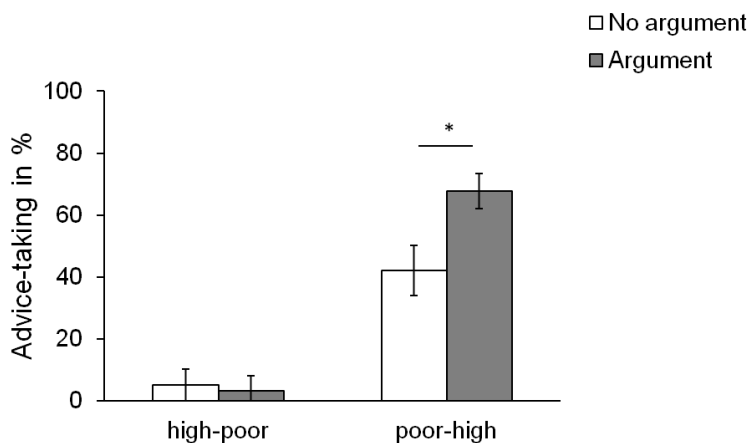
## 8.3.2. Results

**8.3.2.1. Preliminary analyses.** Children's deviation from the true value sitting at the table, behind the clear and behind the blurred window are depicted in Fig. 11. Analysis of the manipulation check showed that visual access of the participants was manipulated as intended. An ANOVA showed a significant main effect of form of visual access ( $F(2) = 150.06, p < .001, \eta_p^2 = 0.81$ ). Post hoc paired sample *t*-tests revealed significant differences between all three forms of visual access (without box vs. clear window:  $t(36) = -2.67, p = .011, d_z = -0.44$ ; without box vs. blurred window:  $t(36) = -17.94, p < .001, d_z = -2.99$ ; clear vs. blurred window:  $t(36) = 10.67, p < .001, d_z = 1.78$ ).



**Figure 11.** Deviation from the true value as a function of form of visual access (without box, clear window and blurred window). Error bars indicate standard error of the means.

**8.3.2.2. Main analyses.** Performance in advice-taking as a function of condition and argumentation is depicted in Fig. 12. A  $2 \times 2$  ANOVA for repeated measures with condition (*high-poor* or *poor-high*) and argumentation (*with* or *without argument*) as within-subjects variables was calculated. The ANOVA showed a significant main effect of condition ( $F(1, 74) = 87.97, p < .001, \eta_p^2 = 0.54$ ) and an interaction of condition and argumentation ( $F(1, 74) = 4.48, p = .038, \eta_p^2 = 0.06$ ). Post hoc tests revealed a significant increase of advice-taking when advice was accompanied by the argument in the *poor-high* condition only (*high-poor* condition:  $t(74) = -0.10, p = .922, d = -0.23$ ; *poor-high* condition:  $t(74) = -2.46, p = .017, d = -0.60$ ).



**Figure 12.** Advice-taking as a function of condition and argumentation. Error bars indicate standard error of the means.



### 8.3.3. Discussion

The main finding of Experiment 2 demonstrated that children varied in their degree of advice-taking as a function of argument quality implemented through high or poor visual access of the advisor. An advisor having good visual access and supporting advice with a strong argument significantly increased advice-taking by children. In contrast, an advisor having poor visual access and supporting advice with an identically phrased argument failed to increase advice-taking. Advice-taking remained at the same level as if advice was not supported by an argument at all. More importantly, in the current design of the task an decrease in advice-taking in the *high-poor* condition would be irrational. Children looked through the clear window to make their initial judgment, thus, the correct solution was already revealed. In order to perform negative advice-taking (to change the initial judgment into the opposite direction of the advice) children would have to downgrade their initial judgment. The question of whether children would actively discount an informant based on lacking argumentative skills was investigated in Experiment 3.

### 8.4. Experiment 3

The main aim of Experiment 3 was to investigate whether preschooler were able to judge wrong argumentation negatively and whether children use argumentative skills of informants to make inferences about their future credibility. If so, the question arises how strongly lacking skills affect children's selective trust in informants? And furthermore, do children prefer judgments of the informant who had previously provided strong over weak arguments or do they even actively discount informants based on previous weak argumentation?

In order to answer these questions a new decision task was designed following a classical selective trust paradigm. Traditionally such paradigms present two informants giving conflicting testimony and comparing children's learning from an accurate versus inaccurate speaker. This design has been adopted in the first part of Experiment 3, in which reliability of speakers was established in form of strong or weak argumentative skills.

The next part of Experiment 3 investigated how children judge speaker's trustworthiness based on argumentative skills more in detail. There are two possibilities: children may consider both speakers as generally trustworthy but prefer information from the speaker that previously provided improved support of their judgment. Alternatively, children may perceive a speaker who gave insufficient evidence for own judgments as unreliable such that they reject information from this informant. In order to distinguish between these two possible

accounts the same decision task was presented as single-informant design. The first account should predict whether children learn from both speakers if they were the only source of information available. The second account should predict whether children acquire information only from a single accurate informant but reject information from a single inaccurate informant.

#### **8.4.1. Method**

**8.4.1.1. Participants.** Forty-one 4- to 6-year-old children (age range: 48 - 71 months,  $M = 59,5$  months, 20 girls) were included in the final sample. One additional girl lost interest during the familiarization phase and did not participate in the experimental trials. Subjects were native German speakers with mixed socioeconomic backgrounds. Children were recruited from a database of families who had previously given consent to their participation. Participants were tested individually in the lab.

**8.4.1.2. Material.** We presented participants with a forced choice design in which they were to decide about the color or softness of a hidden object. Six plush toys of 7 cm length were used (see Fig. 13, c), two of them were always of the same color (blue, pink or white) but different in softness (soft or hard). Toys were placed in a triangular box (edge length of triangle: 57 cm, height: 44 cm). The box was a modified version of the material used in Experiment 1 and 2. It consisted of a clear window side for seeing (window measurements: 27.5 cm x 23.5 cm), a side with an opening for feeling (radius of opening: 12 cm, Fig. 13, a) and an opaque backside. Little cases were used to move the toys secretly into the triangular box. Assistants presenting the two choices were wearing identical t-shirts that differed in color, the first assistant was wearing a green and the second a yellow t-shirt. To allow the children to visualize choice options and avoid working memory to limit intentional choices, presented information and arguments were illustrated using memory cards (see Fig. 13, b): The three colors (blue, pink or white) that were used were depicted as filled circle of each color, where as the two levels of softness were depicted as pillow for being soft and stones for being hard. The modalities “seeing” and “feeling” were visualized using memory cards depicting a pair of eyes or a hand.



**Figure 13.** Three-sided box with opening for feeling (a), memory cards (b) and plush toys (c).

**8.4.1.3. Design.** This experiment employed a 2 (informant: accurate vs. inaccurate)  $\times$  2 (modality: visibility vs. touch) within subject design. The order in which the informants presented their judgment, on which side each informant was standing when presenting the judgment (left or right), and which of the two assistants was playing the accurate or inaccurate informant was counter-balanced. All children received three kinds of practice trials (modality, box, and ignorance) within the familiarization phase, ten test trials (four argumentation trials, four single-informant trials, and two two-informant trials) within the test phase, and three explicit judgment questions at the end of the test phase.

**8.4.1.4. Procedure.** Participants started with a familiarization to the set-up material. In this familiarization phase it was ensured that children knew the three different colors requested in the test trials and were able to assign the two levels of softness to the toys. Furthermore, children demonstrated their understanding of the toys' characteristics and the modality necessary to assign the characteristics correctly in three kinds of practice trials used as manipulation check to verify the validity of the set-up. Subsequently to the familiarization phase, children participated in the test phase. Every test session was video recorded and lasted approximately 30 min.

**8.4.1.4.1. Introduction phase.** In a first step, we introduced toys, memory cards and hiding boxes. Children were presented with six toys located on a table and asked to name their colors. The three memory cards for the colors blue, pink and white were placed on the table and children were asked to classify the six toys into three pairs of different color. Children were instructed as follows: the experimenter placed the eye-card on the table and said while pointing at the eye-card, "Now look at these toys and put all the blue ones onto the blue-color-card.". The experimenter repeated these instructions for the pink and the white

colored toys. After the children had sorted the toys, it was indicated that two toys of one color are not identical but differ in softness: one is soft and the other one is hard. The two memory cards for softness, a picture of stones and a picture of a pillow, were placed on the table using the following phrasing: “These are stones, they are hard and this is a pillow, it is soft.” Subsequently, the experimenter asked the children to sort the toys in terms of softness. The hand-card was placed on the table and children were instructed by pointing to “feel the toys and put all the hard ones on the stone-card and all the soft ones on the pillow-card”. The first step of the familiarization phase was completed by practice trial 1 and 2 (modality trials). In practice trial 1, two toys of the same softness but differing in color were hidden in two separate small opaque boxes. One of the boxes was secretly selected under the table, put in front of the child and the child was asked, “If you want to know the color of the toy in the box, do you need to look at it or do you need to feel it?” In practice trial 2, the procedure was repeated with two toys of the same color but differing in softness. If a child failed to answer a question correctly, the practice trial was repeated once, before the experimenter moved on.

In a second step, children were introduced to the triangular box, invited to explore the box from all sides and asked to identify from which side one can look into the box and from which side one can feel something in the box. Subsequently, children participated in the practice trials 3 – 6 to control whether they established a causal connection between modality and characteristic. At the beginning of each trial children were seated at the backside of the box not able to see or feel inside the box. Two toys were presented and one was secretly moved into the box. The two toys were either of the same softness and different color (trial 3 and 5) or of the same color and different softness (trial 4 and 6). In trial 3 and 4 (box trials) children were firstly asked for the modality necessary to identify the specific characteristic of the hidden toy (seeing or feeling) and, secondly, which side of the box would allow for the informational access necessary. After answering these questions, children were allowed to approach the announced side of the box and identify the characteristic of the toy. In trial 5 and 6 (ignorance trials) children were offered the wrong informational access, for example, in the softness trial they were placed at the window side and allowed to look at the toy but asked for the softness in form of the forced choice question, “Is it hard, is it soft or do you not know?”. If a child failed to answer a question correctly, a practice trial was repeated once, before the experimenter moved on.

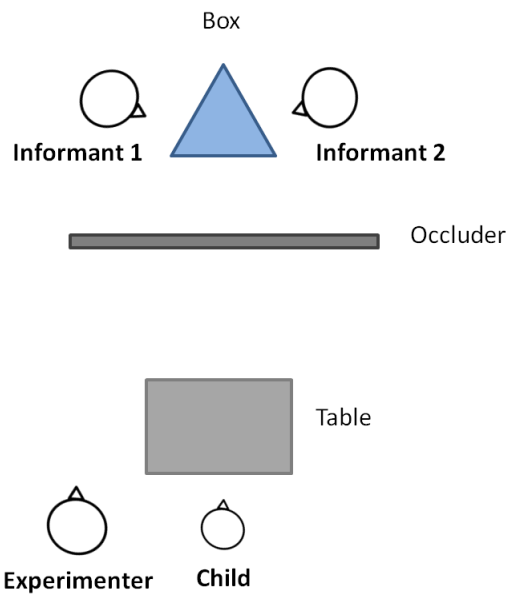
The dependent measures were as follows:

- *Modality*: Represents the number of trials for each of the two categories: right and wrong answer. An answer was coded as “right” when announcing the informative

modality (looking vs. feeling) for the characteristic in question (color vs. softness). The “wrong” category included answers that mismatched modality and characteristic.

- *Box*: Represents the number of trials for each of the two categories: right and wrong answer. Answers announcing the informative side of the box for the characteristic in question (window side for color and opening side for softness) were coded as “right” and answers announcing the not informative side of the box for the characteristic in question (window side for softness and opening side for color) were coded as “wrong” answer.
- *Ignorance*: Represents the number of trials for each of the two categories: right and wrong answer. “I do not know” answers for the forced choice questions were coded as “right” other answers were coded as wrong.

*8.4.1.4.2. Test phase.* In the beginning of the test phase children were informed that they participate in a quiz show and will answer questions about which of two toys was hidden in the triangular box. Two girls (one wearing a green and one wearing a yellow t-shirt) will join the game to assist. The test phase consisted of three kinds of trials in the following order: four argumentation trials, four single-informant trials, and two two-informant trials. The overall procedure was identical for all trials (set-up see Fig. 14). Children sat at a table with an occluder board in front of them hiding the triangular box. The assistants were placed side by side with the child. The experimenter presented two toys. In color-trials the toys were of the same softness but differed in color and in softness-trials toys were of the same color but differed in softness. Both toys were moved into small boxes and by walking behind the occluder board the experimenter announced to “secretly hide one of the toys in the triangular box”. Being behind the occluder the experimenter called one (or both) assistant(s) behind the occluder. Assistants were instructed to get information from one side of the box. For children the instruction was hearable but not visible, therefore, they knew assistants were allowed to see or feel inside the box but did not know which of the two had visual or tactile access to the toy. Assistants again appeared in front of the occluder and presented their information about the characteristic in question. In the end of each trial children answered a forced choice question about the two possible answers. The specific procedure of the three kinds of test trials was as follows.



**Figure 14.** Information phase. Obtaining visual and tactile information of both informants, not visible for the child.

In argumentation trials two informants presented conflicting answers accompanied by an argument. The rationale of argumentation trials was to investigate children's understanding of arguments differing in quality and to establish one assistant as a high quality informant and the other assistant as low quality informant. Two color-trials and two softness-trials were conducted. Behind the occluder board assistants were instructed by "you go to the window side and you go to the opening side". The following phrasing was used to present conflicting information: "I think it is *color/softness*, because I *looked/felt* into the box". Informants used memory cards to illustrate the information presented. Visual information stayed present upon the child's decision. The phrasing of arguments stayed consistent across the trials but differed in quality as a function of the characteristic in question. For example, the argument "because I looked into the box" was a high quality argument in color-trials but a disqualifying argument in softness-trials. Informants stayed consistent in argument quality across four trials, informant A presented high quality arguments and informant B presented disqualifying arguments.

In single-informant trials children were faced with one informant only. In four trials each assistant gave information twice, once in a color-trial and once in a softness-trial. Behind the occluder board the single informant was instructed with the sentence "you go to this one side of the box". The information was given, phrased as "I think it is *color/softness*", after announcing the statement the assistant disappeared behind the occluder.

In two-informant trials two informants presented conflicting answers. Two trials were conducted, one color-trial and one softness-trial. Behind the occluder board, assistants were instructed by “You go to the window side and you go to the opening side.”. Both assistants presented their information phrased as “I think it is *color/softness*.”. For example, in the color-trial when a blue and a pink toy were used, one informant stated “I think it is blue.” and the second informant stated “I think it is pink.”.

At the end of the test phase, the assistants left the test room and the experimenter asked three explicit judgment questions. Firstly, children were asked “How did the assistant in the green t-shirt give reasons for her judgment, good or not so good?”. Secondly, the experimenter asked “How did the assistant in the yellow t-shirt give reasons for her judgment, good or not so good?”. And finally, children were asked to answer the question “Who gave better reasons, the assistant in the green or the one in the yellow t-shirt?”.

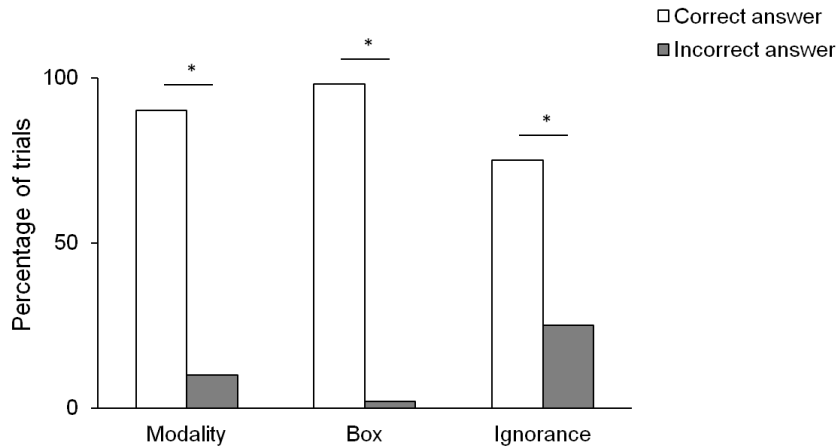
The dependent measures were as follows:

- *Contrast of argument*: Represents the number of trials out of two trials in which children chose the judgment accompanied by the right or wrong argument. In color-trials the argument “because I looked into the box” was coded as right argument and the argument “because I felt into the box” was coded as wrong argument. In softness-trials correctness was coded vice versa.
- *Choice of single-informant*: Represents the percentage of trials in which children responded testimony-consistent and testimony-inconsistent (choice of the other alternative, not recommended by the informant).
- *Contrast of informants*: Represents the percentage of trials in which children chose testimony of the informant previously presenting right or wrong arguments.
- *Explicit judgments of single-informants*: Represents the percentage of answers for the two categories: “good” and “not so good” for each informant.
- *Contrast of explicit judgments*: Represents the percentage of answers favoring the informant previously presenting right or wrong arguments.

## 8.4.2. Results

**8.4.2.1. Preliminary analyses.** Performance in the three kinds of familiarization trials is depicted in Fig. 15. Binomial tests were used to compare children’s performance with chance level (50 %) and revealed that children answered correctly above chance level for all three kinds of questions. This manipulation check ensured validity of the experimental set-up. Firstly, children correctly chose the right modality (looking or feeling) to figure out about the

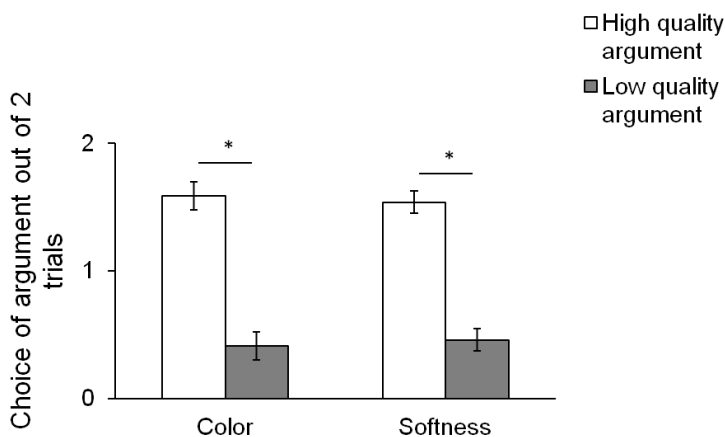
characteristic in question (binomial,  $ps < .001$ ). Secondly, when presented with the triangular box, they chose the correct informational access (window or opening) to find out about the color or softness of the toy (binomial,  $ps < .001$ ) and thirdly, acknowledged ignorance about the characteristic in question when presented with the wrong informational access above chance level (binomial,  $ps = .002$ ).



**Figure 15.** Proportion of familiarization questions answered correctly or incorrectly.

#### 8.4.2.2. Main analyses.

8.4.2.2.1. *Contrast of arguments.* Children's choice of the high or low quality argument as a function of modality is depicted in Fig. 16. Paired  $t$ -tests revealed that children preferred the informant giving a high quality argument compared to the informant giving a low quality argument for both characteristics color ( $t(40) = 5.31, p < .001, d_z = 0.84$ ) and softness ( $t(40) = 5.77, p < .001, d_z = 0.91$ ). This preference did not differ in degree between both characteristics ( $t(40) = 0.35, p = .728, d_z = 0.05$ ).

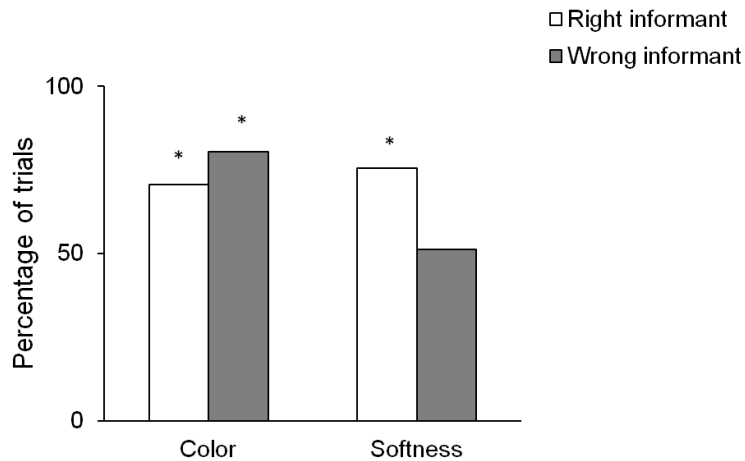


**Figure 16.** Mean number of chosen high quality arguments out of two trials (compared to choices of the low quality argument) as a function of characteristic. Error bars indicate standard error of the means.



8.4.2.2.2. *Choice of single-informant.* The proportion of testimony-consistent responses as a function of characteristic is depicted in Fig.17. The hypothesis was that children's preference for testimony-consistent choices would not differ between modalities for the informant previously giving high quality arguments (right informant) and the informant previously giving low quality arguments (wrong informant). To examine these two predictions, testimony-consistent responses were coded as 1 and testimony-inconsistent responses were coded as 0. The proportions of responses across participants were calculated as dependent variable. Binomial logistic regressions predicted children's responses with the predictor modality (visibility-color vs. touch-softness). The model predicting testimony-consistent choices when informed by the informant previously giving high quality arguments did not predict children's choices ( $R^2 = 0.004$ ,  $\chi^2(1) = 0.25$ ,  $p = .618$ ,  $f = 0.06$ ). Consistent with the hypothesis, children's willingness to accept the testimony of the right informant did not differ between color-trials and softness-trials. In contrast, the model predicting testimony-consistent choices when informed by the wrong informant did predict children's choices ( $R^2 = 0.13$ ,  $\chi^2(1) = 8.00$ ,  $p = .005$ ,  $f = 0.34$ ). There was a significant effect of modality ( $b = -1.37$ ,  $p = .007$ , *odds ratio* = 0.25, 95 % *CI* [0.09, 0.68]). However, children's willingness to accept the testimony of the informants previously giving low quality arguments did differ between color-trials and softness-trials.

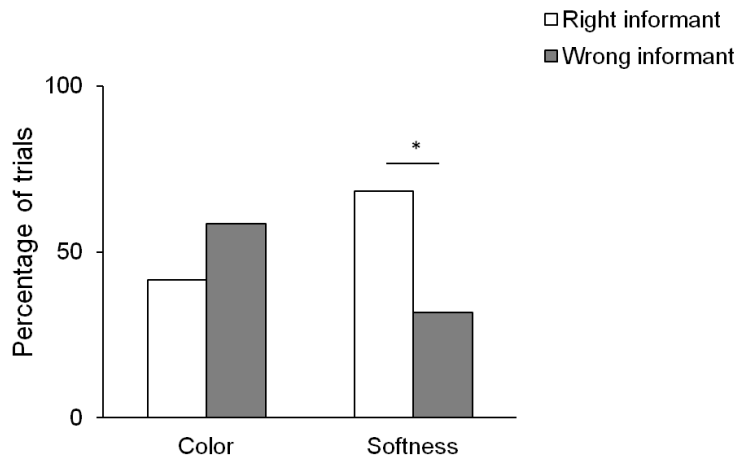
Binomial tests revealed that children responded testimony-consistent about chance level (50 %) in color-trials and softness-trials when presented with the right informant (color: binomial,  $ps = .012$ ; softness: binomial,  $ps = .001$ ). However, children responded testimony-consistent about chance level (50 %) when presented with the wrong informant in color-trials only (binomial,  $ps < .001$ ). This was not true for softness-trials, here children's testimony-consistent responses were at chance level (binomial,  $ps = 1.000$ ). These results are inconsistent with the hypothesis that children reject an informant previously given low quality arguments.



**Figure 17.** Proportions of testimony-consistent responses as a function of informant and modality.

8.4.2.2.3. *Contrast of informants.* Children's choices of the right (previously high quality argument) and wrong informant (previously low quality argument) as a function of characteristic is depicted in Fig. 18. The hypothesis was that children prefer the right informant in color- and softness-trials to the same degree. To examine this prediction, choices of the right informant were coded as 1 and choices of the wrong informant were coded as 0. The proportions of choices across participants were calculated as dependent variable. A binomial logistic regression predicted children's choices with the predictor modality (visibility-color vs. touch-softness). The examined models predicted children's choices of the right informant as a function of modality ( $R^2 = 0.09$ ,  $\chi^2(1) = 6.04$ ,  $p = .014$ ,  $f = 0.32$ ). There was a significant effect of modality ( $b = 1.11$ ,  $p = .016$ , *odds ratio* = 3.04, 95 % *CI* [1.23, 7.51]). In contrast to the hypothesis, results of the logistic regression showed that children's preference for the right informant differed between color- and softness-trials.

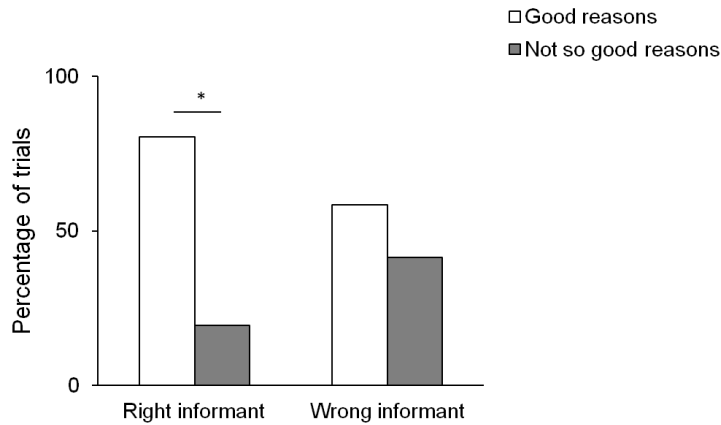
Furthermore, the tendency of children to choose the right informant was compared with chance (50 %) for each modality. In color-trials children chose the right informant at chance level (binomial,  $ps = .349$ ). In contrast, in softness-trial children preferred the right informant more often than chance (binomial,  $ps = .028$ ).



**Figure 18.** Preference of the right informant (previously giving the high quality argument) as a function of modality.

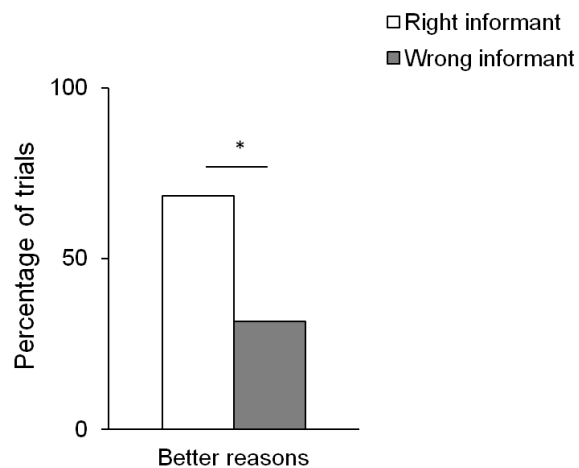
8.4.2.2.4. *Explicit judgments of single-informants.* Explicit judgments as a function of informant are depicted in Fig. 19. The hypothesis was that children judge the right informant more often as giving good reasons for her testimony while judging the wrong informant as giving more often not so good reasons for her testimony. To examine this prediction, “good” judgments were coded as 1 and “not so good” judgments were coded as 0. The proportions of judgments across participants were calculated as dependent variable. A binomial logistic regression predicted children’s judgments with the predictor informant (right vs. wrong informant). The examined model predicted children’s judgments as a function of informant ( $R^2 = 0.08$ ,  $\chi^2(1) = 4.74$ ,  $p = .029$ ,  $f = 0.29$ ). There was a significant effect of informant ( $b = -1.07$ ,  $p = .034$ , *odds ratio* = 0.34, 95 % *CI* [0.13, 0.92]). Consistent with the hypothesis, children judged the right informant more often as giving “good reasons” for testimony than the wrong informant.

Furthermore, the tendency of children to judge the right informant as giving good reasons for her testimony was compared with chance (50 %). Children judged the right informant as giving good reasons for her testimony more often than chance level (binomial,  $ps < .001$ ). In contrast, children’s positive judgment of the wrong informant was at chance level (binomial,  $ps = .349$ ). However, children did not judge the wrong informant as giving „not so good“ reasons for testimony.



**Figure 19.** Children’s explicit judgments of informants (informant previously giving high quality arguments vs. informant previously giving low quality arguments) as giving “good” or “not so good” reasons for their testimonies.

8.4.2.2.5. *Contrast of explicit judgments.* Children’s judgments of the right or wrong informant when presented in a contrast question are depicted in Fig. 20. Proportions of answers were analyzed using a binomial test. Analysis revealed a significant preference to judge the right informant (previously giving high quality arguments) for giving better reasons for testimony compared to the wrong informant (previously giving low quality arguments).



**Figure 20.** Proportions of positive judgments (as informant giving better reasons for her testimony) as a function of informant (informant previously giving high quality arguments vs. informant previously giving low quality arguments).

### 8.4.3. Discussion

As demonstrated in the manipulation check phase participants were able to solve the decision task presented in the current paradigm. They distinguished modalities, matched them correctly with characteristics of toys and admitted ignorance when presented with the

insufficient perceptual access. The main findings of Experiment 3 were the following: Firstly, children did weigh information as a function of argument quality. Arguments phrased similarly were endorsed differently as a function of context. Children preferred information on color or softness when supported by evidence of the correct modality. For example, they endorsed the information “it is a blue toy” when the informant supported own judgments with the argument “because I looked into the box” but did not so when information on color was supported with the argument “because I felt insight the box”. However, they preferred the later argument when asked for the softness of the toy. This result supports the fundamental skill of argument evaluation in 4- to 6-year-old children. Argument structure was kept constant across conditions leading to the same argument being a weak opposed to a strong argument depending on the context of presentation. This supports children’s ability to judge arguments based on the epistemic relationship between information and argument instead of just blindly following connectives and discourse markers like *because* or *and* (Bernard et al., 2012; Leman, 2002).

Secondly, in single-informat trials children did not actively reject information from speakers previously giving weak arguments. However, to show this kind of response behavior in the current design would have required children to reject the only information available, since only one informant participated in these trials and children had no perceptual access themselves. Subsequently, children would need to develop a strong negative judgment about the weak informant based on own mismatching argumentation in the first four trials of the experiment. They would need to transfer this kind of negative evaluation to a situation in which this informant did not give an argument at all and was presented without contrast or reference point for reliability of the own judgment. In the color trials children chose both informants significantly more than 50 % and consistently with testimony. In contrast, in the softness trials children preferred the judgment of the previously strong argumentator. When faced with the previously weak argumentator children did not prefer given testimony over the alternative answer (e.g., the *soft* toy when the informant recommended the *hard* toy). This result suggests that children might use their argumentative ability as cue for the informant’s reliability in some circumstances but not in others. The developing ability might also be still fragile and easily masked by performance factors.

Thirdly, when presented with conflicting testimony of the previously strong versus weak argumentator children preferred the accurate informant compared to the inaccurate informant, however, only for the characteristic softness. In the color trials children did not develop any preference for one or the other speaker. This finding, again, supports different response patterns in children as a function of modality: when deciding about the softness of

a hidden object children transferred previously established trust based on argumentative ability and used it to infer about informants credibility even when speakers did not give arguments to support their information. In contrast, with respect to the color of the toys children did not develop a preference for either one of the informants. This difference in modality is surprising and calls for more research data to disentangle factors and mechanisms underlying this gradual trust in informants as function of their argumentative skills.

Finally, children were significantly more likely to judge the accurate instead of the inaccurate informant as having good argumentative skills in the single-informant question. Consistent with this answer pattern they judged the accurate informant as the one giving “better” reasons for the own testimony versus the inaccurate informant. Thus, children’s explicit judgments showed their ability of remembering the informants differently with respect to their argument quality and by whom arguments were presented (accurate or inaccurate informant). However, children seemed to explicitly avoid judging the inaccurate informant as the one giving “not so good” reasons but rather displayed no systematic preference for the informant’s argument skills.

## 9. Study 3: Collective decision-making

### 9.1. Introduction

The aim of Study 3 was to investigate preschooler's ability to collectively solve a judgment task. A tool important for decision-making is reasoning. In its narrow sense it was described as social skill enabling people to produce and evaluate arguments in order to reach joint decisions (Mercier & Sperber, 2011). Study 2 demonstrated preschooler's ability to evaluate argument quality in two contexts: firstly, in an advice-taking scenario in which children aggregated their own judgment with the judgment of the advisor. Results of Experiment 2 demonstrated that children accepted more advice when the advice was supported by an argument and moreover, effectively weighted identically phrased arguments as a function of the advisor's epistemic status. Secondly, in a selective trust scenario in which children did not have own access to information.

However, in social contexts advice-taking situations often do not end with an advisor giving a reason for own advice but they rather result in interpersonal dialogs in which arguments and counterarguments are being exchanged in a for-and-back fashion. Thus, a classical JAS paradigm in which judges make individual decisions by aggregating two judgments transforms into a joint decision based on a discussion about various solutions for a specific problem. In order to make an optimal collective decision partners need to exchange evidence for each others individual judgments and collectively weigh individual judgments for their reliability. It is an advanced metacognitive skill to provide a decision partner with an evaluation about the reliability of one's own judgment. In order to base a collective decision on evidence it is insufficient to only present the content of one's individual judgment but rather should give information on one's informational access and confidence in the own judgment. Moreover, effective reasoning requires partners to combine this information and produce arguments that reliably reflect on the quality of individual judgments. To date only a few studies have been investigating whether preschoolers use meta-argumentative strategies and meta-talk to collectively solve decision tasks. No study so far analysed the functionality of meta-talk as a parameter for decision quality.

In Study 3, therefore, dyads of preschoolers were presented with the visual perception task used in Study 1 and 2, however, this time in an interactive fashion. Dyads were asked to solve the problem collectively to reach a correct joint decision.

In particular, four main aspects of preschooler's collective decision-making were investigated:

- (1) The quality of collective decisions: Is the collective decision functional? Are 4- to 6-year-old children able to collectively decide for the correct judgment?
- (2) The decision style and quality of collective decisions: How do they do it? Does the quality of preschooler's joint decisions vary as a function of decision style?
- (3) The structure of dialogs: Do preschoolers produce meta-talk, including information on their individual perceptual access, the reliability of their individual judgments, and justifications for specific judgments?
- (4) The quality of collective decision as a function of dialog structure: Do dyads performing meta-talk solve the collective decision task more successfully? What content of meta-talk might result in more successful joint decision?

Moreover, following their collective decision in 50 % of all trials dyads were asked to give reasons for their final decision. Consequently, a sufficient amount of data was generated to investigate whether children's ability to correctly justify their final decision can predict the successfulness of collective decisions. The structure of requested reasons and the quality of collective decision as a function of structure of requested reasons were analyzed.

## **9.2. Method**

### **9.2.1. Participants**

Forty adults (age range: 18 - 41 years,  $M = 23.5$  years, 20 women) and forty-two 4- to 6-year-old children (age range: 48 - 71 months,  $M = 59.6$  months, 26 girls) participated in the study. Subjects were native German speakers with mixed socioeconomic backgrounds. Adults had answered our study announcement on public notice-boards. Children were recruited from a database of families who had previously given consent to their participation. One dyad lost interest during the test session and only three trials were included in the analysis.

### **9.2.2. Material**

The following material as in Study 1 was used: three-sided box with one clear and one blurred window, colored wooden sticks, and analogue scales (see Fig. 1, p. 38).

### **9.2.3. Design and procedure**

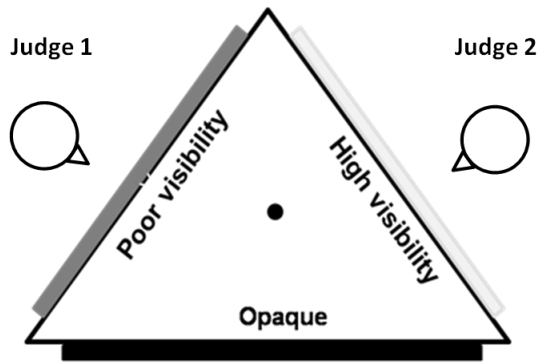
In this experiment participants were tested in dyads. For children, members of one dyad did not differ more than three months in age. Dyads were presented with a collective choice task:



they were asked to decide collectively for a final judgment based on two individual judgments. Individual judgments differed in quality and were made previously by the two judges, respectively. There was one condition (*high-poor*) with one participant making the individual judgment behind the clear window and the other judge making the individual judgment behind the blurred window. Age group (children and adults) was a between factor in the experimental design.

**9.2.3.1. Introduction phase.** Every session started with an introduction phase prior to the test phase. Participants were introduced individually, one after the other, to the set-up material and the visual perception task. After entering the test room participants were invited to explore the three-sided box. The two windows were indicated and their perceptual quality assessed. All children indicated correctly the quality of both windows. To familiarize participants with the stimuli they judged one stick without box. Participants' understanding of the scale as an identical image of the stimuli was ensured. Subsequently, participants judged one stimuli first at the blurred and second at the clear window. It was illustrated that the two judgments differ in quality.

**9.2.3.2. Test phase.** The two judges were each placed on a different side of the box (for illustration, see Fig. 21). One stick was put into the box (in such a way that subjects could not observe the placement), and both judges were asked to make their initial judgments. Then, judges were seated together at the opaque side of the box and both windows were occluded. Judges were instructed as follows: "Show each other your individual judgments, compare and discuss them to agree on a final collective judgment.". For children we added the phrasing "You are allowed to help each other.". Judges chose freely who marked the collective decision on the final scale. We conducted eight test trials. Every two trials participants switched their starting position (the clear or blurred window) where they made their individual judgment. After four trials both children answered again the manipulation check questions and indicated the window of high versus the one of low visual access. In the last four trials of each session we requested reasons from the judges for why they chose a specific individual judgment as final judgment. All test sessions were video recorded and lasted approximately 35 min.



**Figure 21.** Set-up, starting position of every test trial, when judges made individual judgments.

**9.2.3.3. Coding.** Firstly, all three judgments (two individual and one final judgment) were measured using a ruler on each scale and collected as judgment data. Based on this data we calculated the following dependent measures.

To access the quality of decisions it was coded:

- *Choice of judgment:* Represents the number of collective decisions for each of the three categories (high quality judgment, low quality judgment, and between). Final judgments within 2 cm around the high or low quality judgment were coded, respectively. Final judgments in the interval between both radii were coded as “between”. If individual judgments were less than 4 cm apart, the final judgment was coded as a function of closest distance to the high or the low quality judgment. Additionally, final judgments not located between the individual judgments but on one side of the high or low quality judgment were coded, respectively.
- *Deviation from the true value:* Represents the distance of the marks drawn by the judges from the true location of the mark on the stick (measured in mm).

Next, we coded children’s decision style using video recordings of all test trials and calculated the following dependent measures:

- *Form of decision:* Represents the number of collective decisions for each of the two categories: collective and individual. In collective decisions both children agreed verbally or gesturally (e.g., by nodding one’s head or pointing on the individual judgment they want to choose) on the final judgment. In individual decisions the elided decision partner showed verbally or gesturally signs of disagreement (e.g., pointing on the own individual judgment opposed to the judgment plotted as final judgment). In children a low number of trials were coded as undefined. In these trials

coders could not clearly identify these criteria (e.g., the head of the child was not visible or voices were not hearable on recordings).

- *Plotting of final judgment*: Represents the number of final marks that were plotted by each of the two judges (judge with high visual access and judge with poor visual access). In a low number of trials both judges plotted the final judgment, e.g., judge 1 draw the line on the third scale and judge 2 retraced the same line again. These trials were coded as undefined.

Finally, conversations between the children and requested reasons were transcribed verbatim. A full list of expressions coded for each category is provided in the appendix (dialogs: Tab. 1, p. 136, requested reasons: Tab. 1, p. 136). Based on these codings we calculated the following dependent measures:

- *Ratings*: Represents the number of dialogs including testimony about eyesight, state of knowledge, correctness of individual judgments, and level of difficulty.
- *Uncertainty*: Represents the number of dialogs including expressions of uncertainty about the correctness of individual judgments.
- *Reasons*: Represents the number of dialogs in which participants supported their individual judgments by reasons.
- *Requested reasons*: Represents the number of trials where participants gave reasons for each of the four categories: right, wrong, ambiguous, and irrelevant reasons.

### 9.3. Results

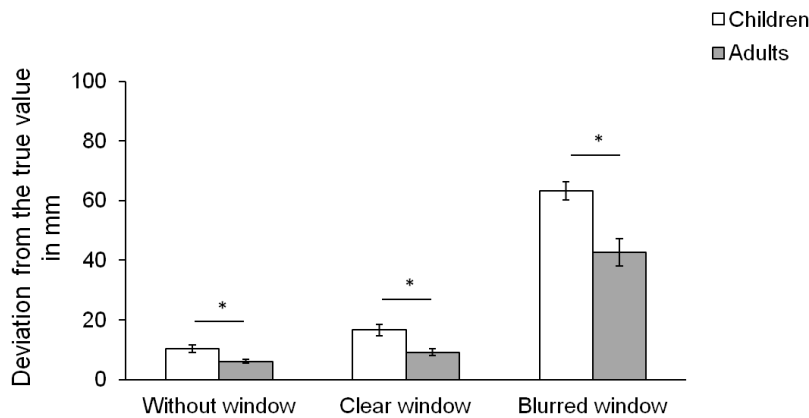
#### 9.3.1 Preliminary analyses.

The deviation from the true value in the three conditions of the familiarization phase (without the box, when looking through the clear window or when looking through the blurred window) are depicted in Fig. 22. The design of the set-up was valid. The visibility of participants was successfully manipulated by presenting them with a clear and a blurred window.

An ANOVA for repeated measures with form of visual access (without box, clear or blurred window) as within-subjects variable and age group (children or adults) as between-subjects factor was used to analyze whether there was a significant difference in the deviation from the true value between the two windows. The deviation from the true value without the box was included for reasons of completeness. An ANOVA with Greenhouse-Geisser correction revealed a significant main effect of visual access ( $F(1.3, 158) = 202.16, p$

< .001,  $\eta_p^2 = 0.72$ ), age group ( $F(1, 79) = 24.93$ ,  $p < .001$ ,  $\eta_p^2 = 0.24$ ) and a significant interaction of both factors ( $F(1.3, 158) = 6.95$ ,  $p < .005$ ,  $\eta_p^2 = 0.08$ ).

Post hoc paired sample  $t$ -tests revealed significant differences between all three forms of visual access in children (without box vs. clear window:  $t(40) = -2.72$ ,  $p = .010$ ,  $d_z = -0.43$ ; without box vs. blurred window:  $t(41) = 16.90$ ,  $p < .001$ ,  $d_z = 2.64$ ; clear vs. blurred window:  $t(40) = 12.90$ ,  $p < .001$ ,  $d_z = 2.04$ ) and adults (without box vs. clear window:  $t(39) = -2.63$ ,  $p = .012$ ,  $d_z = -0.42$ ; without box vs. blurred window:  $t(39) = -8.00$ ,  $p < .001$ ,  $d_z = -1.28$ ; clear vs. blurred window:  $t(39) = 6.94$ ,  $p < .001$ ,  $d_z = 1.11$ ). Furthermore, the values from the children's judgments deviated to a higher degree from the true value than the ones from adults. That was true for all three forms of visual access (without box:  $t(58.12) = 3.60$ ,  $p = .003$ ,  $d = 0.94$ , clear window:  $t(66.65) = 3.26$ ,  $p = .002$ ,  $d = 0.80$ , blurred window:  $t(71.24) = 3.64$ ,  $p = .001$ ,  $d = 0.86$ ).



**Figure 22.** Deviation from the true value as a function of form of visual access (without box, clear window and blurred window) in children and adults. Error bars indicate standard error of the means.

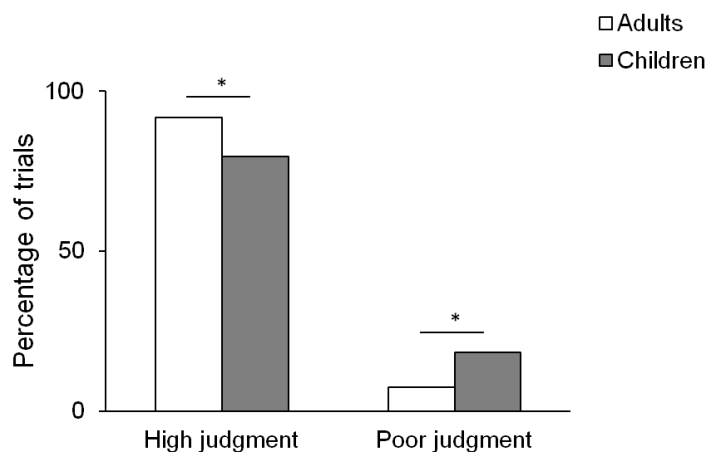
### 9.3.2. Main analyses.

**9.3.2.1. Number of collective decisions.** In children 132 decisions and in adults 121 decisions were included in the analysis. These decision trials fulfilled the following requirements: Firstly, the collective decision process was completed, including all three judgments (two individual judgments, one from the clear and one from the blurred window, and one collective judgment). Secondly, individual judgments were at least 2 cm apart, to ensure that they were perceived as differing and thus an agreement had to be formed. Furthermore, trials with a smaller distance of individual judgments were included when the individual judgments were at least 1 cm apart and additionally, the final judgment was

located not between the individual judgments but outside of this range and could thus be clearly assigned to the high or low quality individual judgment.

**9.3.2.2. Quality of collective decisions.** Two measures were used to investigate the quality of collective decisions: choice of judgment (whether the judges collectively choose the individual judgment of high or low quality) and the deviation from the true value in mm of the collective judgment.

*9.3.2.2.1. Choice of judgment.* The choice of final judgments is depicted in Fig. 23. Only 0.03 % of trials fell in the “between” category and were not included in further analyses. The hypothesis was that adults will choose the high quality judgment as final judgment more often than children. To examine this prediction, collective choices of the high quality judgment were coded as 1 and collective choices of the low quality judgment were coded as 0. Decision data were analyzed using a Generalized Linear Mixed Model (GLMM) with a binary response term (high quality judgment = 1, low quality judgment = 0) (Bolker et al., 2009). Age group was entered as fixed effect and dyad identification number was entered as random effect given the repeated measures design of the experiment. The proportion of high quality judgments as collective choices increased with age (*estimate* ± *SE*:  $1.01 \pm 0.47$ ,  $t_{246} = 2.15$ ,  $p = .032$ ). Consistent with the hypothesis, adults were more likely to collectively decide for the high quality judgment than children.

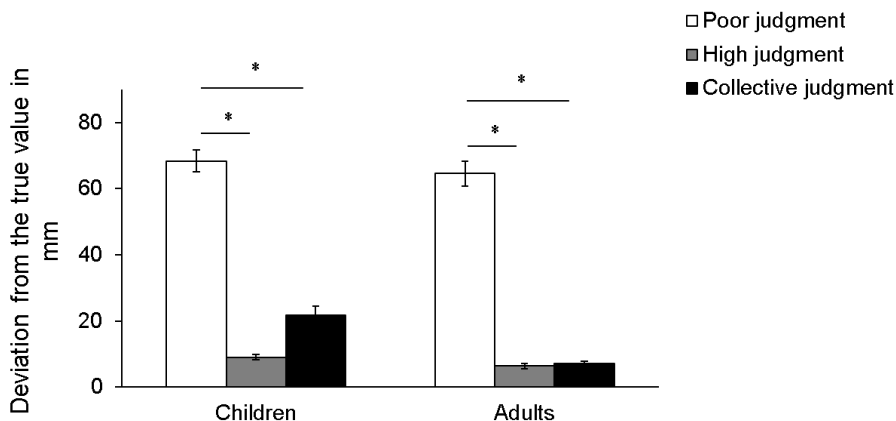


**Figure 23.** Choice of the collective judgment as proportions of high and low quality judgments in children and adults.

*9.3.2.2.2. Deviation from the true value.* The deviation from the true value as a function of judgment and age group is depicted in Fig. 24. A 3 (poor quality judgment, high quality judgment and collective judgment) × 2 (children and adults) ANOVA for repeated measures with Greenhouse-Geisser-correction showed a significant main effect of judgment

( $F(1.47, 366.85) = 374.59, p < .001, \eta_p^2 = 0.60$ ), age group ( $F(1, 250) = 11.56, p = .001, \eta_p^2 = 0.04$ ) and a significant interaction of both factors ( $F(1.47, 366.85) = 3.94, p = .032, \eta_p^2 = 0.02$ ).

Post hoc tests revealed a significant higher deviation from the true value in the case of the poor quality judgment compared to the high quality judgment (children:  $t(131) = 16.53, p < .001, d_z = 1.44$ ; adults:  $t(119) = 14.49, p < .001, d_z = 1.33$ ). However, deviations of the children did not differ from deviation of the adults in case of the poor quality judgements ( $t(250) = 0.74, p = .457, d = 0.09$ ) but adults plotted high quality judgments more accurately ( $t(249.89) = 2.54, p = .012, d = 0.32$ ). Children's collective judgments deviated significantly more from the true value than their individual high quality judgments ( $t(131) = -4.74, p < .001, d_z = -0.41$ ) but significantly less than their individual low quality judgments ( $t(131) = 12.45, p < .001, d_z = 1.09$ ). Children's collective judgments deviated significantly more from the true value than adult's collective judgments ( $t(155.72) = 5.91, p < .001, d = 0.83$ ), adult's collective judgments were equivalent to their individual high quality judgments ( $t(119) = -0.73, p = .467, d_z = -0.07$ ).

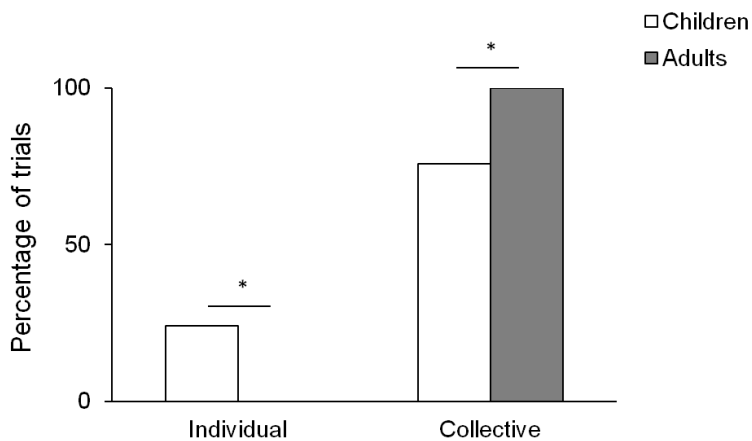


**Figure 24.** Deviation from the true value in mm for the two individual judgments (plotted at the clear or blurred window) and the collective judgments in children and adults. Error bars indicate standard error of the means.

### 9.3.2.3. Decision style and quality of decision.

9.3.2.3.1. *Form of decision.* The form of decision in children and adults is depicted in Fig. 25. Only 0.02 % of all trials were coded as undefined in terms of whether the decision was collective or individual, these trials were not included in further analyses. It was hypothesized that children, in contrast to adults, make a certain number of decisions individually instead of collectively. Decisions made collectively were coded as 1 and decisions made individually were coded as 0. The proportion of each form of decision across

trials were calculated as dependent variable. Decision data were analyzed using a Generalized Linear Mixed Model (GLMM) with a binary response term (collective decisions = 1, individual decisions = 0) (Bolker et al., 2009). Age group was entered as fixed effect and dyad identification number was entered as random effect given the repeated measures design of the experiment. The proportion of individual decision decreased with age (*estimate*  $\pm$  *SE*:  $3.56 \pm 1.13$ ,  $t_{246} = 3.22$ ,  $p = .001$ ). Consistent with the hypothesis, adults were more likely to collectively decide for the high quality judgment than children.

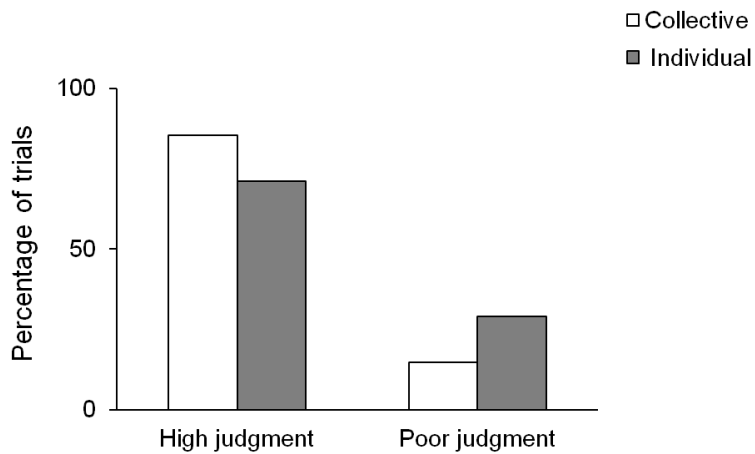


**Figure 25.** Form of decision. Proportion of collective and individual decisions in children and adults.

9.3.2.3.2. *Quality of decision as a function of form of decision.* Two measures were used to investigate the quality of decisions as a function of whether dyads decided collectively or whether one judge decided individually. Firstly, the proportion of successful decisions (high quality judgment was chosen as final judgment) and secondly, deviation from the true value of final judgments as a function of form of decision.

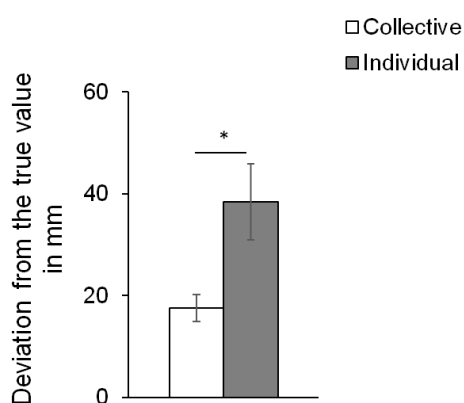
*Proportion of successful decisions.* The proportion of successful trials as a function of form of decision is depicted in Fig. 26. The following analyses are for children only, since adults decided collectively in all trials. The hypothesis was that in children collective and individual decisions would differ qualitatively. Decision data were analyzed using a Generalized Linear Mixed Model (GLMM) with a binary response term (high quality judgment = 1, low quality judgment = 0) (Bolker et al., 2009). Form of decision was entered as fixed effect and dyad identification number was entered as random effect given the repeated measures design of the experiment. The proportion of high quality judgments as collective choices was no different in collective compared to individual decisions (*estimate*  $\pm$  *SE*:  $-0.64 \pm 0.52$ ,  $t_{126} = -1.21$ ,  $p = .227$ ). The hypothesis need to be rejected since data indicate that

children are not more or less successful as a function of whether they decided collectively or individually.



**Figure 26.** Proportion of high and low quality judgments as a function of form of decision (collective versus individual) in children.

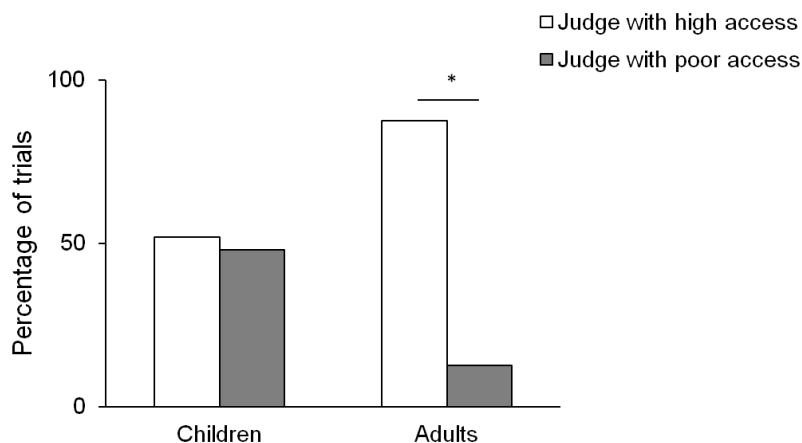
Deviation from the true value. The deviation from the true value of the final judgment as a function of form of decision (collective versus individual) is depicted in Fig. 27. The hypothesis was that in children collective and individual decisions would differ in deviation from the true value. Data were analysed using a Linear Mixed Model (LMM) given that the response variable was a continuous measure and the repeated measures design of the experiment. Form of decision was entered as fixed effect and dyad identification number was entered as random effect. Form of decision was a binary predictor term (collective = 1, individual = 0). Individual decisions deviated significantly more from the true value than collective decisions ( $estimate \pm SE: -13.77 \pm 0.57, t_{15,66} = -23,98 p < .001$ ).



**Figure 27.** Deviation from the true value of the final judgment as a function of form of decision (collective versus individual) in children. Error bars indicate standard error of the means.



**9.3.2.3.3. Plotting of final judgments.** The plotting of the final judgments (who of the two judges plotted the collective decision) is depicted in Fig. 28. In 0.02 % of all trials both judges plotted the same final judgment (e.g. the second child retraced the plotting of the first child), these plottings were coded as undefined in terms of who plotted the collective judgment and did not include them in our further analyses. Plottings of the judge with high visual access were coded as 1 and plottings of the judge with poor visual access were coded as 0. Decision data were analyzed using a Generalized Linear Mixed Model (GLMM) with a binary response term (judge with high visual access = 1, judge with poor visual access = 0) (Bolker et al., 2009). To analyze whether one of the judges (judge with high or poor visual access) dominated in marking the final judgment, we calculated a basic GLMM without predictor for children and adults and dyad identification number as random effect given the repeated measures design of the experiment. Children did not differ in who of the two judges marked the final decision (*estimate*  $\pm$  *SE*:  $-0.08 \pm 0.24$ ,  $t_{127} = -0.35$ ,  $p < .726$ ). In contrast, in adults the final decision was more often marked by the judge with high visual access compared to the judge with low visual access (*estimate*  $\pm$  *SE*:  $-2.15 \pm 0.43$ ,  $t_{118} = -5.05$ ,  $p < .001$ ). To analyze whether the final decision was plotted more often by one of the judges in children and adults a GLMM with age group entered as fixed effect was calculated. The proportion of final judgments plotted by the judge with high visual access increased with age (*estimate*  $\pm$  *SE*:  $-1.91 \pm 0.45$ ,  $t_{250} = -4.25$ ,  $p < .001$ ).



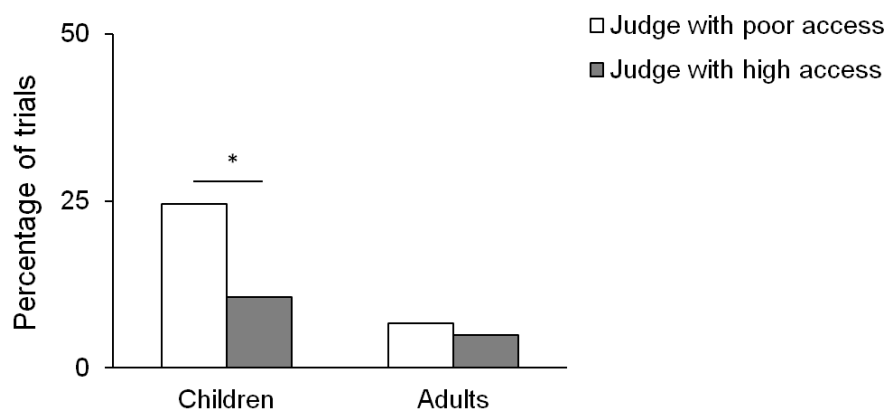
**Figure 28.** Percentage of trials in which the judge with high compared to the judge with poor visual access plotted the final judgment.

*Quality of decision as a function of plotting.* Two measures were used to investigate the quality of judgments plotted by the judge with high or low informational access: firstly, the plotting of each judge (whether a judge plotted the high or low quality judgment, meaning the

own judgment or the judgment of the other judge) and secondly, the deviation from the true value of the collective judgment as a function of who plotted the collective judgment.

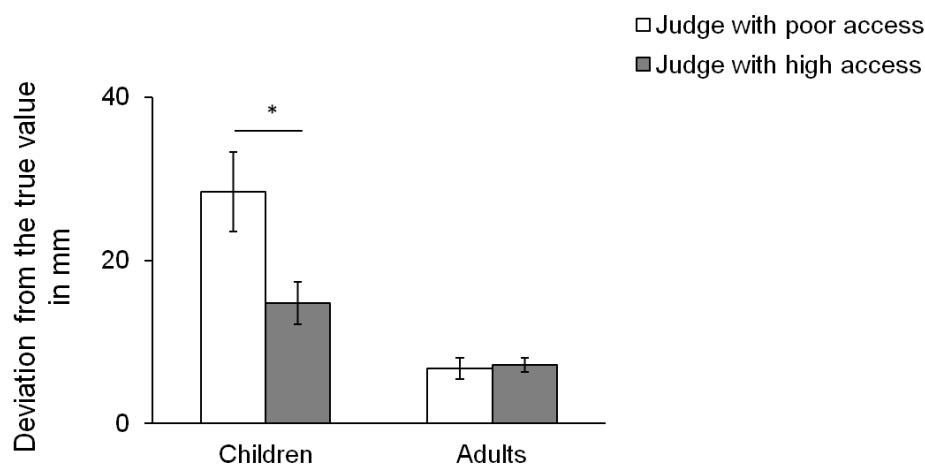
#### Plottings of judges with high or low informational access.

The number of times judges plotted their own individual judgement as final decision is depicted in Fig. 29. The hypotheses were that in children a poor judge, when plotting the collective decision, is more often plotting the own individual poor judgment than in adults and that plottings of high judges would not differ between children and adults. To examine this predictions, plottings of the low quality judgment as final judgment were coded as 1 and plottings of the low quality judgment as final judgment were coded as 0. Decision data were analyzed using a Generalized Linear Mixed Model (GLMM) with a binary response term (high quality judgment = 1, low quality judgment = 0) (Bolker et al., 2009). Visual access of the judge who plotted the final judgment was entered as fixed effect and dyad identification number was entered as random effect given the repeated measures design of the experiment. In children the proportion of poor judgments as final judgement increased for final decisions plotted by the judge with poor visual access compared to final decisions plotted by the judge with high visual access (*estimate*  $\pm$  *SE*:  $1.12 \pm 0.50$ ,  $t_{125} = 2.25$ ,  $p = .026$ ). In adults the proportion of poor judgments as final judgement did not increase for final decisions plotted by the judge with poor visual access compared to final decisions plotted by the judge with high visual access (*estimate*  $\pm$  *SE*:  $0.65 \pm 0.85$ ,  $t_{116} = 0.77$ ,  $p = .441$ ). Consistent with the hypotheses, in children, if the judge who had poor visual access plotted the final judgment the own judgment was plotted more often (the judgment of low visual access) than when the judge with high visual access plotted the final judgment. This was not true for adults.



**Figure 29.** Proportion of low quality judgments as final judgments as a function of who plotted the final judgment (the judge with high or poor visual access).

Deviation from the true value. The deviation from the true value of the final judgment as a function of who plotted the judgment is depicted in Fig. 30. The hypothesis was that in children final judgments plotted by the judge with poor visual access would deviate more from the true value than final judgments plotted by the judge with high visual access. Data were analysed using Linear Mixed Models (LMMs) given that the response variable was a continuous measure and the repeated measures design of the experiment. The judge who plotted the final judgment was entered as fixed effect and dyad identification number was entered as random effect. The judge who plotted the final judgment was a binary predictor term (judge with high visual access = 1, judge with poor visual access = 0). In children final judgments plotted by the judge with poor visual access deviated significantly more from the true value than final judgments plotted by the judge with high visual access (*estimate*  $\pm$  *SE*:  $16.18 \pm 1.42$ ,  $t_{132} = 11.37$ ,  $p < .001$ ). In adults the deviation from the true value of the final judgment did not differ as a function of who plotted the final judgment (*estimate*  $\pm$  *SE*:  $-1.85 \pm 0.02$ ,  $t_{120} = -79.67$ ,  $p = 1.000$ ).

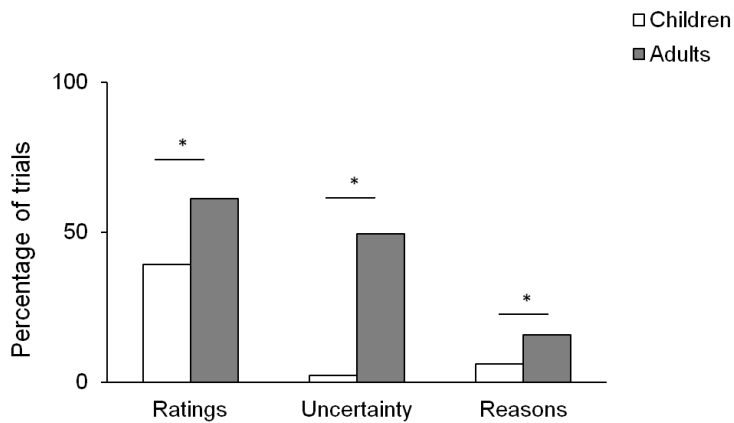


**Figure 30.** Deviation from the true value of the final judgment as a function of who plotted the final judgment (the judge with high or poor visual access). Error bars indicate standard error of the means.

#### 9.3.2.4. Structure of dialogs and requested reasons

9.3.2.4.1. *Dialogs.* The structure of dialogs is depicted in Fig. 31. It was hypothesized that adults incorporate ratings, uncertainty, and reasons more often in their dialogs and coded occurrence of a measure as 1 and absence of the measure as 0. Generalized Linear Mixed Models (GLMMs) with a binary response term (occurrence of measure = 1, absence of measure = 0) (Bolker et al., 2009) were calculated for all three measures. Age group was entered as fixed effect and dyad identification number was entered as random effect given the repeated measures design of the experiment. All three measures increased with age

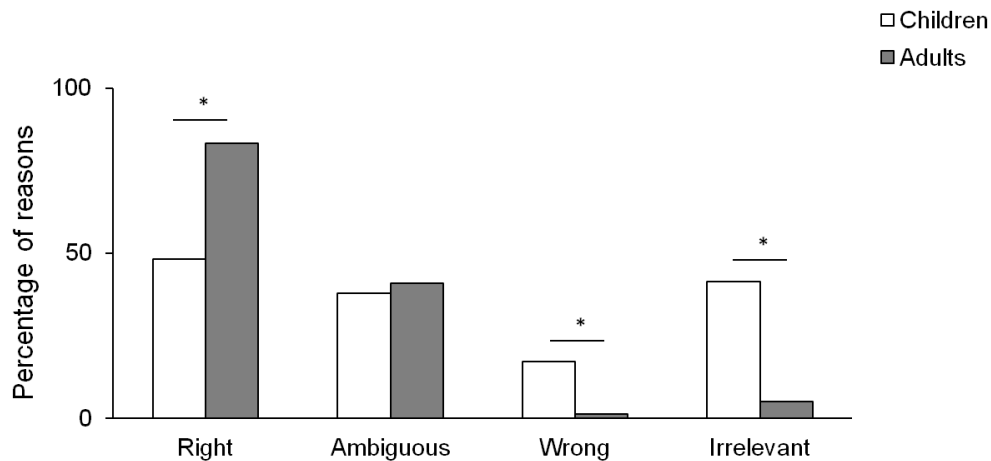
(ratings:  $estimate \pm SE: 0.67 \pm 0.31$ ,  $t_{251} = 2.16$ ,  $p = .032$ ; uncertainty:  $estimate \pm SE: -3.77 \pm 0.66$ ,  $t_{251} = -5.70$ ,  $p < .001$ ; reasons:  $estimate \pm SE: -1.10 \pm 0.54$ ,  $t_{251} = -2.04$ ,  $p = .042$ ). Consistent with the hypothesis, dyads of adults discussing their collective decision about the final judgment rated the quality of their individual judgments, announced concerns about their certainty and gave reasons for choosing a certain individual judgment more often than children did.



**Figure 31.** Dialog structure, including the three measures ratings, uncertainty, and reasons in children and adults.

**9.3.2.4.2. Requested reasons.** The structure of requested reasons is depicted in Fig. 32. It was hypothesized that adults justify their individual and collective judgments more often with right and ambiguous reasons while children give more wrong and irrelevant reasons. The occurrence of each type of reason was coded as 1 and absence of each type of reason as 0. Generalized Linear Mixed Models (GLMMs) with a binary response term (occurrence of measure = 1, absence of measure = 0) (Bolker et al., 2009) were calculated for all four measures. Age group was entered as fixed effect and dyad identification number was entered as random effect given the repeated measures design of the experiment. Analysis revealed that right and ambiguous reasons increased with age (right reasons:  $estimate \pm SE: -1.95 \pm 0.59$ ,  $t_{134} = -3.33$ ,  $p = .001$ ) while wrong and irrelevant reasons decreased with age (wrong reasons:  $estimate \pm SE: -2.68 \pm 1.25$ ,  $t_{134} = -2.14$ ,  $p = .034$ ; irrelevant reasons:  $estimate \pm SE: -2.60 \pm 0.69$ ,  $t_{134} = -3.77$ ,  $p < .001$ ). In contrast, ambiguous reasons did not increase with age ( $estimate \pm SE: 0.16 \pm 0.46$ ,  $t_{134} = 0.34$ ,  $p = .735$ ).

Consistent with the hypothesis, adults justified their decisions using more right reasons while children used more wrong and irrelevant reasons. In contrast with the hypothesis, children and adults did not differ in their usage of ambiguous reasons.



**Figure 32.** Proportions of the four types of requested reasons (right, ambiguous, wrong, and irrelevant) in children and adults.

### 9.3.2.5. Quality of collective decisions as function of dialog structure and requested reasons

#### 9.3.2.5.1. Dialogs

Quality of decision as a function of testimony. Two measures were used to investigate the quality of collective decisions as a function of testimony. Firstly, the proportion of successful decisions (decisions favoring high quality judgments) as outcome of dialogs that comprised each dialog parameter or did not comprise the parameter, respectively. As a second measure the deviation from the true value of final judgments following dialogs with or without dialog parameter was compared.

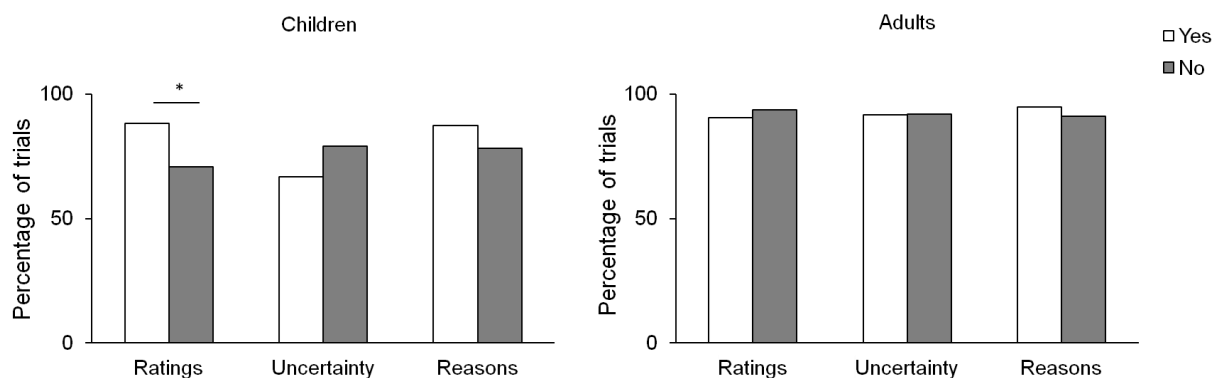
Proportions of successful decisions. Proportions of successful decisions are depicted in Fig. 33. The hypothesis was that in children, opposed to adults, the usage of certain dialog parameters can predict the successfulness of collective decisions. In adults we did not have such a prediction since their was no difference in successfulness between trial (adults were successful in nearly all trials). Analysis on adult date are presented for the matter of completeness. To examine this prediction, collective decisions favoring high quality judgment were coded as 1 and collective decisions favoring low quality judgment were coded as 0. Generalized Linear Mixed Models (GLMMs) with a binary response term (high quality judgment = 1, low quality judgment = 0) (Bolker et al., 2009) were calculated for each dialog parameter (ratings, uncertainty, and reasons) in children and adults. Occurrence of dialog parameters was entered as fixed effect and dyad identification number was entered as random effect given the repeated measures design of the experiment.

Ratings. In children the success of the final decision increased as a function of whether judges used ratings of their individual judgment throughout the dialog or not (*estimate*  $\pm$  *SE*:  $1.10 \pm 0.49$ ,  $t_{130} = 2.25$ ,  $p = .026$ ). In adults, in contrast, the success of the final decision did not increase as a function of whether judges used ratings of their individual judgment throughout the dialog or not (*estimate*  $\pm$  *SE*:  $0.34 \pm 0.69$ ,  $t_{119} = 0.50$ ,  $p = .619$ ).

Uncertainty. In children and adults the success of the final decision did not increase as a function of whether the judges used expressions of uncertainty about their individual judgment throughout the dialog or not (children: *estimate*  $\pm$  *SE*:  $0.75 \pm 1.30$ ,  $t_{130} = 0.57$ ,  $p = .568$ ; adults: *estimate*  $\pm$  *SE*:  $-0.01 \pm 0.65$ ,  $t_{119} = -0.02$ ,  $p = .982$ ).

Reasons. In children and adults the success of the final decision did not increase as a function of whether judges gave reasons why their individual judgment should be chosen throughout the dialog or not (children: *estimate*  $\pm$  *SE*:  $-0.58 \pm 1.13$ ,  $t_{130} = -0.52$ ,  $p = .61$ ; adults: *estimate*  $\pm$  *SE*:  $-0.42 \pm 1.01$ ,  $t_{119} = -0.41$ ,  $p = .679$ ).

In summary, consistent with the hypothesis, successfulness of their collective decision differed as a function of dialog structure. Dialogs that included ratings resulted in a higher number of successful trials than dialogs without ratings. Ratings of uncertainty and given reasons did not correlate with successfulness.



**Figure 33.** Proportion of successful trials (high quality judgment as outcome of collective decision) with dialogs comprising each dialog parameter (ratings, uncertainty, and reasons) or not.

Deviation from the true value as a function of testimony. The deviation from the true value of the final judgment as a function of whether each dialog parameter occurred in the dialog or not is depicted in Fig. 34. The hypothesis was that in children, opposed to adults, the usage of dialog parameters can predict the accuracy of the final judgment. Linear Mixed

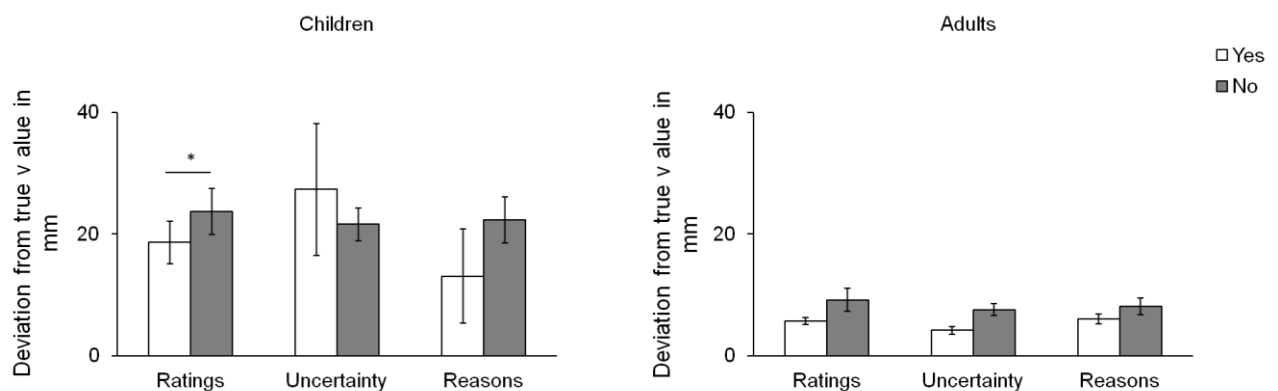
Models (LMMs) for each dialog parameter (ratings, uncertainty, and reasons) were calculated.

**Ratings.** In children collective judgments deviated significantly less from the true value when dyads included ratings of their individual judgments in their collective decision processes ( $estimate \pm SE: 9.21 \pm 1.68, t_{132} = 5.48, p = .016$ ). In adults the deviation from the value of the collective judgment did not differ as function of ratings ( $estimate \pm SE: 1.23 \pm 0.01, t_{121} = 260.90, p = 1.000$ ).

**Uncertainty.** In children and adults the deviation from the value of the collective judgment did not differ as function of uncertainty (children:  $estimate \pm SE: -3.00 \pm 0.03, t_{132} = -113.63, p = .131$ ; adults:  $estimate \pm SE: 0.88 \pm 0.09, t_{121} = 10.12, p = .220$ ).

**Reasons.** In children and adults the deviation from the value of the collective judgment did not differ as function of uncertainty (children:  $estimate \pm SE: -20.31 \pm 0.17, t_{132} = -118.22, p = .440$ ; adults:  $estimate \pm SE: 1.39 \pm 0.01, t_{121} = 141.32, p = .985$ ).

In summary, children dyads increased their accuracy of their final judgment when discussing the quality of individual judgments. Rating uncertainty or giving reasons did not influence the accuracy of collective decisions in children. But it is notable that only in a few trials children gave reasons spontaneously.



**Figure 34.** Deviation from the true value of the collective judgment as a function of whether the dialog parameter (ratings, uncertainty, or reasons) occurred in the collective decision making dialog. Error bars indicate standard error of the means.

#### 9.3.2.5.2. Requested reasons.

Quality of decision as a function of requested reasons. Two measures were used to investigate the quality of collective decisions as a function of requested reasons. Firstly, the proportion of successful decisions (decisions favoring high quality judgments) as a function

of type of reasons dyads gave to justify their collective judgments was used. As a second measure the deviations from the true value for collective decisions as a function of whether a certain type of reason was given or not was compared.

Proportions of successful decisions. Proportions of successful decisions as function of type of requested reason are depicted in Fig. 35. The hypothesis was, firstly, that collective decisions justified by right and ambiguous reasons are more successful than collective decisions not justified by right and ambiguous reasons, and secondly, that collective decisions justified by wrong and irrelevant reasons are less successful than collective decisions not justified by wrong and irrelevant reasons. To examine these predictions, collective decisions of the high quality judgment were coded as 1 and collective decisions of the low quality judgment were coded as 0. We calculated Generalized Linear Mixed Models (GLMMs) with a binary response term (high quality judgment = 1, low quality judgment = 0) (Bolker et al., 2009) for each type of reason (right, ambiguous, wrong, and irrelevant) in children and adults. Occurrence of each type of reason was entered as fixed effect and dyad identification number was entered as random effect given the repeated measures design of the experiment.

**Right reasons.** In children and adults the success of the final decision did not increase as a function of whether dyads justified their final decisions giving right reasons (children:  $estimate \pm SE: 0.89 \pm 0.59, t_{56} = 1.51, p = .138$ ; adults:  $estimate \pm SE: -0.55 \pm 0.88, t_{76} = -0.62, p = .534$ ).

**Ambiguous reasons.** In children the success of the final decision did increase as a function of whether dyads justified their final decisions giving ambiguous reasons ( $estimate \pm SE: 2.04 \pm 0.80, t_{56} = 2.54, p = .014$ ). In adults the success of the final decision did not increase as a function of whether dyads justified their final decisions giving ambiguous reasons ( $estimate \pm SE: -0.15 \pm 0.76, t_{76} = -0.19, p = .846$ ).

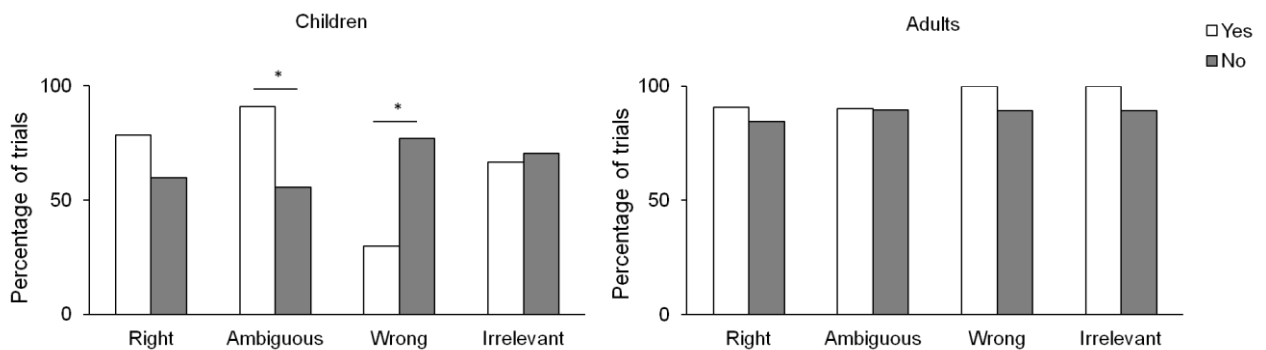
**Wrong reasons.** In children the success of the final decision did decrease as a function of whether dyads justified their final decisions giving wrong reasons ( $estimate \pm SE: -2.29 \pm 0.83, t_{56} = -2.76, p = .008$ ). In adults the success of the final decision did not increase as a function of whether dyads justified their final decisions giving wrong reasons ( $estimate \pm SE: -1.38 \pm 6.00, t_{76} = -0.23, p = .818$ ).

**Irrelevant reasons.** In children and adults the success of the final decision did not increase as a function of whether dyads justified their final decisions giving irrelevant



requested reasons (children:  $estimate \pm SE: 0.18 \pm 0.57, t_{56} = 0.32, p = .753$ ; adults:  $estimate \pm SE: -1.42 \pm 3.01, t_{76} = -0.47, p = .638$ ).

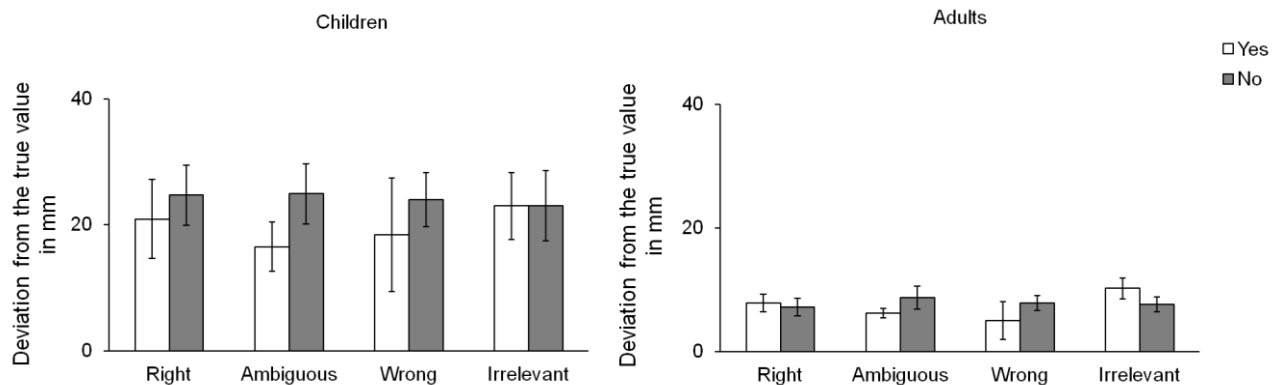
In summary, consistent with the hypothesis children's collective decisions were more successful if justified with ambiguous reasons and less successful when justified with wrong reasons. In contrast, giving right or irrelevant reasons did not interplay with the successfulness of joint decisions.



**Figure 35.** Proportion of success for each type of requested reason (right, ambiguous, wrong, irrelevant) as a function of whether each type of reason was given to justify the collective judgment or not.

Deviation from the true value as a function of testimony. The deviation from the true value of the final judgment as a function of type of requested reason is depicted in Fig. 36. The hypothesis was that in children the accuracy of the collective judgment would differ as a function of reasons dyads give if asked to justify their decisions. It was expected that the final judgment justified by right and ambiguous reasons deviates less from the true value than the final judgment justified by wrong and irrelevant reasons. Linear Mixed Models (LMMs) for each type of requested reason (right, ambiguous, wrong, and irrelevant) in children and adults was calculated. Analysis revealed no significant difference in the deviation from the true value as a function of occurrence of any of the four types of requested reasons in children or adults: right requested reasons (children:  $estimate \pm SE: 5.49 \pm 3.47, t_{58} = 1.58, p = .134$ ; adults:  $estimate \pm SE: -0.60 \pm 2.01, t_{78} = -0.30, p = .767$ ), ambiguous requested reasons (children:  $estimate \pm SE: -1.13 \pm 4.45, t_{58} = -0.25, p = .802$ ; adults:  $estimate \pm SE: 1.56 \pm 1.23, t_{78} = 1.24, p = .220$ ), wrong requested reasons (children:  $estimate \pm SE: 3.87 \pm 5.13, t_{58} = 0.75, p = .461$ ; adults:  $estimate \pm SE: 7.00 \pm 4.05, t_{78} = 1.73, p = .101$ ), irrelevant

requested reasons (children:  $estimate \pm SE: 0.53 \pm 3.86, t_{58} = 0.14, p = .892$ ; adults:  $estimate \pm SE: -4.39 \pm 3.05, t_{78} = -1.44, p = .160$ ).



**Figure 36.** Deviation from the true value as a function of type of requested reason. Error bars indicate standard error of the means.

#### 9.4. Discussion

Four aspects of preschooler's collective decisions were investigated: the quality, the decision style and quality, the dialog structure, and the quality as a function of dialog structure. The main findings of Study 3 included the following.

Firstly, dyads of children, similar to adults, preferred the individual judgment made behind the clear window as final judgment. The deviations from the true value for the three judgments mirrored this preference. In children and adults the collective judgment deviated significantly less from the true value than the judgment made behind the blurred window. In contrast, the collective judgment did not deviate more from the true value than the judgment made behind the clear window. Thus, in most of their joint decisions children decided for the high quality judgment.

Secondly, children, unlike adults, decided individually for a final judgment in a certain number of trials. In individual decisions children drew the final judgment without approval of their decision partners. In these trials the elided decision partner showed verbally or gesturally signs of disagreement (e.g. pointing on the own individual judgment opposed to the judgment determined as final judgment). Individual decisions did not differ in their proportion of successful trials, however, final judgments based on individual decisions deviated significantly more from the true value than in joint decisions. This indicated the tendency of judges initially sitting behind the blurred window ("poor judge") to more often overrule their decision partners compared to judges initially sitting behind the clear window

("high judge"). Furthermore, analysis of who determined the final judgments showed that children took turns when drawing the final mark independently of the judge's starting position. In contrast, adult judges initially sitting behind the clear window drew the final judgment in around 90 % of all trials. More detailed analysis showed that in children versus adult trials, "poor judges" plotted poor judgment more often as final judgment than "high judges". Similarly, in children trials only final judgments plotted by the "poor judge" deviated more from the true value than final judgments plotted by the "high judge".

Thirdly, dialogs of children and adult dyads included ratings about the visibility of the stimulus, the correctness of individual judgments, and the difficulty of the perception task. Furthermore, children and adults raised concerns about the reliability of individual judgment and gave reasons to justify their proposals of specific judgments. The proportion of trials including these dialog parameters was significantly lower with children than with adults. In particular, children reflected on their uncertainty and gave reasons for their proposal only in a few trials, representing 13 % and 6 %, respectively. When being asked by the experimenter to justify their collective decisions children dyads gave more often wrong and irrelevant reasons than adults. In contrast, adults gave more often right reasons than children.

Fourthly, children's dialogs that included ratings resulted in a higher number of successful trials than dialogs without ratings. This is consistent with the hypothesis that successfulness of children's collective decisions differs as a function of dialog structure. Furthermore, when being asked for reasons children more often justified successful decisions with ambiguous reasons and unsuccessful decisions with wrong reasons.

In summary, results of Study 3 showed that children and their partners who accessed information from sources of differing quality trusted the information supported by the strong evidence rather than by the weak evidence in their joint decisions. Thus, 4- to 6-year-old children were able to reach the "correct" joint decisions and endorse high quality information contradicting their initial judgment when their initial judgment resulted from a less reliable perceptual access.

An interesting question arose as to how children managed to decide successfully? Children seemed to be able to intuitively reflect on the quality of, at least, their own informational access. In around 50 % of the trials they explicitly informed their decision partner about this informational status by rating their visibility or the difficulty of the task. However, children's understanding of the causal relationship between the quality of the visual access and the subsequent quality of individual judgments showed to be still fragile in 4- to

6-year-old children. This finding is consistent with previous research suggesting limits in preschooler's ability to access their own ignorance (Kuhn, 2001). They seemed to subjectively turn a feeling of knowing into knowledge limiting their ability to grade certainty effectively (Kloo & Rhower, 2012).

In the current experiment correct joint decision may have resulted from two scenarios: firstly, both decision partners held accurate evaluation of their own knowledge (the "poor judge" was aware of the low quality of their own individual judgment and the "high judge" was aware of the high quality of their own judgment) or secondly the "high judge" was more dominant within the decision process and did more of the interactive work than the "poor judge". However, in trials where the "poor judge" was confident in own knowledge and not willing to revise individual poor judgment, "high judges" were not able to use meta-talk and argumentative strategies. Thus, they failed to make the evidence supporting their high quality judgments and the rationality of choosing this judgment transparent to their decision partners. Therefore, they were not successful in convincing them to collectively decide for the better judgment. The solely rating of visibility and task demands did not proof strongly enough as argument to convince a confident "poor judge". More promising seemed the use of meta-talk to illustrate a causal connection between the degree of visibility and the subsequent quality of the proposed judgment. Based on this understanding argumentative strategies can be used to justify a specific choice of judgment and cause agreement about a correct joint decision. This in return can then be based on evidence as a means to optimize decision outcomes.

In conversations with peers children spontaneously gave reasons to justify their proposals only in a few trials. However, children's requested reasons (when asked by the experimenter) mirrored limits in their ability to reflect on the quality of individual judgments. Decisions justified by ambiguous reasons were more often decisions that resulted in the better judgment than decisions not justified by this type of reasons. Reasons were coded as ambiguous when they transferred the correct message about why the high quality judgment was chosen, for example, „because this judgment is the right one“ or „because I know it better“. However, ambiguous reasons (opposed to right reasons) did not clarify the causal connection between the visual access and the quality of judgments, for example, „mine is better, because I saw better“ or „mine is worse, because I was sitting behind the blurred window“. Decisions justified by wrong reasons were more often unsuccessful than decisions not justified by wrong reasons. Reasons were coded as „wrong“ when children reflected incorrectly on their visual access, for example, when a „poor judge“ claimed to have sufficient

visual access. It seemed that only the wish to know the answer or the limited ability to identify their poor judgment as relevant guess (since they had visual access, but a poor one) turned subjectively into perfect knowledge. However, this claim could not be justified by arguments based on logic or evidence which might have led to the apparently “second best approach”, namely to claim to have the better evidence. These findings are consistent with research showing children’s understanding of their own epistemic states to be fully developed late in their childhood, and in a reflective/explicit form even delayed until school age (Kloo & Rhower, 2012).

## 10. General discussion

The aim of this dissertation was to investigate the development of socially induced individual and collective belief revision in young children. In order to do so an interdisciplinary approach was used and four separate lines of research were combined: studies on advice-taking in social psychology, on the ontogeny of selective social learning (selective trust), on argumentation, and on collective judgment formation. In particular, in a first step a method classically used in social psychology research was adapted and a judge-advisor-system suitable for 4- to 6-year-old children designed to subsequently investigate: (1) preschoolers' advice-taking as a function of their own and the advisor's state of knowledge, (2) preschoolers' ability to evaluate arguments and their subsequent advice utilization as a function of argument quality, and (3) preschoolers' collective decision-making as a function of dialog structure and argumentative strategies.

In the first study participants solved a perceptual judgment task, received advice, and subsequently made final decisions. The informational access (perceptual quality) of participants and advisor was experimentally manipulated. Adults revised their judgments systematically as a function of both their own and the advisors' informational access while children based their adjustments only on their own informational access. Two follow-up experiments suggested, however, that this pattern of results in children reflected performance rather than competence limitations: In suitably modified tasks, children did consider both their own information and that of the advisor in their selective social belief revision. These results will be discussed with focus on limiting performance factors and developmental trajectories.

Study 2 included three experiments. In Experiment 1 and 2 the same visual perception task as in Study 1 was used. However, advisors supported their judgment by giving a reason for why they recommend the final judgment to be consistent with the advice. Children were sensitive to arguments and took more advice when accompanied by arguments. However, they varied in their degree of advice-taking as a function of argument quality only if cues of argument quality were unambiguous. Experiment 3 went one step further and investigated whether preschoolers infer a speaker's future credibility based on the argumentative skills and whether information by poor argumentators would even be rejected. For this purpose a newly designed decision task was developed and presented in two different formats of a selective trust paradigm: a two-informant and a single-informant version. Children weighed new information as a function of argument quality and endorsed identically phrased arguments as a function of context. These findings supported children's

ability to judge arguments based on the epistemic relationship between information and argument. However, children did not actively reject information from speakers previously providing weak arguments. Study 2 will be discussed in terms of factors that might have increased task demands for children and the use of arguments to determine informant credibility.

Finally, Study 3 presented dyads of children with a visual perception task. Children started solving the task individually and subsequently compared these individual judgments finally to be integrated into a collective judgment. Importantly, judge's informational access was manipulated resulting in individual judgments differing in reliability. Children's joint decisions were analyzed for functionality. Furthermore, children's conversations leading to the collective decision were coded and children were additionally asked to justify their collective judgments. Children's joint decision-making was functional. Dyads of children, similar to adults, generally preferred the individual judgment with higher reliability as the collective judgment. In contrast to adults, however, children performed a number of failed joint decisions, deciding for the individual judgment marked at the blurred window. During their decision-making children engaged in various kinds of meta-talk including ratings about their own visual access and their individual difficulty to solve the decision task. Furthermore, however to a lesser degree, children communicated the confidence in their individual judgments and provided their decision partner with reasons to justify the correctness of a specific joint decision. Moreover, dialogs that included specific parameters led to a higher number of successful decisions. These results suggested that preschoolers were able to engage in collaborative decision-making successfully. However, the ability to reflect on the process by stepping back and jointly examining the evidence still remained fragile in 4- to 6-year-old children. Findings of Study 3 will be discussed in terms of limits in metacognitive awareness that may mislead dyads in their decision process.

## 11. Selective social belief revision

Study 1, which investigated ontogenetic roots of social belief revision (advice-taking), was motivated by two lines of research: social psychological research on advice-taking in adults (for reviews see Bonaccio & Dalal, 2006) and developmental research on selective social learning in children (for reviews see Harris, 2012). Work in social psychology has shown that adults take advice selectively as a function of the advisor's relevant attributes (e.g., competence, reliability, etc.). However, although systematically revising beliefs mainly in response to competent rather than incompetent advisers adults did not use advice in optimally rational ways. In particular, adults overweighted their own initial judgments and in relation underweighted advice - a phenomenon called *egocentric advice discounting* (e.g., Yaniv, 2004). Several accounts have been put forward to explain such discounting behavior in adults, for example, "privacy of thoughts" (Yaniv & Kleinberger, 2000) or the "mere ownership effect" (Beggan, 1992). However, so far there is no general agreement about one mechanism underlying EAD. By investigating the developmental roots of social belief revision it might be explained how and why adult judges fall short of perfect rationality. Developmental research, however, focused on selective learning of new information. From these studies it is known that children learn new words and facts similar to adults heeding advice and they prefer informants as a function of several attributes such as knowledgability, confidence or reliability. In contrast, the capacity to coordinate advices of others with own judgments and to take advice as a function of epistemic status was investigated in children in only one previous study (Rakoczy et al., 2015). The rationale of Study 1 was, thus, to investigate the early development of selective social belief revision systematically.

### 11.1. Belief revision as a function of epistemic status

In Study 1 participants were given a visual perception task in the format of a judge-advisor system (JAS) that was suitable for both adults and children. In this paradigm, subjects made an initial judgment, then heard about an advisor's judgment, and lastly were asked to make their final judgment – allowing the parametric calculation of the degree to which subjects take the advice into account. In different conditions, the quality of the subject's and the advisor's perceptual access was manipulated by having them look through clear or blurred windows.

The results of Experiment 1 were the following: adults engaged in selective advice-taking consistent with previous research: they took into account both their own informational access (revising their initial judgments more when they themselves saw poorly) as well as



the advisors' (revising their initial judgments more in response to advisors with better visual access). Children, in contrast, only revised their beliefs selectively as a function of their own informational access, but did not seem to differentiate according to the advisor's information access.

This finding seems to contradict previous research on advice-taking in children showing that children placed stronger weight on more competent advisors (Rakoczy et al., 2015). Consequently, two additional experiments were conducted to investigate whether the failure of children to consider their advisor's visual access in Experiment 1 reflected true limitations in young children's competence of selective belief revision or demonstrated performance limitations due to specific demands of the current task. Comparing the current task to that used by Rakoczy et al. (2015), three characteristics were identified that may have made the task in Experiment 1 cognitively too demanding for children. Firstly, participants had to keep track of both their own and another agent's perceptual access, which varied across trials. Secondly, they made continuous judgments rather than choosing one out of a small set of responses. Thirdly, they could choose from a wide range of revision strategies in response to the advisor's judgment.

Experiment 2 and 3 tested for children's performance by removing some of these complexities. Experiment 2 addressed the question whether children may not have differentiated between good and poor advice because advice quality varied within the advisor and across trials. Therefore, children were faced with a very similar task which, however, required belief *formation* rather than belief *revision*. Subjects were confronted with two informants who had different perceptual access (one looked through the clear, the other through the blurred window) and were asked to choose between them. Children now had little difficulty in choosing the agent with the better informational access, even though it varied within advisors such as in Experiment 1. Thus, Experiment 2 demonstrated the ability of young children to track the quality of advice by considering relevant situational factors. Experiment 3 addressed the question whether the continuous response format and the subsequent requirement to integrate two continuous judgments may have made the task in Experiment 1 so difficult that children were no longer willing or able to track the advisor's visual access or to consider it when making their final judgments. Accordingly, children were faced with a very similar task only differing in that the belief revision required was categorical rather than continuous. Results revealed that children now revised their judgments in a functional manner, heeding advice more often when they themselves saw poorly, and also when the advisor saw more clearly. Thus, children in Experiment 3 behaved qualitatively

similar to the adults in Experiment 1, and the way they selectively heeded advice also mirrored results of previous research on advice-taking in young children (Rakoczy et al., 2015).

In summary, these results suggest that young children are capable, in principle, of selective social advice-taking, revising their beliefs quite sensibly as a function of their own and the informants' perceptual access. This capacity, however, seems still fragile and easily becomes masked by performance factors.

### **11.2. Methodical considerations and limiting performance factors**

Children's performance in Experiment 1 might have been limited by the pragmatic structure and demands of the task. The JAS, by its nature, involves repeated requests of a participants' judgment within the same task. Children might have unsystematically revised their judgments on pragmatic ground simply because they thought the repeated question implied that their first answer was wrong. Indeed, previous research suggested that children tended to revise judgments when being asked repeatedly (Siegal et al., 1988). In addition, the task in Experiment 1 may have been too difficult for children since they needed to take into account and consider a lot of background knowledge about conditions for visual perception and its relation to information and knowledge. Recent theoretical and empirical work suggested that sophisticated selective learning builds on and presupposes specific background knowledge (Hermes et al., 2018a; Sobel & Kushnir, 2013).

Another limiting factor might have been the fact that, at any time, subjects in Experiment 1 were only confronted with one single advisor. The lack of contrast to another informant perhaps promoted to indiscriminately take into account this only advice independently of its quality. This seems plausible considering existing research: Most studies that documented competence for selective trust have confronted children with two informants differing in their quality of information. Contrarily to studies with only one informant, children often had difficulties in overcoming uncritical trust even if informants were obviously unreliable or incompetent (Jaswal et al., 2010; Vanderbilt et al., 2011; 2014; for critical discussion see Jaswal & Kondrad; 2016).

As a consequence, this demand to selectively suppress or modulate naïve trust towards the advisor (depending on epistemic position) may have posed too high demands on children's executive function. Related research suggested that any tasks that required the coordination of diverging perspectives, in particular those in which one of these perspectives

pertains to one's own current beliefs, were demanding in terms of executive function (Fizke et al., 2014; Rakoczy, 2010). In line with this, previous research found that the ability to overcome indiscriminate trust towards single informants builds on the development of executive function (Jaswal et al., 2014). Similar research suggested that children both had sophisticated rational and less sophisticated heuristic strategies for social learning in their cognitive repertoire in parallel; and used the former if the executive and other demands of the tasks allowed, but reverted to the latter otherwise (Hermes et al., 2018b). Similar aspects may have played a role in the current Experiment 1: Due to overwhelming task complexity (presenting a continuous measure in combination with general demands of the task), children may have followed simpler heuristics such as "If you see clearly, take little advice; if you see poorly, take more advice".

In summary, the present findings, thus, showed that young children were capable, in principle, of selective social belief revision or advice-taking: Under the right circumstances, they were capable of revising their judgments more sufficient when their own initial information was of poor quality, and more often in response to an advisor whose informational access was better. These findings go beyond previous studies on selective social learning in two crucial ways: they suggest that children do not only selectively acquire new beliefs from others, but also engage in selective belief *revision* – a more complex and demanding epistemic practice. In addition they suggest that children do so in contexts in which their initial judgments reflect proper beliefs (of varying certainty) rather than mere guesses. Future studies might investigate more in detail under which circumstances this fragile early competence becomes realized, or remains masked by performance factors.

### **11.3. Developmental trajectories of selective advice-taking**

Interesting questions for future research can be posed about the trajectories and the underlying motors, of the development from these early basic forms of selective advice-taking to fully-fledged, and more sophisticated adult forms of social belief-revision. For example, do children simply become more accurate and epistemically sensitive over time, or are there qualitative developmental leaps? One particularly interesting aspect in this context is whether there seemingly paradoxical developmental patterns exist such that social belief-revision, in some sense, worsens over developmental time. From other research in social psychology it is known that adults weighted advice in a selective and systematic but not fully rational fashion (for reviews, see Bonaccio & Dalal, 2006; Feng & Mac George, 2006;

Snizek & Buckley, 1995). In particular, adult advice-taking is limited by the fact that subjects weighted others' advice less than optimal; a phenomenon known as *egocentric advice discounting* (EAD, Yaniv & Kleinberger, 2000; Yaniv, 2004).

From a cognitive point of view, it is not fully understood what the basis of EAD in adults is. And from a developmental point of view, it remains unclear how these limits of advice-taking develop. Broadly, two courses of development are possible: Firstly, and in line with naïve intuition, children may become more rational and less egocentric over time (following the general trend of declining egocentrism). Secondly, and much more counter-intuitively, the reverse may be true: in a sort of developmental paradox, children may actually start out more rational than adults and decline their rationality over development time. This possibility agrees with an exciting idea recently discussed in cognitive and comparative psychology. According to this idea, animals are sometimes more rational than humans because they lack some of the experience it needs to develop less-than-fully-rational heuristics (Stanovich, 2013).

Analogous patterns may apply to the development of advice-taking: The cognitive foundations of EAD may be complex and thus require protracted development. One promising account in the advice-taking literature – the so-called “privacy of thoughts” approach - explains EAD in terms of metacognitive processes. The basic claim describes that adults prefer to stick with their own judgment because they have more insight into their own reasons compared to those of the advisor (Yaniv & Kleinberger, 2000). Another account proposes the *mere ownership effect* (Beggan, 1992) as mechanism underlying EAD. According to this, adults see opinions as intellectual property, with increased subjective value of one's own ideas (Abelson & Prentice, 1989). Advice might thus be perceived as devaluing one's own ideas and discounted as a consequence. From a developmental point of view, it is known that both metacognitive abilities (Kloo & Rohwer, 2012) and a basic notion of intellectual property (Shaw et al, 2012) develop in protracted fashion. If these accounts were right, the development of advice-taking may indeed reveal more complex and seemingly paradoxical patterns. Systematic future research that implement comparable paradigms in various age-groups is needed to test for conclusive developmental patterns.

## 12. Argument evaluation in preschoolers

Study 2 investigated the impact of the justification of advice on judge's decisions regarding advice acceptance. As supported by findings of Study 1, 4- to 6-year-old children preferred advice as a function of the advisor's competence and weighted this competence against their own epistemic status. In real life situations, however, advice-taking is more complex than aggregating two judgments only. Advice is often given in the form of interpersonal communication and supported by arguments. Discounting of advice was explained by judges having limited access to advisor's reasons for their proposals (Yaniv, 2004). This led to the hypothesis that making their state of knowledge more salient to judges would increase rates of advice-taking.

The account of epistemic vigilance (Sperber et al., 2010) describes successful decision makers as vigilant listeners weighing information and proposed justification as a function of quality. Following this argumentation advice-taking would not increase only by the presence of justification alone. Accurate reasoning may result in some arguments to increase advice-taking while others might be ignored or even have a decreasing effect. Irrational arguments could be interpreted not just as limited argumentative skills but rather as limited reasoning skills. Thus, expose an advisor as not understanding the logical structure of task demands and subsequently devalue their advice for specific tasks or in general. In social psychology research adults have been proven to weigh advice for its quality and evaluate arguments for their strength to present conclusive evidence. Subsequently, one would assume that an argument presented to support a specific piece of advice is processed for its strength of evidence just like other cues in the advice-taking context are processed to assess the quality of the advice (Yaniv & Kleinberger, 2000). So far the effects of advice justification on advice-taking in children has not been studied. The rationale of Study 2 was, thus, to investigate in children firstly, whether advice with supportive arguments is followed more than without advice justification. Secondly, whether the degree of advice-taking is influenced by the quality of arguments, and thirdly, whether an informant disqualifies by demonstrating insufficient argumentation skills leading children to even discount the informant's credibility in the future.

### 12.1. Advice-taking as a function of argument quality

In Experiment 1 and 2 children participated in the same visual perception task as in Study 1. In a JAS paradigm participants started with an initial judgment about the location of a black marking on a colored wooden stick, they then were presented with an advisor's judgment, and subsequently were asked to make their final judgment. The measurement of these three judgments allowed advice-taking, meaning the weight of judges giving advice, which is being calculated based on an equation introduced by Harvey and Fischer (1997). Using a within subject design two independent variables were manipulated: firstly, the informational access of the judge and advisor and secondly, the support of advice by arguments. Under different conditions, advice was offered with or without the justification by an argument. Arguments were phrased identically but manipulated in quality by having the advisor look through a clear or blurred window. In Experiment 1 advice was supported with the argument "*I think it is here because I looked through here.*" while in Experiment 2 the argument "*I really want us to take this, I am absolutely sure because I looked through here.*" was used. In Experiment 3 a new decision task was designed following a classical selective trust paradigm and presented in two formats: a single-informant design and a two-informant design.

The results of Experiment 1 were the following: firstly, children engaged in selective advice-taking consistent with findings of Study 1. They were sensitive to their own and the advisor's state of knowledge, revising their initial judgments more when the advisor was better informed and less when they themselves were better informed. Secondly, when advice was justified by an argument children followed the advice significantly more than when advice was unjustified. However, there was no significant difference in the increase of advice utilization as a function of argument quality. Advice justification increased advice-taking similarly in both conditions: when the advisor was seeing clearly or poorly. The results of Experiment 1 are both similar and different from findings of previous studies on advice-taking and argumentation.

Similarly to the current experiment, one study on advice justification in adults (Tzioti et al., 2014) reported judges to significantly update their opinion more towards the direction of the advice compared to a situation of no advice justification (control group). Developmental work on argument evaluation showed preschoolers from around four years of age to favour new information supported by an argument over unsupported information (Mercier et al., 2014). Moreover, 4- to 5-year-old children, just like adults, are sensitive to the presence of connectives. If only one out of two informants used *because* to link the own argument to the statement, participants preferred choices that contained *because* (Bernard et al., 2012). In

contrast, findings of Experiment 1 differed from evaluation behavior shown in adults. Participants in a decision task in JAS format followed justifications as the best fit various test conditions: A similar argument increased advice-taking when offered by a competent advisor, whereas it decreased when offered by an inexperienced advisor. Thus, weights that decision makers attached to the justified advice agreed to what one would expect to be most effective in terms of the advice situation. Furthermore, previous findings on preschooler's argument evaluation showed that children placed more weight on information supported by strong versus weak arguments (e.g., Castelain et al., 2018; Mercier et al., 2014).

Due to the fact that outcomes of Experiment 1 seemed to contradict previous research two additional experiments were designed. The aim of these studies was to investigate whether the failure of children at the age of four to six years to distinguish between argument quality as a function of visual access of the advisor resulted from limited argumentative skills or whether it was due to the specific design of Experiment 1. Two characteristics may have been responsible in this respect: the perceived quality of argument and the consistency of the advisor. Firstly, phrasing of the argument in Experiment 1 might have been too neutral to be judged as weak argument in the *high-poor* condition and was, therefore, rather interpreted by children in a reasonable fashion (as a supporting argument). Secondly, advice was given by the same advisor across conditions. Children might not experience problems to form an overall impression of one's argumentative skills, however, they may do have difficulties adjusting the argumentative competence of an advisor as a function of condition. Furthermore, based on the design of the current task, decreasing advice-taking in the *high-poor* condition would not have been functional to optimize the accuracy of the final judgment. Children started by judging the stimulus at the clear window and making their initial judgment already based on optimal information access. In order to perform negative advice-taking – to change the initial judgment into the opposite direction of the advice – would have demanded children to change their initial judgment to the worse. Experiment 2 and 3 tested for children's performance by removing some of these difficulties.

Experiment 2 investigated whether phrasing the argument more clearly would allow children to reflect more effectively on the quality of the argument as a function of condition. Accordingly, argument strength was increased by enhancing the speaker's commitment for the own advice. In the *high-poor* condition this means the advisor increased confidence in the advice although judging the stimulus by looking through a blurred window. This questioned the ability of the advisor to correctly calibrate the evidence the advice is based on. Consequently, one would expect less difficulties to distinguish argument quality between

the two conditions. Additionally, children were confronted with two different advisors, i.e. one acting in the *poor-high* condition and the other in the *high-poor* condition.

As a result of this, children distinguished between conditions clearly. An advisor with high visual access and justification of the own advice by a new argument led children to increase their advice-taking significantly more than without advice justification. In contrast, children ignored an advice generated from the advisor at the blurred window although justified with an identical argument. In Experiment 2 children were able to evaluate arguments as a function of quality and subsequently utilized advice functionally. Thus, results of Experiment 1 can not be explained purely by a general inability of children to inhibit a bias to trust other's testimony (Jaswal, Croft, Setia & Cole, 2010; Jaswal, Pérez- Edgar, Kondrad, Palmquist, Cole & Cole, 2014) or cognitive limits to process false testimony. Previous research often explained children's trust in inaccurate informants by the fact that they would not yet be sensitive enough to the concept that someone might deceive them, become confused about the communicative intent of inaccurate informants (Heyman, 2014; Mascaro & Morin, 2014), or might not yet be able to evaluate other's claims based on evidence (Sobel & Kushnir, 2013).

Experiment 3 addressed three questions: firstly, whether children prefer new information justified by strong versus weak arguments. Secondly, whether children discount unskilled argumentators, and thirdly, whether children use argumentative skills as cue to infer an informant's reliability in the future. Accordingly, a new decision task in the form of a classical selective trust paradigm was designed. The task was presented to children in a single- and two-informant design. Children preferred information supported by strong arguments compared to weak arguments. This finding confirmed findings of Experiment 1. Moreover, Experiment 3 extended these findings by firstly, using a new decision task and secondly, by speaking on the selection of new information without children having informational access themselves. However, participants did not go one step further and discounted an informant previously lacking argumentative skills by actively rejecting the judgment. Furthermore, results indicated that preschoolers determined the informant's credibility based on the quality of explanations.

Overall results revealed that children from around four years of age were capable of evaluating fundamental arguments as well as preferring strong over weak arguments in situations of new information learning and belief revising. This capacity, however, might not have been fully observable in Experiment 1 and 3. Both experimental designs might have imposed difficulties on children and masked their true competence on a performance level.



## 12.2. Methodical considerations and limiting performance factors

One factor that might have increased task demands in Experiment 1 was phrasing of the argument. Ambiguous phrasing might have made it more difficult for children to distinguish the strong from the weak argument as a function of their advisor's visual access. As a result children judged the explanation "*I think it is here because I looked through here*" as a strong argument in both situations when the advisor saw clearly or poorly. Argument quality was implemented based on the advisor's epistemic status to allow presenting identical wording of an argument either in form of a sound or absurd judgment justification as a function of context. Previous research has shown children's sensitivity to the length of explanations (Corriveau & Kurkul, 2014), argument structure (Leman, 2002), and usage of connectives (Bernard et al., 2012). This sensitivity might have distracted children in their performance to judge arguments by epistemic content. Instead, they might have lost track due to complex structures of sentences or got misled by the signalling effect of connectives like *because*. However, the aim of Experiment 1 was to investigate children's ability to select for arguments as a function of epistemic content. The ambiguity of the argument might have affected children's reasoning on argument quality in several ways.

Firstly, children might have increased their advice-taking in the *high-poor* condition because they interpreted the advisor's explanation in a rational way like, "*This is good advice, because I have very good eyes.*" or as sign of good intentions in the sense of "*This is good advice, because I tried the best I can.*". Both, competence and good intentions are characteristics of advisors known to have an increasing effect on advice utilization (Bonaccio & Dalal, 2006). Moreover, the *principle of charity* (e.g., Wilson, 1959; Davidson, 1984) describes listeners as being motivated to interpret speaker's statements in the most rational way possible and, in the case of arguments, considering the most benevolent interpretation. The principle may be invoked to make sense of a speaker's explanations, especially when one is unsure of the precise meaning. In the current task the precise meaning in a way to be understood by children and to be used as basis for subsequent advice utilization was the epistemic value of the argument; i.e. evidence the advisor grounded the judgment on. Thus, children might have believed that the advisor had a good reason for the judgment, even though this reason was not clearly pronounced by the advisor.

Secondly, lacking clear interpretability of argument quality in Experiment 1 children might have used other heuristics that did not fit well with basic skills of argumentation, such as reviewing arguments for their epistemic content. The advisor giving a reason for own judgment might have been interpreted as a mark of dominance. Regarding children the

ambiguous argument “because I looked through here” under the *high-poor* condition seemed similar to other expressions, such as “because I say so” and “that’s the way it is”, often offered to children by adults or other dominant individuals (Grusec & Goodnow, 1994). Children might not have interpreted the expression as weak argument but rather as simple restatement of the advisor’s position used by the advisor to remind the child of their authority. A recent experiment suggested that 3- to 5-year-old children used dominance relations in their selection of testimony (Bernard et al., 2014). One cue applied by children to infer dominance was decisional power in the form of an informant repeatedly imposing one’s own goals onto others. Ambiguous arguments could, therefore, affect advice-taking through an indirect route - by indicating the dominance of the speaker, which in turn would then influence the evaluation of argument justifying the advice. Alternatively, children could attempt to be polite or to ingratiate with the informant (Jaswal & Kondrad, 2016). Several recent studies suggested that children sometimes endorsed information from previously inaccurate speakers to pursuit social rather than epistemic goals, including a desire to belong (Song, Over, & Carpenter, 2015), maintain positive relationships (Hartup, 1989), and be liked (Fawcett & Liszkowski, 2012).

One further factor may have made the task in Experiment 1 cognitively demanding for the children. Advice and advice justifications was given by only one advisor across all conditions. Previous research so far has not investigated the circumstances under which children may use argumentative skills as cue to infer informant reliability and which task demands might cause difficulties for children to contrast arguments as a function of epistemic content. Research that applied a single-informant paradigm to study selective trust, however, appears to provide evidence of children’s limited ability or willingness to distinguish accurate from inaccurate informants in such specific test situations (Kim et al., 2017; Vanderbilt et al., 2014). When children’s ability to evaluate other’s argumentative skills were not fully developed at the age of four to six years it perhaps became too demanding to judge arguments presented by the same advisor to be strong versus weak as a function of condition. In contrast, it might become easier for children to distinguish between argument quality when advisors are consistent in their ability to argue.

### 12.3. Argumentative skills as cue for selective learning and future credibility

In Experiment 3 preschoolers showed a preference for strong versus weak explanations when learning new information. This finding is consistent with previous research indicating that from around three years of age children possess basic skills of argument evaluation and use these skills to selectively learn from testimony of others (Mercier et al., 2014). In classical selective trust paradigm different degrees of reliability of informants become established via labelling of objects – one informant is labelling the object with correct names whereas the other informant uses incorrect names. The current Experiment showed that preschoolers were able to process not only single words to access the quality of informants but used complete causal subclauses to judge others competences based on their utterances (for similar findings see Corriveau & Kurkul, 2014). Children did not develop their preference for the more accurate informant just depending on the target word *because*, since both the strong and the weak argument were introduced using this causal connective. Moreover, the accuracy of the argument was not determined by different reasons. The strong as well as the weak argument were phrased identically but differed in quality as a function of the situation in which they were presented to justify testimony. Additionally, previous research used circularity to establish weak arguments. A circular argument does not present any new information, but instead is mostly a pure repetition of the informant's judgment (e.g., „I want us to take this one, because I want this.“). Thus, children may succeed by applying the simple heuristic to evaluate various arguments only for new informational content without the verification of its quality. The current experiment most likely for the first time in research contrasted two arguments that both contained epistemic information and thus demanded children to weigh arguments based on different degrees of epistemic values without setting the epistemic value of one argument to equal zero.

Furthermore, it was examined whether preschoolers use argumentative skills of informants as cue to infer credibility of testimony in subsequent tasks. In Part 2 and 3 of Experiment 3 informants gave testimony without justification, firstly, presented as single-informant paradigm and secondly presented as two-informant paradigm. The single-informant-paradigm (part 2) aimed at investigating whether children actively reject information from an informant previously presenting weak arguments. When being presented with both informants at the same time (two-informant paradigm, part 3) children selectively endorsed the claims about softness of the informant who had previously used strong explanations. This finding is in accordance with children explicitly judging this informant as “better” in the contrast judgment question. Experiment 3 supported the children's ability not

only to monitor explanations for quality but also to use this information to make judgments about an informant's future credibility, however, surprisingly this was true only regarding the modality touch. It seems that preschoolers can use the monitoring of explanation quality as strategy to selectively learn from informants but only in certain situations and under certain conditions. This indicates that children's ability to monitor explanations for quality seems still fragile at the age of four to six years. Future research might explore developmental sequences of the cues used for explanation monitoring and reveal whether younger children are more limited in their use of multiple strategies to selectively endorse information. Furthermore, reasons should be examined systematically why children's information-seeking behavior might differ across various modalities of perception within the same task design.

#### **12.4. Discounting of poor argumentators**

Children did not actively reject information from an informant previously given weak arguments. In a single-informant paradigm where children choose between testimony of the informant and an alternative solution they did not selectively endorse the alternative. This finding is in agreement with research on children's monitoring of single-word utterances as a function of the number of informants. In a learning context in which children were faced with conflicting testimony from two informants they relied on the prior reliability of the informants and selectively preferred to trust the more reliable source versus the less reliable one (e.g., Koenig & Harris, 2005). In tasks in which children were faced with testimony from a single informant and not having prior knowledge or proof of the informant's testimony (e.g., to see or touch a stimulus themselves) they generally trusted the single informant regardless of the informant's past accuracy (e.g., Kim et al., 2017; Vanderbilt et al., 2014). The current experiment extended this knowledge by showing that children were willing to endorse information even though the informant presented entirely inaccurate explanations, which might be perceived as stronger signals of inaccuracy than just single false labels.

Children's performance on the explicit contrast-judgment questions suggested that they did not forget about this informant's weak arguments, since children explicitly judged the high-quality informant as the one giving the "better reasons". However, similarly to the single-testimony endorsement-task, children did not judge the weak argumentator as someone providing "not so good reasons". Children's general hesitation to, firstly, reject novel information from unreliable sources and secondly, explicitly judge informants as unreliable implied that preschoolers may have a natural aversion to perceive a source of information

negatively. Thus, Experiment 3 contributed to the discussion of two conflicting accounts that might explain why children selectively endorse testimony of accurate informants in a situation where two contrasting sources are pitted against each other: Do they prefer the accurate informant or do they reject the inaccurate informant? The current results showed the tendency of children to consider both accurate and inaccurate informants as generally trustworthy rather than seeing inaccurate informants as being absolutely unreliable; hence, they have been willing to learn from the inaccurate informant if the informant was the only source of information available.

Another question was why children might hesitate to explicitly judge an informant negatively or to actively reject testimony from speakers with a history of inaccuracy. One possible explanation is that children used different strategies as a function of number of informants (one-informant paradigm versus two-informant paradigm). According to the dual-process account of selective trust (Hermes et al., 2018) children's reasoning about testimony involves two qualitatively different processes: Type I processes, which operate fast and yield heuristic default judgments, and Type II processes, which are relatively slow and yield rational judgments based on inferences. Specifically, Type I heuristics may be overwritten by the more sophisticated Type II rationalizations, especially when the heuristics provide no conclusive solutions to the existing problem. Thus, in the single-informant task of Experiment 3, children may have applied a general trust heuristic (e.g., Jaswal et al., 2010) such that they generally trusted social information regardless of the informant's history. However, when confronted with multiple competing informants, the general trust heuristic no longer worked and children may have engaged in Type II inferences based on informants' history of reliability to evaluate and respond to testimony.

It is also possible that children avoided to reject testimony from speakers with a history of inaccuracy because their judgment was part of a broader tendency to view people in a positive light (e.g., Boseovski, 2010; Lockhart, Chang & Story, 2002) For example, children required less behavioral evidence to make positive attributions versus negative attributions (Boseovski & Lee, 2006). Children may be reluctant to view people negatively in the absence of very strong evidence. A related possibility is that children have learned about the usual accuracy of people's testimony. Consequently, their default assumption was to accept what they hear in the absence of explicit alternatives.

### **13. Collective decision-making**

The goal of Study 3 was to investigate preschooler's decision-making in an interactive scenario. The final study of this dissertation builds upon findings of Study 1 and 2 that revealed the following: Firstly, presented with a classical JAS paradigm 4- to 6-years-old children revised their perceptual judgments selectively as a function of their own and an advisor's epistemic status. Secondly, children were sensitive to arguments and revised beliefs more sufficiently when advice was justified by explanations. However, when arguments were not instantly phrased soundly enough they were not weighted effectively. Study 3 finally gave children the opportunity to inquire more information in order to eliminate ambiguities and to discuss the quality of evidence supporting specific judgments in form of an interpersonal communication. Such social interactions, as proposed by Mercier (2016) are ideal situations to unfold reasoning capabilities in full measure because participants have the possibility to produce and to evaluate each other's arguments. In particular, in dialogs of conflict challenging each others argumentations leads to an increase in cognitive efforts and produces the best outcomes: during back and forth argumentation and counterargumentation, weak arguments are rejected and good ones finally convince the decision partners.

Successful argumentation between decision partners results from effective meta-talk that allows to assess the quality of arguments and/or evidence directly. In order to implement effective argument strategies one has to illustrate own reflection about the quality of own judgment based on informational access and confidence in the own judgment. This leads to the hypothesis that dyads engaging in more advanced meta-talk and producing stronger arguments achieve a higher number of successful joint decisions. However, so far no studies seem to have been investigating the functionality of meta-talk as a parameter for decision quality. The rationale of Study 3, thus, was to analyze the successfulness of preschooler's collective decisions-making, the dialogs leading to joint decisions, and the effectiveness of meta-talk strategies.

#### **13.1. Collective decision-making in children and adults**

In Study 3 children and adults were presented with the same visual perception task already used in Study 1 and 2 except that dyads of subjects solved the task jointly. In the described design dyad members firstly solved the task individually, secondly they compared and discussed their individual judgments, and thirdly they jointly decided for a final judgment. The

quality of individual judgments was manipulated by restricting dyad members to either high or low visual access regarding the stimulus. In dialogs conflicting judgments needed to be proven jointly for supporting evidence and accordingly be integrated to reach the best possible decision outcome, finally allowing to analyze the quality of joint decisions, the communicative dialog this decisions were based on, and a possible causal interplay of these two factors.

The main findings of Study 3 included the following: Firstly, collective decisions favoured the high quality judgment. In 92 % of all decisions adults favored the judgment based on high visual access. In tasks involving children, partners mainly agreed on the judgment supported by the strong evidence. Judges originally sitting behind the clear window were able to present their better judgment successfully and judges behind the blurred window seemed to be willing and able to endorse this contradicting judgment even without the possibility to validate it's quality directly. However, in 20 % of collective decision-makings children failed to choose the better judgment. Unsuccessful decisions were unlikely to result from limited understanding of their partner's quality of visual access, since several manipulations check parameters validated a full comprehension of the experimental design for all participants.

Secondly, analysis of the dialog structure in children and adults revealed that both informed their decision partners about their own visual access or the difficulty of the perception task. Additionally, in 50 % of dialogs adults shared ratings of subjective confidence with the decision partner. Previous research suggested that the functional role of such shared metacognition is to provide information on the strength of an individual's opinion and it's reliability based on situational evidence (Bahrami et al., 2012). Moreover, in a certain number of trials adults linked the quality of judgments with the perceptual evidence they resulted from by giving reasons for their evaluation of individual judgments (e.g., "My judgment is of higher quality because I had clear visual access."). In contrast to adults, children included significantly less ratings of visual access, subjective confidence, and explanations for judgment proposals. In particular, children reflected on their uncertainty and gave reasons for proposals only in a few dialogs (13 % and 6 %). Being asked for reasons supporting joint decisions adults expressed more often right reasons to justify final judgments than children while children presented more wrong and irrelevant reasons than adults .

Finally, decision processes in children that included sharing the quality of each other's perceptual access more often resulted in successful collective decisions. This finding supports a potential interplay of meta-talk strategies and effective decision-making of dyads,

however, the occurrence of further dialog parameters (e.g., sharing of confidence ratings) did not result in differing degrees of successfulness. Furthermore, children justified successful decisions more often with ambiguous reasons while unsuccessful decisions were more often justified by wrong reasons.

### **13.2. Successful collective decisions in children**

In 80 % of joint decisions dyads of children were able to reliably settle on the judgment with the best supporting evidence. Children were presented with a complex task requiring to mentally map two perspectives of the same stimulus in parallel. These two perspectives, firstly, were formed by different dyad members (child A or child B), secondly they held opposite degrees of informational access and accordingly strengths of evidence, and thirdly, thus, resulted in judgments differing in quality. The success rate of joint decisions suggested that preschoolers can engage in collaborative reasoning successfully even under such demanding test conditions. Children were able and willing to collaborate in order to solve a schematic version of a numerical task. This goes beyond previous studies on collective decision-making presenting children at this age with rather engaging playful scenarios like building a zoo together. Much research has demonstrated that there are important differences in how children think and communicate about animals and artifacts (Rakison & Poulin-Dubois, 2001; Brandone & Gelman, 2013). A zoo-setting may have been an especially facilitative context for children to collaborate and find nice homes for animals. Animals are living creatures that may fast catch children's attention and are easier to relate to. In contrast, transferring a mark on a stick onto a picture of the stick seemed to be a more theoretical, less exciting task that demanded higher concentration and motivation to engage into a joint reasoning process.

Reasoning with a peer to make a joint decision involves making a proposal and justifying it with relevant facts based on common ground assumptions. Children participated in the task equipped with skills of joint attention allowing dyad members to build up a joint attentional frame that included joint assumptions such as: "Somewhere on the stick is a mark.", "We jointly decide where the mark is located.", or "When you look at the stick you can see the mark.". The establishment of a joint attentional frame allows decision partners to cooperatively exchange information and accordingly form judgments jointly about potential solutions and their quality. Judges who initially sat behind the blurred window were introduced to an alternative solution by their partners and willing to incorporate their partner's



perspective into a joint judgment of higher quality by reaching a mutual agreement. Accordingly, “poor judges” were willing to revise their own prior judgment as soon as it was acquired based on an unreliable informational access and they then accepted the alternative proposed by their peer. This implies changes at an epistemic level (knowing what oneself knows compared to what others know), in which „poor judges“ adopt the „high judge’s“ beliefs as superior to achieve a good collective performance.

However, the question remains why dyads failed to reach the correct joint decision in 20 % of cases. Phenomena discussed in social psychology to explain collective failure are difficult to apply in the current design. “Social loafing” describes collective situations in which individuals invest less effort in the presence of others leading to reduced overall group performance. Shared responsibility for possible failures and accordingly, the lack of holding a specific member directly responsible was indicated as explanation (Karau & Williams, 1993). This mechanism, however, can not explain unsuccessful joint decisions in the current experiment because all judgments (individual and joint ones) were precisely documented for the dyad and both participants left little room for sharing the responsibility of wrong joint decisions. The dyad member who would have led the group to the wrong decision easily could have been identified. This feature of the experimental design was also implemented to avoid motivation loss of individuals participating in group performances (Williams, Harkins, & Latané, 1981).

“Groupthink” (Turner & Pratkanis, 1998) is another case of collective failure and was explained by interdependence of individual decisions (Raafat, Chater, & Frith, 2009). When individuals are not given the opportunity to make their own decisions privately, they subsequently fail to express disagreeing opinions. This phenomenon cannot account for failed decisions in the current experiment because individual decisions were always in the first place made privately and independently.

Finally, the “hidden profile” paradigm (e.g., Stasser & Titus, 1985) is another extensively studied case of collective failure. This happens in situations in which one individual holds unshared but highly relevant information to reach optimal outcomes. In the current experiment dyads were actually composed of members with dissimilar knowledge profiles; one member having clear visual access to the stimulus and another having restricted visual access. However, the difference in participants’ informational access was not exclusive but common knowledge because both dyad members were introduced to the set-up during an extensive familiarization phase and even remembered within test sessions by participating in manipulation check procedures.

The exact kind of mechanism that may have caused children dyads to fail significantly more often than adult dyads need to be investigated more in detail in future studies. One possibility to be addressed for further explanations might be the influence of interpersonal competition between dyad members. Although individuals were not differentially rewarded for their decisions and there was no incentive for competition it can not fully be ruled out that children from time to time fell back to a more egocentric form of thinking. In return competitive rather than cooperative decision strategies might have been activated, combined with a still fragile understanding of the causal relationship between the quality of the visual access and the subsequent quality of individual judgments. Consequently, dyads may not have been willing or able to decide for the judgment with the best supporting evidence.

### **13.3. Structure of dialogs leading to collective decisions**

Both, children and adults explicitly briefed their decision partners about the own informational status using ratings of visibility or task difficulty in 40 % or 60 % of trials, respectively. Participants announced their subjective state of knowledge, although this evaluation could have been assumed by the other dyad member. This is true because it was part of the dyad's situational common ground and did not need to be explicitly stated, since both partners were introduced to the clear and blurred window situation. Such testimony perhaps did not result from the limited ability of participants to tailor conversation based on the knowledge state of their dialog partner or the common ground knowledge that they shared with each other. Some children dyads even explained the whole test scenario by providing information about which person is sitting where, the location of the clear and blurred windows, the degree to which each person has visual access, and so forth. This kind of testimony may have served to demonstrate agreement between subjective evaluations and the expected objective evaluation of informational access. Thus, functioned to show commitment to the situational common ground and willingness to reach joint decisions based on rational evaluation of individual judgments. Moreover, these evaluations of member's visual ability might be used as everyday expressions of certainty. A testimony such as "I did not see well" might be used alternating or as a substitute for verbal evaluations of judgment reliability. Previous studies undertaking linguistic analysis of the conversations that led to collective solutions of visual tasks showed that adults used expressions of the verb „be sure“ and „seeing“ interchangeably and tended to align with each other's confidence expressions (Bahrami et al., 2012; Fusaroli et al., 2012). For example, if one started the conversation with 'I did not see anything', the other person would most likely respond with some expression

using 'see'. This might especially be true for 4- to 6-year-old children whose ability to clearly distinguish different markers of unreliability, such as limited knowledge access and expressions of uncertainty, is still fragile.

On the other hand, in their testimony children used signal words of uncertainty such as "maybe", "not sure" or "might, could, should" scarcely accounting for only 13 % of all trials. Children might have referred to their visual access only because it was perhaps easier for them to simply state visibility rather than in a more complex context reflecting on the subsequent quality of the judgment. Precise evaluations of confidence would require participants to actively introspect about their perceptual experience and then indicate their internal, metacognitive estimate of the reliability of their judgment. This is a costly cognitive task that requires allocation of top-down attention and therefore its implementation would neither be automatic nor free of costs (Corallo, Sackur, Dehaene, & Sigman, 2008).

Adults on the other hand expressed significantly more uncertainty than children. These evaluations seemed not to conform with the quality of their individual judgments. Judges initially sitting behind the clear window regularly presented their high quality, individual judgment accompanied with markers of uncertainty. Such behavior might be a strategy to demonstrate willingness to work collaboratively and value their partner's contribution to a successful joint decision rather than demonstrating uncertainty about the quality of a judgment drawn behind the clear window. This observation agrees with concepts discussed in the framework of the *Politeness theory*, originally proposed by Brown and Levinson (1987). It is argued that every member of a society has a face, which is defined as one's public self-image. A positive labeled face means the desire of people to be appreciated or approved of by other members of a society. Contradicting testimony can potentially carry the risk to cause other group members to lose their face. In order to minimize this risk speakers tend to use politeness strategies, for example, linguistic politeness. Furthermore, dyad members were instructed by the experimenter to solve the task collectively and thus, might have felt socially obligated to treat decision partners equally.

Joint decisions did not result from argumentative discourses between dyad members in the current study. Decision partners did not discuss individual judgments in a back-and-forth-fashion by exchanging arguments and counterarguments. This finding might be caused by the overall design of the task. Both dyad members were extensively introduced to the two qualities of visual access the individual judgments were based on and could easily determine the distribution of informational access between both partners in every test trial. In this respect a simple strategy such as always deciding for the judgment drawn behind the clear

window would have been fully successful. However, participants were instructed by the experimenter to compare their individual judgment, discuss it, and jointly decide for a final judgment. That brought up the question why participants hardly gave reasons spontaneously.

On one side this is surprising since the possibility to engage in discussions was proposed to increase reasoning efforts for providing convincing arguments and support own positions or challenge positions of dialog partners (Mercier, 2012). On the other side live, real-time discourse is a sophisticated interaction between different perspectives or opinions, which especially applies to 4- to 6-year-old children. At each moment one was either speaking or processing what the other person said to potentially criticize the opponent's statement. Moreover, several egocentric biases of communication were identified that may have impaired decision outcomes resulting from verbal dialogs (e.g., overestimation of transparency according to Gilovich et al. (1998). Considering how cognitively demanding and error-prone it is to produce a reasonable argument based on common ground assumptions it seems more feasible (or even rational) to simply disambiguate informational access and enter argumentative discourses only if essential for successful joint decisions. Both, children and adults were similar in not producing a high number of arguments spontaneously. However, when invited by the experimenter to present arguments to justify joint decisions for specific judgments, adults expressed significantly more right reasons than children who expressed more often wrong and irrelevant reasons. This supports the conclusion that children's understanding of the causal relationship between the quality of the visual access and the subsequent quality of individual judgments is still fragile at the age of four to six years.

#### **13.4. Success as a function of meta-talk and argumentative strategies**

Adults decided successfully regarding joint decisions in 92 % of trials, thus, communicative strategies that might have led to more or less successful joint decision were not identified in the current experiment.

Children dyads unlike adults differed in the successfulness of their joint decisions as a function of meta-talk strategies. Dialogs that included ratings of individual visual access, correctness of individual judgments, or the difficulty of the task more often resulted in successful joint decisions than dialogs lacking these parameters. This testimony might have functioned to remind each other of the difference in visual access between decision partners and to demonstrate each other's commitment to this situational common ground. These

simpler meta-talk strategies proved successful in cases where they fell on fertile ground, namely that dyad partners intuitively reflected correctly on the quality of their own judgments and “poor judges” were accordingly willing to revise their individual poor judgment. In contrast, in situations where “poor judges” were confident in own knowledge and not willing to revise their individual poor judgment the more advanced meta-talk strategies might have been necessary to increase the success rate of joint decisions. Such strategies may reflect on the situational evidence to clarify why the individual judgment of the “high judge” is preferable and would constitute the best possible outcome of joint decisions. However, the ability to reflect on the process by stepping back to jointly examine the evidence was proposed to emerge only during school years (Kuhn et al., 2013).

Furthermore, the ability to correctly justify joint decisions was found to interplay with the success of joint decisions. Decisions justified by wrong reasons were more often unsuccessful than decisions not justified by wrong reasons. Reasons were coded as „wrong“ if children reflected incorrectly on their visual access, for example, if a „poor judge“ claimed to have better visual access. It seemed that in this cases the „poor judges“ lacked the ability to identify their poor judgment as relevant guess opposed to correct knowledge based on adequate informational access. This interpretation would be consistent with research showing children’s understanding of their own epistemic states to fully develop late in their childhood, in a reflective/explicit form even not before school age (Kloo & Rhower, 2012). However, further research using multiple decision tasks is necessary to disentangle limited competences and potential confining performance factors in 4- to 6-year-old children.

Decisions justified by ambiguous reasons were more often taken by the better judgment than decisions not justified by this type of reasons. Reasons were coded as ambiguous if they transferred the correct message why the high quality judgment was chosen (e.g., „because this judgment is right one“) and, thus, showed the ability to at least intuitively reflect on the quality of individual judgments. Moreover, children presented right reasons in around 50 % of trials that referred to the quality of visual access an individual judgment was based on, for example, „because I saw better“. Right reasons, however, did not correlate with correct joint decisions. Surprisingly, collective decisions favouring the high quality judgment were not significantly more often given to justify decision success versus failure. Even stronger arguments could have been phrased by referring to the reliability of the source of information (looking through a clear opposed to looking through a blurred window), and accordingly emphasising this fact as supporting evidence for specific individual judgments, for example, „Mine is better, because I saw better.“. Giving reasons that reflect

intuitive knowledge but lack the precision of explicitly referring to situational facts as supporting evidence might be the first trajectory step in the development of advanced skills to engage in complex argumentative discourses. However, the current study is to my knowledge the first one that investigated potential correlations between the structure of children's discussions and the functionality of resulting joint decisions. Many questions still remain open and future research will need to uncover more systematically how skills of argumentative discourse and meta-talk strategies become more solid over subsequent development and describe their broader ontogenetic trajectories.

---

## 14. References

- Abelson, R. P., & Prentice, D. A. (1989). Beliefs as possessions: A functional perspective. *Attitude structure and function*, 3, 361-381.
- Bahrami, B., Olsen, K., Bang, D., Roepstorff, A., Rees, G., & Frith, C. (2012). What failure in collective decision-making tells us about metacognition. *Phil. Trans. R. Soc. B*, 367(1594), 1350-1365.
- Bakeman, R., & Adamson, L. B. (1984). Coordinating attention to people and objects in mother-infant and peer-infant interaction. *Child development*, 1278-1289.
- Bascandziev, I., & Harris, P. L. (2014). In beauty we trust: Children prefer information from more attractive informants. *British Journal of Developmental Psychology*, 32(1), 94-99.
- Baum, L. A., Danovitch, J. H., & Keil, F. C. (2008). Children's sensitivity to circular explanations. *Journal of Experimental Child Psychology*, 100(2), 146-155.
- Beggan, J. K. (1992). On the social nature of nonsocial perception: The mere ownership effect. *Journal of Personality and Social psychology*, 62(2), 229.
- Begus, K., & Southgate, V. (2012). Infant pointing serves an interrogative function. *Developmental science*, 15(5), 611-617.
- Bergelson, E., & Swingle, D. (2012). At 6–9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences*, 109(9), 3253-3258.
- Bernard, S., Castelain, T., Mercier, H., Kaufmann, L., Van der Henst, J. B., & Clément, F. (2016). The boss is always right: Preschoolers endorse the testimony of a dominant over that of a subordinate. *Journal of experimental child psychology*, 152, 307-317.
- Bernard, S., Mercier, H., & Clément, F. (2012). The power of well-connected arguments: Early sensitivity to the connective because. *Journal of Experimental Child Psychology*, 111(1), 128-135.
- Birch, S. A., Akmal, N., & Frampton, K. L. (2010). Two- year- olds are vigilant of others' non- verbal cues to credibility. *Developmental science*, 13(2), 363-369.
- Birnbaum, M. H., & Stegner, S. E. (1979). Source credibility in social judgment: Bias, expertise, and the judge's point of view. *Journal of Personality and Social Psychology*, 37(1), 48.
- Bonaccio, S., & Dalal, R. S. (2006). Advice taking and decision-making: An integrative literature review, and implications for the organizational sciences. *Organizational Behavior and Human Decision Processes*, 101(2), 127-151.

- Bond, R. (2005). Group size and conformity. *Group processes & intergroup relations*, 8(4), 331-354.
- Bond, R., & Smith, P. B. (1996). Culture and conformity: A meta-analysis of studies using Asch's (1952b, 1956) line judgment task. *Psychological bulletin*, 119(1), 111.
- Boseovski, J. J. (2010). Evidence for "rose-colored glasses": An examination of the positivity bias in young children's personality judgments. *Child Development Perspectives*, 4(3), 212-218.
- Boseovski, J. J., & Lee, K. (2006). Children's use of frequency information for trait categorization and behavioral prediction. *Developmental psychology*, 42(3), 500.
- Boyd, R., & Richerson, P. J. (2005). *The origin and evolution of cultures*. Oxford University Press.
- Brand, R. J., & Shallcross, W. L. (2008). Infants prefer motionese to adult-directed action. *Developmental science*, 11(6), 853-861.
- Brandone, A. C., & Gelman, S. A. (2013). Generic language use reveals domain differences in young children's expectations about animal and artifact categories. *Cognitive development*, 28(1), 63-75.
- Brehmer, B., & Hagafors, R. (1986). Use of experts in complex decision making: A paradigm for the study of staff work. *Organizational Behavior and Human Decision Processes*, 38(2), 181-195.
- Brosseau-Liard, P., Cassels, T., & Birch, S. (2014). You seem certain but you were wrong before: Developmental change in preschoolers' relative trust in accurate versus confident speakers. *PloS one*, 9(9), e108308.
- Budescu, D. V., & Rantilla, A. K. (2000). Confidence in aggregation of expert opinions. *Acta psychologica*, 104(3), 371-398.
- Budescu, D. V., Rantilla, A. K., Yu, H. T., & Karelitz, T. M. (2003). The effects of asymmetry among advisors on the aggregation of their opinions. *Organizational Behavior and Human Decision Processes*, 90(1), 178-194.
- Burge, T. (1993). Content preservation. *The Philosophical Review*, 102(4), 457-488.
- Cadinu, M. R., & Rothbart, M. (1996). Self-anchoring and differentiation processes in the minimal group setting. *Journal of personality and social psychology*, 70(4), 661.
- Call, J., & Carpenter, M. (2002). Three sources of information in social learning. *Imitation in animals and artifacts*, 211-228.
- Call, J., & Tomasello, M. (1999). A nonverbal false belief task: The performance of children and great apes. *Child development*, 70(2), 381-395.



- Castelain, T., Bernard, S., & Mercier, H. (2018). Evidence that Two- Year- Old Children are Sensitive to Information Presented in Arguments. *Infancy*, 23(1), 124-135.
- Chow, V., Poulin- Dubois, D., & Lewis, J. (2008). To see or not to see: Infants prefer to follow the gaze of a reliable looker. *Developmental science*, 11(5), 761-770.
- Clément, F., Koenig, M., & Harris, P. (2004). The ontogenesis of trust. *Mind & Language*, 19(4), 360-379.
- Clement, R. W., & Krueger, J. (2000). The primacy of self- referent information in perceptions of social consensus. *British Journal of Social Psychology*, 39(2), 279-299.
- Cooper, R. P., & Aslin, R. N. (1990). Preference for infant- directed speech in the first month after birth. *Child development*, 61(5), 1584-1595.
- Corallo, G., Sackur, J., Dehaene, S., & Sigman, M. (2008). Limits on introspection: Distorted subjective time during the dual-task bottleneck. *Psychological Science*, 19(11), 1110-1117.
- Corner, A., Hahn, U., & Oaksford, M. (2006). The Slippery Slope Argument–Probability, Utility & Category Reappraisal. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 28, No. 28).
- Corriveau, K., & Harris, P. L. (2009). Choosing your informant: Weighing familiarity and recent accuracy. *Developmental science*, 12(3), 426-437.
- Corriveau, K. H., & Harris, P. L. (2010). Preschoolers (sometimes) defer to the majority in making simple perceptual judgments. *Developmental psychology*, 46(2), 437.
- Corriveau, K. H., & Kurkul, K. E. (2014). “Why does rain fall?”: Children prefer to learn from an informant who uses noncircular explanations. *Child development*, 85(5), 1827-1835.
- Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends in cognitive sciences*, 13(4), 148-153.
- Davidson, D. (1984). On the very idea of a conceptual scheme. *Inquiries into truth and interpretation*, 183, 189.
- Domberg, A., Köymen, B., & Tomasello, M. (2018). Children's reasoning with peers in cooperative and competitive contexts. *British Journal of Developmental Psychology*, 36(1), 64-77.
- Dunn, J., Bretherton, I., & Munn, P. (1987). Conversations about feeling states between mothers and their young children. *Developmental psychology*, 23(1), 132.
- Eisenberg-Berg, N., & Geisheker, E. (1979). Content of preachings and power of the model/preacher: The effect on children's generosity. *Developmental Psychology*, 15(2), 168.

- Evans, J. S. B. (2002). Logic and human reasoning: An assessment of the deduction paradigm. *Psychological bulletin*, 128(6), 978.
- Fawcett, C., & Liskowski, U. (2012). Mimicry and play initiation in 18-month-old infants. *Infant Behavior and Development*, 35(4), 689-696.
- Feng, B., & MacGeorge, E. L. (2006). Predicting receptiveness to advice: Characteristics of the problem, the advice-giver, and the recipient. *Southern Communication Journal*, 71(1), 67-85.
- Fischer, I., & Harvey, N. (1999). Combining forecasts: What information do judges need to outperform the simple average?. *International journal of forecasting*, 15(3), 227-246.
- Fizke, E., Barthel, D., Peters, T., & Rakoczy, H. (2014). Executive function plays a role in coordinating different perspectives, particularly when one's own perspective is involved. *Cognition*, 130(3), 315-334.
- Fusaro, M., & Harris, P. L. (2008). Children assess informant reliability using bystanders' non-verbal cues. *Developmental science*, 11(5), 771-777.
- Fusaroli, R., Bahrami, B., Olsen, K., Roepstorff, A., Rees, G., Frith, C., & Tylén, K. (2012). Coming to terms: quantifying the benefits of linguistic coordination. *Psychological science*, 23(8), 931-939.
- Ganea, P. A. (2005). Contextual factors affect absent reference comprehension in 14-month-olds. *Child Development*, 76(5), 989-998.
- Ganea, P. A., Shutts, K., Spelke, E. S., & DeLoache, J. S. (2007). Thinking of things unseen: Infants' use of language to update mental representations. *Psychological Science*, 18(8), 734-739.
- Gardner, P. H., & Berry, D. C. (1995). The effect of different forms of advice on the control of a simulated complex system. *Applied cognitive psychology*, 9(7), S55-S79.
- Gilbert, D. T. (1991). How mental systems believe. *American psychologist*, 46(2), 107.
- Gilbert, D. T., Krull, D. S., & Malone, P. S. (1990). Unbelieving the unbelievable: Some problems in the rejection of false information. *Journal of personality and social psychology*, 59(4), 601.
- Gilovich, T., Savitsky, K., & Medvec, V. H. (1998). The illusion of transparency: biased assessments of others' ability to read one's emotional states. *Journal of personality and social psychology*, 75(2), 332.
- Gino, F., Brooks, A. W., & Schweitzer, M. E. (2012). Anxiety, advice, and the ability to discern: Feeling anxious motivates individuals to seek and use advice. *Journal of personality and social psychology*, 102(3), 497.

- Gino, F., & Moore, D. A. (2007). Effects of task difficulty on use of advice. *Journal of Behavioral Decision Making*, 20(1), 21-35.
- Gino, F., & Schweitzer, M. E. (2008). Blinded by anger or feeling the love: How emotions influence advice taking. *Journal of Applied Psychology*, 93(5), 1165.
- Grossmann, T., Johnson, M. H., Lloyd-Fox, S., Blasi, A., Deligianni, F., Elwell, C., & Csibra, G. (2008). Early cortical specialization for face-to-face communication in human infants. *Proceedings of the Royal Society of London B: Biological Sciences*, 275(1653), 2803-2811.
- Grusec, J. E., & Goodnow, J. J. (1994). Impact of parental discipline methods on the child's internalization of values: A reconceptualization of current points of view. *Developmental psychology*, 30(1), 4.
- Hahn, U., & Oaksford, M. (2007). The rationality of informal argumentation: A Bayesian approach to reasoning fallacies. *Psychological review*, 114(3), 704.
- Harris, P. L. (2007). Trust. *Developmental Science*, 10(1), 135-138.
- Harris, P. L. (2012). *Trusting what you're told: How children learn from others*. Harvard University Press.
- Harris, P. L., & Corriveau, K. H. (2011). Young children's selective trust in informants. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 366(1567), 1179-1187.
- Harris, P. L., & Koenig, M. A. (2006). Trust in testimony: How children learn about science and religion. *Child development*, 77(3), 505-524.
- Hartup, W. W. (1989). Social relationships and their developmental significance. *American psychologist*, 44(2), 120.
- Harvey, N., & Fischer, I. (1997). Taking advice: Accepting help, improving judgment, and sharing responsibility. *Organizational Behavior and Human Decision Processes*, 70(2), 117-133.
- Hastie, R. (1986). Experimental evidence on group accuracy. In B. Grofman & G. Owen (Eds.), *Decision Research* (Vol. 2). Greenwich, CT: JAI Press.
- Haun, D., & Over, H. (2015). Like me: a homophily-based account of human culture. In *Epistemological dimensions of evolutionary psychology* (pp. 117-130). Springer, New York, NY.
- Haun, D. B., & Tomasello, M. (2011). Conformity to peer pressure in preschool children. *Child development*, 82(6), 1759-1767.
- Haun, D. B., Van Leeuwen, E. J., & Edelson, M. G. (2013). Majority influence in children and other animals. *Developmental cognitive neuroscience*, 3, 61-71.

- Hermes, J., Behne, T., Bich, A. E., Thielert, C., & Rakoczy, H. (2018a). Children's selective trust decisions: rational competence and limiting performance factors. *Developmental Science, 21*(29).
- Hermes, J., Behne, T., & Rakoczy, H. (2018b). The Development of Selective Trust: Prospects for a Dual- Process Account. *Child Development Perspectives, 12*(2), 134-138.
- Heyman, G. D. (2014). Children's reasoning about deception: A cross-cultural perspective. In *Trust and Skepticism* (pp. 91-102). Psychology Press.
- Hill, G. W. (1982). Group versus individual performance: Are N+ 1 heads better than one?. *Psychological bulletin, 91*(3), 517.
- Hoffman, M. L. (1970). Conscience, personality, and socialization techniques. *Human development, 13*(2), 90-126.
- Hollenbeck, J. R., Ilgen, D. R., Sego, D. J., Hedlund, J., Major, D. A., & Phillips, J. (1995). Multilevel theory of team decision making: Decision performance in teams incorporating distributed expertise. *Journal of Applied Psychology, 80*(2), 292.
- Horner, V., & Whiten, A. (2005). Causal knowledge and imitation/emulation switching in chimpanzees (*Pan troglodytes*) and children (*Homo sapiens*). *Animal cognition, 8*(3), 164-181.
- Jaswal, V. K. (2010). Believing what you're told: Young children's trust in unexpected testimony about the physical world. *Cognitive psychology, 61*(3), 248-272.
- Jaswal, V. K., Croft, A. C., Setia, A. R., & Cole, C. A. (2010). Young children have a specific, highly robust bias to trust testimony. *Psychological Science, 21*(10), 1541-1547.
- Jaswal, V. K., & Kondrad, R. L. (2016). Why children are not always epistemically vigilant: Cognitive limits and social considerations. *Child Development Perspectives, 10*(4), 240-244.
- Jaswal, V. K., Lima, O. K., & Small, J. E. (2009). Compliance, conversion, and category induction. *Journal of Experimental Child Psychology, 102*(2), 182-195.
- Jaswal, V. K., & Malone, L. S. (2007). Turning believers into skeptics: 3-year-olds' sensitivity to cues to speaker credibility. *Journal of Cognition and Development, 8*(3), 263-283.
- Jaswal, V. K., Pérez- Edgar, K., Kondrad, R. L., Palmquist, C. M., Cole, C. A., & Cole, C. E. (2014). Can't stop believing: Inhibitory control and resistance to misleading testimony. *Developmental Science, 17*(6), 965-976.
- Johnson, B. T., & Eagly, A. H. (1990). Involvement and persuasion: Types, traditions, and the evidence.

- Kahneman, D. (2003). A perspective on judgment and choice: mapping bounded rationality. *American psychologist*, *58*(9), 697.
- Karau, S. J., & Williams, K. D. (1993). Social loafing: A meta-analytic review and theoretical integration. *Journal of personality and social psychology*, *65*(4), 681.
- Kennedy, J., Kleinmuntz, D. N., & Peecher, M. E. (1997). Determinants of the justifiability of performance in ill-structured audit tasks. *Journal of Accounting Research*, *35*, 105-123.
- Kerr, N. L., & Tindale, R. S. (2004). Group performance and decision making. *Annu. Rev. Psychol.*, *55*, 623-655.
- Keysar, B., Lin, S., & Barr, D. J. (2003). Limits on theory of mind use in adults. *Cognition*, *89*(1), 25-41.
- Kim, S., Paulus, M., & Kalish, C. (2017). Young children's reliance on information from inaccurate informants. *Cognitive science*, *41*, 601-621.
- Kinzler, K. D., Corriveau, K. H., & Harris, P. L. (2011). Children's selective trust in native- accented speakers. *Developmental science*, *14*(1), 106-111.
- Kloo, D., & Rohwer, M. (2012). The development of earlier and later forms of metacognitive abilities: Reflections on agency and ignorance. *Foundations of metacognition*, 167-180.
- Koenig, M. A., Clément, F., & Harris, P. L. (2004). Trust in testimony: Children's use of true and false statements. *Psychological Science*, *15*(10), 694-698.
- Koenig, M. A., & Harris, P. L. (2005). Preschoolers mistrust ignorant and inaccurate speakers. *Child development*, *76*(6), 1261-1277.
- Koenig, M. A., & Woodward, A. L. (2010). Sensitivity of 24-month-olds to the prior inaccuracy of the source: possible mechanisms. *Developmental psychology*, *46*(4), 815.
- Köymen, B., Mammen, M., & Tomasello, M. (2016). Preschoolers use common ground in their justificatory reasoning with peers. *Developmental psychology*, *52*(3), 423.
- Köymen, B., Rosenbaum, L., & Tomasello, M. (2014). Reasoning during joint decision-making by preschool peers. *Cognitive Development*, *32*, 74-85.
- Köymen, B., & Tomasello, M. (2018). Children's meta-talk in their collaborative decision making with peers. *Journal of experimental child psychology*, *166*, 549-566.
- Kuczynski, L. (1982). Intensity and orientation of reasoning: Motivational determinants of children's compliance to verbal rationales. *Journal of Experimental Child Psychology*, *34*(3), 357-370.
- Kuhn, D. (2001). How do people know?. *Psychological science*, *12*(1), 1-8.
- Kuhn, D. (2015). Thinking together and alone. *Educational Researcher*, *44*(1), 46-53.

- Kuhn, D., Zillmer, N., Crowell, A., & Zavala, J. (2013). Developing norms of argumentation: Metacognitive, epistemological, and social dimensions of developing argumentative competence. *Cognition and Instruction, 31*(4), 456-496.
- Kunda, Z. (1990). The case for motivated reasoning. *Psychological bulletin, 108*(3), 480.
- Landrum, A. R., Mills, C. M., & Johnston, A. M. (2013). When do children trust the expert? Benevolence information influences children's trust more than expertise. *Developmental Science, 16*(4), 622-638.
- Langer, E. J., Blank, A., & Chanowitz, B. (1978). The mindlessness of ostensibly thoughtful action: The role of "placebic" information in interpersonal interaction. *Journal of personality and social psychology, 36*(6), 635.
- Latané, B., Williams, K., & Harkins, S. (1979). Many hands make light the work: The causes and consequences of social loafing. *Journal of personality and social psychology, 37*(6), 822.
- Larrick, R. P., & Soll, J. B. (2006). Intuitions about combining opinions: Misappreciation of the averaging principle. *Management science, 52*(1), 111-127.
- Laughlin, P. R., Bonner, B. L., & Miner, A. G. (2002). Groups perform better than the best individuals on letters-to-numbers problems. *Organizational Behavior and Human Decision Processes, 88*(2), 605-620.
- Laughlin, P. R., Hatch, E. C., Silver, J. S., & Boh, L. (2006). Groups perform better than the best individuals on letters-to-numbers problems: effects of group size. *Journal of Personality and social Psychology, 90*(4), 644.
- Laughlin, P. R., Zander, M. L., Knievel, E. M., & Tan, T. K. (2003). Groups perform better than the best individuals on letters-to-numbers problems: Informative equations and effective strategies. *Journal of Personality and Social Psychology, 85*(4), 684.
- Leman, P. J. (2002). Argument structure, argument content, and cognitive change in children's peer interaction. *The Journal of genetic psychology, 163*(1), 40-57.
- Levine, J. M., Resnick, L. B., & Higgins, E. T. (1993). Social foundations of cognition. *Annual review of psychology, 44*(1), 585-612.
- Li, Q. G., Heyman, G. D., Xu, F., & Lee, K. (2014). Young children's use of honesty as a basis for selective trust. *Journal of experimental child psychology, 117*, 59-72.
- Lim, J. S., & O'Connor, M. (1995). Judgemental adjustment of initial forecasts: Its effectiveness and biases. *Journal of Behavioral Decision Making, 8*(3), 149-168.
- Lockhart, K. L., Chang, B., & Story, T. (2002). Young children's beliefs about the stability of traits: Protective optimism?. *Child Development, 73*(5), 1408-1430.

- Mahr, J. B., & Csibra, G. (2018). Why do we remember? The communicative function of episodic memory. *Behavioral and brain sciences*, 41.
- Mascaro, O., & Morin, O. (2014). Gullible's travel: How honest and trustful children become vigilant communicators. In *Trust and skepticism: Children's selective learning from testimony* (pp. 69-83). Psychology Press.
- Mascaro, O., & Sperber, D. (2009). The moral, epistemic, and mindreading components of children's vigilance towards deception. *Cognition*, 112(3), 367-380.
- McGuigan, N., Whiten, A., Flynn, E., & Horner, V. (2007). Imitation of causally opaque versus causally transparent tool use by 3-and 5-year-old children. *Cognitive Development*, 22(3), 353-364.
- Meltzoff, A. N., & Moore, M. K. (1977). Imitation of facial and manual gestures by human neonates. *Science*, 198(4312), 75-78.
- Meltzoff, A. N. (1988a). Infant imitation after a 1-week delay: long-term memory for novel acts and multiple stimuli. *Developmental psychology*, 24(4), 470.
- Meltzoff, A. N. (1988b). Infant imitation and memory: Nine-month-olds in immediate and deferred tests. *Child development*, 59(1), 217.
- Meltzoff, A. N., & Moore, M. K. (1983). The origins of imitation in infancy: Paradigm, phenomena, and theories. *Advances in infancy research*.
- Mercier, H. (2017). How gullible are we? A review of the evidence from psychology and social science. *Review of General Psychology*, 21(2), 103.
- Mercier, H. (2011). Reasoning serves argumentation in children. *Cognitive Development*, 26(3), 177-191.
- Mercier, H. (2016). The argumentative theory: Predictions and empirical evidence. *Trends in Cognitive Sciences*, 20(9), 689-700.
- Mercier, H., Bernard, S., & Clément, F. (2014). Early sensitivity to arguments: How preschoolers weight circular arguments. *Journal of Experimental Child Psychology*, 125, 102-109.
- Mercier, H., & Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. *Behavioral and brain sciences*, 34(2), 57-74.
- Moll, H., & Tomasello, M. (2004). 12- and 18- month- old infants follow gaze to spaces behind barriers. *Developmental science*, 7(1), F1-F9.
- Nagell, K., Olguin, R. S., & Tomasello, M. (1993). Processes of social learning in the tool use of chimpanzees (*Pan troglodytes*) and human children (*Homo sapiens*). *Journal of Comparative Psychology*, 107(2), 174.

- Neuman, Y., Glasner, A., & Weinstock, M. (2004). The effect of a reason's truth-value on the judgment of a fallacious argument. *Acta Psychologica*, *116*(2), 173-184.
- Neuman, Y., Weinstock, M. P., & Glasner, A. (2006). The effect of contextual factors on the judgement of informal reasoning fallacies. *The Quarterly Journal of Experimental Psychology*, *59*(2), 411-425.
- Nielsen, M., & Tomaselli, K. (2010). Overimitation in Kalahari Bushman children and the origins of human cultural cognition. *Psychological science*, *21*(5), 729-736.
- Nucci, L. P. (1985). Children's conceptions of morality, societal convention, and religious prescription. *Moral dilemmas: Philosophical and psychological issues in the development of moral reasoning*, 137-174.
- O'Neill, D. K., Astington, J. W., & Flavell, J. H. (1992). Young children's understanding of the role that sensory experiences play in knowledge acquisition. *Child Development*, *63*(2), 474-490.
- Oostenbroek, J., Suddendorf, T., Nielsen, M., Redshaw, J., Kennedy-Costantini, S., Davis, J., & Slaughter, V. (2016). Comprehensive longitudinal study challenges the existence of neonatal imitation in humans. *Current Biology*, *26*(10), 1334-1338.
- Pasquini, E. S., Corriveau, K. H., Koenig, M., & Harris, P. L. (2007). Preschoolers monitor the relative accuracy of informants. *Developmental psychology*, *43*(5), 1216.
- Perlman, M., & Ross, H. S. (2005). If-then contingencies in children's sibling conflicts. *Merrill-Palmer Quarterly (1982-)*, 42-66.
- Price, P. C., & Stone, E. R. (2004). Intuitive evaluation of likelihood judgment producers: Evidence for a confidence heuristic. *Journal of Behavioral Decision Making*, *17*(1), 39-57.
- Raafat, R. M., Chater, N., & Frith, C. (2009). Herding in humans. *Trends in cognitive sciences*, *13*(10), 420-428.
- Rader, C. A., Larrick, R. P., & Soll, J. B. (2017). Advice as a form of social influence: Informational motives and the consequences for accuracy. *Social and Personality Psychology Compass*, *11*(8), e12329.
- Rakison, D. H., & Poulin-Dubois, D. (2001). Developmental origin of the animate–inanimate distinction. *Psychological bulletin*, *127*(2), 209.
- Rakoczy, H. (2010). Executive function and the development of belief–desire psychology. *Developmental Science*, *13*(4), 648-661.
- Rakoczy, H., Ehrling, C., Harris, P. L., & Schultze, T. (2015). Young children heed advice selectively. *Journal of experimental child psychology*, *138*, 71-87.



- Rakoczy, H., Warneken, F., & Tomasello, M. (2008). The sources of normativity: young children's awareness of the normative structure of games. *Developmental psychology*, 44(3), 875.
- Ram, A., & Ross, H. S. (2001). Problem solving, contention, and struggle: How siblings resolve a conflict of interests. *Child Development*, 72(6), 1710-1722.
- Ram, A., & Ross, H. (2008). 'We Got to Figure it Out': Information-sharing and Siblings' Negotiations of Conflicts of Interests. *Social Development*, 17(3), 512-527.
- Reid, T., & Brookes, D. R. (2000). *An inquiry into the human mind on the principles of common sense: a critical edition* (Vol. 2). Penn State Press.
- Rips, L. J. (2002). Circular reasoning. *Cognitive Science*, 26(6), 767-795.
- Robinson, E. J., & Einav, S. (Eds.). (2014). *Trust and skepticism: Children's selective learning from testimony*. Psychology Press.
- Ross, L. (1977). The Intuitive Psychologist And His Shortcomings: Distortions in the Attribution Process<sup>1</sup>. In *Advances in experimental social psychology* (Vol. 10, pp. 173-220). Academic Press.
- Russell, B., & Paperbacks, R. A. R. C. T. (1921). 1989. *The Analysis of Mind*, London: Routledge.
- Sabbagh, M. A., & Baldwin, D. A. (2001). Learning words from knowledgeable versus ignorant speakers: Links between preschoolers' theory of mind and semantic development. *Child development*, 72(4), 1054-1070.
- Savadori, L., Van Swol, L. M., & Sniezek, J. A. (2001). Information sampling and confidence within groups and judge advisor systems. *Communication Research*, 28(6), 737-771.
- Schultze, T., Mojzisch, A., & Schulz-Hardt (2017). On the inability to ignore useless advice: A case for anchoring in the Judge-Advisor-System *Experimental Psychology*. *Experimental Psychology*, 64, 170–183.
- Searle, J. R., & Willis, S. (1995). *The construction of social reality*. Simon and Schuster.
- Shafir, E., Simonson, I., & Tversky, A. (1993). Reason-based choice. *Cognition*, 49(1-2), 11-36.
- Shaw, A., Li, V., & Olson, K. R. (2012). Children apply principles of physical ownership to ideas. *Cognitive Science*, 36(8), 1383-1403.
- Shaw, M. E. (1932). A comparison of individuals and small groups in the rational solution of complex problems. *The American Journal of Psychology*, 44(3), 491-504.
- Shaw, M. E., & Ashton, N. (1976). Do assembly bonus effects occur on disjunctive tasks? A test of Steiner's theory. *Bulletin of the Psychonomic Society*, 8(6), 469-471.

- Siegal, M., Waters, L.J., & Dinwiddy, L.S. (1988). Misleading children: Causal attributions for inconsistency under repeated questioning. *Journal of experimental Child Psychology*, 45(3), 438-456.
- Smith, L. B., Yu, C., & Pereira, A. F. (2011). Not your mother's view: The dynamics of toddler visual experience. *Developmental science*, 14(1), 9-17.
- Sniezek, J. A., & Buckley, T. (1995). Cueing and cognitive conflict in judge-advisor decision making. *Organizational behavior and human decision processes*, 62(2), 159-174.
- Sniezek, J. A., Schrah, G. E., & Dalal, R. S. (2004). Improving judgement with prepaid expert advice. *Journal of Behavioral Decision Making*, 17(3), 173-190.
- Sniezek, J. A., & Van Swol, L. M. (2001). Trust, confidence, and expertise in a judge-advisor system. *Organizational behavior and human decision processes*, 84(2), 288-307.
- Sobel, D. M., & Kushnir, T. (2013). Knowledge matters: How children evaluate the reliability of testimony as a process of rational inference. *Psychological Review*, 120(4), 779.
- Sperber, D., Clément, F., Heintz, C., Mascaro, O., Mercier, H., Origgi, G., & Wilson, D. (2010). Epistemic vigilance. *Mind & Language*, 25(4), 359-393.
- Sperber, D. (2000). Metarepresentations in an evolutionary perspective. *Metarepresentations: A multidisciplinary perspective*, 117-137.
- Sperber, D., & Wilson, D. (1986). *Relevance: Cognition and communication*. Oxford: Blackwell, 1995.
- Sobel, D. M., & Kushnir, T. (2013). Knowledge matters: How children evaluate the reliability of testimony as a process of rational inference. *Psychological Review*, 120(4), 779.
- Soll, J. B., & Mannes, A. E. (2011). Judgmental aggregation strategies depend on whether the self is involved. *International Journal of Forecasting*, 27(1), 81-102.
- Soll, J. B., & Larrick, R. P. (2009). Strategies for revising judgment: How (and how well) people use others' opinions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35(3), 780.
- Song, R., Over, H., & Carpenter, M. (2015). Children draw more affiliative pictures following priming with third-party ostracism. *Developmental Psychology*, 51(6), 831.
- Stanovich, K. E. (2013). Why humans are (sometimes) less rational than other animals: Cognitive complexity and the axioms of rational choice. *Thinking & Reasoning*, 19(1), 1-26.
- Stasser, G., & Dietz-Uhler, B. (2001). Collective choice, judgment, and problem solving. *Blackwell handbook of social psychology: Group processes*, 3, 31-55.

- Stasser, G., & Titus, W. (1985). Pooling of unshared information in group decision making: Biased information sampling during discussion. *Journal of personality and social psychology*, 48(6), 1467.
- Taylor, M., Cartwright, B. S., & Bowden, T. (1991). Perspective taking and theory of mind: Do children predict interpretive diversity as a function of differences in observers' knowledge?. *Child Development*, 62(6), 1334-1351.
- Tomasello, M. (2014). *A natural history of human thinking*. Harvard University Press.
- Tomasello, M. (1995). Joint attention as social cognition. *Joint attention: Its origins and role in development*, 103130.
- Tomasello, M., & Carpenter, M. (2007). Shared intentionality. *Developmental science*, 10(1), 121-125.
- Tomasello, M., Kruger, A. C., & Ratner, H. H. (1993). Cultural learning. *Behavioral and brain sciences*, 16(3), 495-511.
- Tost, L. P., Gino, F., & Larrick, R. P. (2012). Power, competitiveness, and advice taking: Why the powerful don't listen. *Organizational Behavior and Human Decision Processes*, 117(1), 53-65.
- Turner, M. E., & Pratkanis, A. R. (1998). Twenty-five years of groupthink theory and research: Lessons from the evaluation of a theory. *Organizational behavior and human decision processes*, 73(2-3), 105-115.
- Tzioti, S. C., Wierenga, B., & van Osselaer, S. M. (2014). The effect of intuitive advice justification on advice taking. *Journal of Behavioral Decision Making*, 27(1), 66-77.
- Vanderbilt, K. E., Heyman, G. D., & Liu, D. (2014). In the absence of conflicting testimony young children trust inaccurate informants. *Developmental Science*, 17(3), 443-451.
- Vanderbilt, K. E., Liu, D., & Heyman, G. D. (2011). The development of distrust. *Child development*, 82(5), 1372-1380.
- Van Eemeren, F. H., Grootendorst, R. F., & Henkemans, F. S. (1996). *Fundamentals of argumentation theory: A handbook of historical backgrounds and contemporary applications*.
- Van Swol, L. M., & Sniezek, J. A. (2005). Factors affecting the acceptance of expert advice. *British Journal of Social Psychology*, 44(3), 443-461.
- Walker, M. B., & Andrade, M. G. (1996). Conformity in the Asch task as a function of age. *The Journal of social psychology*, 136(3), 367-372.
- Walton, D. (1999). The appeal to ignorance, or argumentum ad ignorantiam. *Argumentation*, 13(4), 367-377.

- Weinstock, M., Neuman, Y., & Tabak, I. (2004). Missing the point or missing the norms? Epistemological norms as predictors of students' ability to identify fallacious arguments. *Contemporary Educational Psychology, 29*(1), 77-94.
- Whitcombe, E. L., & Robinson, E. J. (2000). Children's decisions about what to believe and their ability to report the source of their belief. *Cognitive Development, 15*(3), 329-346.
- Williams, K., Harkins, S. G., & Latané, B. (1981). Identifiability as a deterrent to social loafing: Two cheering experiments. *Journal of Personality and Social Psychology, 40*(2), 303.
- Wilson, N. L. (1959). Substances without substrata. *The Review of Metaphysics, 12*(4), 521-539.
- Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition, 13*(1), 103-128.
- Yaniv, I. (2004). Receiving other people's advice: Influence and benefit. *Organizational Behavior and Human Decision Processes, 93*(1), 1-13.
- Yaniv, I., & Foster, D. P. (1997). Precision and accuracy of judgmental estimation. *Journal of behavioral decision making, 10*(1), 21-32.
- Yaniv, I., & Kleinberger, E. (2000). Advice taking in decision making: Egocentric discounting and reputation formation. *Organizational behavior and human decision processes, 83*(2), 260-281.
- Zelazo, P. D., Frye, D., & Rapus, T. (1996). An age-related dissociation between knowing rules and using them. *Cognitive development, 11*(1), 37-63.
- Zmyj, N., Daum, M. M., Prinz, W., Nielsen, M., & Aschersleben, G. (2012). Fourteen-month-olds' imitation of differently aged models. *Infant and Child Development, 21*(3), 250-266.



## Appendix

### Study 3 - Coding scheme

**Table 1.** Coding scheme of dialogs and requested reasons.

	Category	English (translation)	German
<b>Ratings</b>			
	Eyesight		
		to see	sehen <sup>1</sup>
	State of knowledge		
		to know	wissen <sup>2</sup>
	Correctness of individual judgments <sup>3</sup>		
		that one is right	das stimmt
		I did it right	ich habe es richtig gemacht
		mine is the right one	meins ist richtig
		it belongs there	da gehört es hin
		here was it really	hier war es wirklich
	Level of difficulty		
		this is hard	das ist schwer
		this is mean	das ist fies
<b>Uncertainty</b>			
		maybe	vielleicht
		might/could/should/would	könnte/sollte/müsste/wäre
		possibly	eventuell
		to believe	glauben
		probably	wahrscheinlich
		to assume	vermuten
		no idea	keine Ahnung
		to guess	geraten
		to suppose	annehmen

		actually	eigentlich
		approximately	ungefähr
		my feeling was	vom Gefühl
		presumably	vermutlich
		somehow	irgendwie
<b>(Requested)<sup>4</sup></b>			
<b>Reasons</b>			
	Right		
		to see	sehen <sup>5</sup>
	Ambiguous <sup>6</sup>		
		because I know better	weil ich es besser weiß
		because it belongs there	weil es dahin gehört
		because mine was right	weil meins richtig ist
	Wrong		
		to see	sehen <sup>7</sup>
	Irrelevant		
		because I want it	weil ich das möchte

<sup>1</sup> Positive and negative forms of the verb were coded as well as self-referred and other-referred forms: I see, I see not, you see, you see not.

<sup>2</sup> Positive and negative forms of the verb were coded as well as self-referred and other-referred forms: I know, I do not know, you know, you do not know.

<sup>3</sup> Only expressions with the judgment as subject in the sentence were coded as well as positive and negative forms, and self-referred and other-referred forms: e.g. mine (for my judgment) is right/wrong, mine is not right/wrong, yours (for your judgment) is right/wrong, yours is not right/wrong.

<sup>4</sup> The coding scheme of *Reasons* and *Requested reasons* was identical, however, *Reasons* were presented in the result section *Dialogs* as overall category, summarizing all four single categories.

<sup>5</sup> Reasons including the verb “to see” were coded, only. Positive and negative forms of the verb were coded as well as self-referred and other-referred forms, but only if the given reason was right from the perspective of the judges’ visual access. If the judge, previously sitting behind the clear window, used expressions like “because, I saw the mark better” or “because, she (the judge from the blurred

window) did not see the mark". Accordingly, from the perspective of the judge from the blurred window: "because, I did not see well" or "because, she (the judge from the clear window) saw better".

<sup>6</sup> Reasons that were basically right (meant the right explanation) but did not include "seeing", e.g. did not precisely say why the judge "know better". The missing explanation here was: "because she saw better".

<sup>7</sup> Reasons including the verb "to see" were coded, only. Positive and negative forms of the verb were coded as well as self-referred and other-referred forms, but only if the given reason was wrong from the perspective of the judges' visual access. If the judge, previously sitting behind the clear window, used expressions like "because, I did not see well" or "because, she (the judge from the blurred window) saw better". Accordingly, from the perspective of the judge from the blurred window: "because, I saw better" or "because, she (the judge from the clear window) did see worse".



## Curriculum Vitae

### Positions

- Oct. 2015 –  
Oct. 2018            Scientific assistant, Developmental Psychology, Georg-August-University, Göttingen
- Oct. 2012 –  
Mar. 2014            Scientific assistant, Research group Evolutionary Psychology, Freie University Berlin
- Oct. 2010 –  
Jul. 2012            Scientific assistant, Department of Developmental and Comparative Psychology, Max Planck Institute for Evolutionary Anthropology, Leipzig
- Dec. 2008 –  
May 2012            Student assistant, Department of Developmental and Comparative Psychology, Max Planck Institute for Evolutionary Anthropology, Leipzig

### Education

- Oct. 2015 –  
Mar. 2019            Graduate studies, PhD program Behavior and Cognition, Georg-August-University, Göttingen
- Feb. 2010            Diploma in Biology, Martin-Luther-University Halle-Wittenberg
- Jul. 2000            A levels school degree, Johann-Gottfried-Herder-Gymnasium, Halle (Saale)