Interpersonal Information Integration in Judgment Revision and Collective Judgment Formation

The Benefits of Distributed Access to Redundant and Complementary Visual Information in a Shared Environment

By

Pavel Valeryevich Voinov

Submitted to

Central European University

Department of Cognitive Science

In partial fulfillment of the requirements for the degree of Doctor of Philosophy in Cognitive Science

First Supervisor: Günther Klaus Knoblich

Secondary Supervisor: Natalie Sebanz

Budapest, Hungary

2017

Declaration of Authorship

I hereby declare that this submission is my own work and to the best of my knowledge it contains no materials previously published or written by another person, or which have been accepted for the award of any other degree or diploma at Central European University or any other educational institution, except where due acknowledgment is made in the form of bibliographical reference.

Pavel Voinov

Abstract

One intellectual problem where collaboration can be helpful is coming up with a quantitative judgment under uncertainty. The common consensus among scholars and researchers, though, is that people generally fail to fully realize the advantage of having plural minds. The aim of the present work is to increase our understanding of psychological and social mechanisms that allow interacting individuals to combine their uncertain knowledge into a judgment, and the causes of collective benefit and collective failure in this process.

The present work addresses collective judgment as an "information integration" problem in analogy with the process of multi-sensory integration that takes place within the brain (Ernst & Bülthoff, 2004), following the original approach suggested by Bahrami et al. (2012a). It extends existing research on two lines. First, it addresses the process of inter-individual information integration under conditions implying a different degree of structural overlap in the individually available information. The second novel aspect is the focus on non-verbal modes of interaction via a shared environment.

The thesis includes two empirical studies, each consisting of a series of behavioral experiments, that investigate how environment-mediated interactions can support the process of inter-individual information integration under conditions of individual access to redundant and complementary information. The two studies address two conceptually different processes of inter-individual information integration: individual judgment revision and joint judgment formation.

The first study investigates how indirect interactions via a shared environment can help individuals to improve their perceptual judgments by observing another's judgments. The main finding is that whether people can properly integrate observable information in the environment produced by another individual depends on their uncertainty about their own judgment. Crucially, when their own uncertainty is high, people do not discriminate between information of high and low

quality in another's judgment. This leads to underperformance in the potentially most beneficial conditions – the ones where people have access to complementary information.

The second study investigates how well pairs of participants can coordinate their joint judgment by means of interactions via a shared environment. It addresses the interplay between feedback on accuracy and verbal communication under conditions of simultaneous access to complementary and redundant visual information. Under conditions of access to redundant information, availability of feedback on accuracy turns out to be critical: without it interactions do not lead to an improved judgment. Verbal communication does not seem to play a crucial role, but it is helpful under conditions of access to complementary information. Furthermore, in the latter situation, a reliable collective benefit from interaction can be obtained in the absence of verbal communication, of feedback, or both.

The reported studies have three major implications. First, they suggest that in a situation of collective judgment in a shared environment reliable collective benefits from interaction can be obtained without verbal communication. Second, they point to a critical role of a shared agreement on a judgment in this process. Third, they highlight the significance of the factor of structural overlap in individually held information as an important determinant of the amount of collective benefit that collaborators are likely to obtain from social interaction.

This dissertation is dedicated to my wife, Marina Oshkina, who has supported me through all these years and exerted much more patience than I could have asked for.

Acknowledgments

I would first like to express gratitude to my supervisors, Günther Klaus Knoblich and Natalie Sebanz, for their guidance, involvement, care, and moral support; for their critical scrutiny, and, at the same time, forbearance and infinite patience on this way. Their contribution goes far beyond plain supervision. I especially appreciate the freedom they allowed me in pursuing my research project. I owe all my researcher skills to them.

I would like to thank Krisztián Gábris for helping me with programming, and Veronika Herendy, Anna Mária Lisincki, and Fruzsina Elekes for help with data collection and for literally saving one of the experiments. Without their assistance much of this work would have been not possible to accomplish.

I would like to thank Máté Lengyel for helping me with the formal aspect of the thesis.

I am thankful to CEU for providing me with various financial support on my graduate track, in particular, for providing me with the write-up grant, that allowed me to complete this thesis.

Finally, I would like to thank all lab visitors who provided me with helpful and insightful commentary on my work.

Table of Contents

List of Figures

List of Tables

Introduction

"Two heads are better than one" according to folk wisdom. This statement reflects the intuition that by combining individually available knowledge, a group of people can make a more informed and reliable decision than solitary individuals can. One type of decision problem where integrating privately held information from different members of a group can be helpful is coming up with quantitative judgments. Examples for such judgments include forecasting the sales of products, predicting revenue, estimating the amount of investment required to attain a certain outcome, or choosing the proper movement direction to chase the prey during a hunt.

However, a large body of literature on decision-making in small groups (Kerr & Tindale, 2004; Lorge, Fox, Davitz, and Brennar, 1958) has raised serious doubts about the claim that two heads are better than one. While, based on the statistical argument (Larrick, Mannes, & Soll, 2012), individual judgments can be easily combined to derive a judgment of superior quality, empirical research on small group decision-making involving judgments has demonstrated that interacting groups systematically fail to effectively combine individual opinions into a judgment that would surpass that of an average, not to say, of the most competent group member (Hastie, 1986). Applying an information processing perspective on the problem, the present thesis aims to increase our understanding of the psychological mechanisms at work, and in this way, to identify factors that make groups fail to benefit from distributed knowledge and factors that increase the chances of maximizing the benefit of group judgments.

In the current investigations we follow a general theoretical framework that conceives of groups as information processors (Hinsz, Tindale, & Vollrath, 1997) and we conceptualize the process of collective judgment in terms of inter-individual "information integration". Specifically, we adopt a recent theoretical approach proposed by Bahrami et al. (2010, 2012a), which draws on a conceptual analogy between the problem of collective decision-making and that of multi-sensory

integration of information from different sensory modalities in the human brain. Building on the "random error" statistical model of individual judgment (Yaniv, 2004b), the theory predicts that integrating judgments across individuals will reduce the amount of idiosyncratic error in a similar way as multi-sensory integration filters out noise in sensory inputs (Ernst & Bülthoff, 2004).

While our theory rests on the same fundamental assumptions on the nature of uncertainty in internal representations of the true environmental state as in Bahrami et al. (2010, 2012a), the two theories diverge in respect to how information is integrated between individuals. In particular, the inter-individual information integration model by Bahrami et al. (2010, 2012a) asserts that individuals cannot communicate separately the perceptual content informing one's decision and reliability associated with it, but rather need to communicate their confidence in individual decisions to achieve collective benefit. Based upon Graesser's (1991) Social Averaging theorem of group compromise, we argue that in a situation of collective judgment group members do not need to communicate their confidence to achieve a collective benefit from interaction. Instead, collective benefit from information integration can be achieved if the minimal means for coordinating the joint judgment are available to the group members.

One important implication that follows from Bahrami et al.'s (2010, 2012a) theory of inter-individual information integration is that individuals with very different levels of competence should avoid collaboration, as it has high chances to lead to a collective failure. This effect is a consequence of the fact that confidence is a sub-optimal measure of individual uncertainty about one's knowledge, and relying on it results in individual contributions to the collective decision not exactly proportional to individual competence. It is questionable, though, whether the conclusion reached by Bahrami et al. (2010), namely that individuals with too different abilities should avoid collaboration, generalizes to more complex situations.

One particular limitation of current research on inter-individual integration of perceptual information is that it has been restricted to scenarios where multiple observers judge the same one-dimensional environmental properties or objects. Thus, it is not clear how well the proposed mechanisms generalize to situations where observers have structural differences in individually available perceptual input, as in a situation where two observers have different viewpoints on where an object is located in space.

In such situations that involve information integration across multiple perceptual judgment dimensions, the quality of collective judgments will depend not only on the relative difference in individual members' absolute task-related abilities (Bahrami et al., 2010), but also on the structural overlap of information available to different group members. To illustrate, consider putting up a Christmas tree[1]. Installing the tree perpendicular to the floor plane is a challenging task for a single individual due to a limitation of human vision that leads to low spatial accuracy in locating the position of an object on the depth dimension (van Beers, Sittig, & van der Gon, 1998; Foley, 1980). How much putting up the tree together can help to avoid the disaster of a skewed Christmas tree will depend not only (and as we shall show, not as much) on the absolute acuity of the individuals putting up the tree, but rather on the relation between the viewpoints they are looking from when putting up the tree. Given the limitation of depth perception it would seem that two observers should have the biggest advantage when they have orthogonal viewpoints on the tree because this would allow them to compensate for each other's inaccurate depth perception.

Accordingly, the first aim of the current research was to investigate groups' abilities to integrate individual judgments under conditions implying a different degree of structural overlap in the individually available perceptual information. The primary question we addressed to achieve this aim was whether individuals and groups can track higher-order geometric properties of information

[1] I'd like to thank Chris and Uta Frith for pointing me to this example.

distributed across group members, and whether they exploit these properties in the process of information integration.

The second aspect of the problem of information integration in collective judgment that we investigated in the current work concerns coordination mechanisms for inter-individual information integration. Going back to the Christmas tree example, two individuals could accomplish the task together in multiple different ways. For instance, they may or may not act on the tree simultaneously, and they may or may not be in the room at the same time. This raises the question of the degree of interactivity that is required to obtain benefits from joint judgments. Is adjusting the tilt of the tree in a sequential iterative way where two family members independently try to set the tree straight (e.g. dad each morning, mum each evening) as beneficial as putting it up together? And how would the accuracy of two individuals putting up the tree together compare to the accuracy of one individual walking around the tree? Do cooperating individuals need to talk to each other to make the tree stand straight?

The modes of interaction exemplified above involve environment-mediated interactions, which have received significantly less attention in research on collective judgment and decision-making than decisions involving explicit verbal communication about symbolic information. Accordingly, the second aim of the present thesis was to investigate how environment-mediated interactions can support the process of inter-individual information integration. In order to achieve this aim, we investigated whether individuals and groups can improve judgment quality by means of environment-mediated interactions and whether immediate feedback on judgment accuracy is a necessary condition for successful integration of information across people. We also asked whether verbal communication can add precision to joint judgments when two people interact in a shared environment.

To address these research questions two empirical studies were conducted, each consisting of a series of experiments. Each study addressed one mechanism for information integration via environment-mediated interactions. The first study investigated how indirect interactions can help individuals to improve their perceptual judgments by simply observing another's judgments. The second study introduced a consensus constraint (i.e., requiring two individuals to agree) to investigate how well dyads can integrate visual information by means of behaviorally coordinating joint judgments in a shared environment. We addressed the research questions regarding the aspect of structural overlap in individually available information in both studies. To these aims, in both studies we investigated the process of inter-individual information integration under conditions where two individuals from a dyad had access to either redundant or complementary visual information.

The thesis is organized as follows. The first chapter places the current work in the context of previous research that used perceptual judgment and decision-making in a social setting as an experimental model for group processes, and provides a general theoretic background for our research questions. Accordingly, the first section of the first chapter introduces the framework of groups as information processors (Hinsz et al., 1997) and formulates the problem of achieving joint decisions and judgments in terms of information integration (Bahrami et al., 2012a). The second section of the first chapter gives an overview on how previous experiments on collective perceptual judgment and decision-making helped to better understand psychological mechanisms at work. In the third section I extend existing theory to a situation of judgments on multiple dimensions that have an inherent structural relationship and introduce the notions of redundancy and complementarity to operationalize the amount of structural overlap in individually held information. By taking a formal approach to complementarity, I introduce the theorem of Maximum Collective Benefit, which specifies the optimal conditions for collaboration of two individuals making

judgments on two dimensions. In the fourth section of the first chapter I review two mechanisms of environment-mediated interactions: *indirect interactions* and *behavioral implicit communication*, and review how these mechanisms function to coordinate information integration within a group. In the fifth section I review the two-stage model of group judgment (Sniezek & Henry, 1990). The model makes a distinction between two qualitatively different processes taking place at the two stages: the process of judgment revision by individuals in the course of inter-individual interactions, and the process of joint judgment formation. Each of the two empirical studies in the current thesis addresses one of the two stages.

The second chapter reports the first empirical study consisting of three experiments that investigated how indirect interactions can support the process of inter-individual information integration at the stage of judgment revision. In this study we introduced a novel perceptual judgment task. In this task participants were localizing a target on a plain in a virtual environment either individually or in an interactive mode. By varying individual viewpoints on the layout we manipulated the amount of structural overlap in information available to each of two individuals providing judgments in a shared environment. The first experiment tested whether individuals took into account the amount of structural overlap in the information distributed between them when integrating observed judgments into their own judgments. The second experiment investigated the process of judgment revision under conditions of complementary information distribution where two participants had orthogonal viewpoints on a shared environment. It also tested whether people can integrate their own judgments containing complementary information better than judgments available from another individual. The third experiment investigated the process of judgment revision under conditions of redundant information distribution where two participants shared the same viewpoint on a shared environment.

The third chapter reports the second empirical study consisting of four experiments that investigated how environment-mediated interactions can support the process of inter-individual information integration at the stage of joint judgment formation. In this study individuals grouped into dyads provided joint judgments of a target location. We investigated the interplay between feedback on accuracy and verbal communication under conditions of complementary and redundant information distribution. Experiments 1 and 4 investigated the process of inter-individual information integration under conditions of complementary information distribution where two participants had orthogonal viewpoints on a shared environment, while Experiments 2 and 3 investigated the process of judgment formation under conditions of redundant information distribution where two participants had the same viewpoint on a shared environment. In Experiments 1 and 2 participants were provided with immediate feedback about the accuracy of their individual and joint judgments, while in Experiments 3 and 4 participants were not provided with feedback. The availability of verbal communication was manipulated in all four experiments. The results of the second study are discussed in the light of previous studies on collective perceptual decision-making conceptualized in terms of information integration (Bahrami et al., 2012a). In the concluding section implications of the present findings for research on collective judgment and decision-making are discussed.

1. Theoretical Background

1.1 Groups as Information Processors

One way to investigate the processes that enable individuals to provide joint judgments is to adopt the framework of groups as information processors (Brauner & Scholl, 2000; Hinsz et al., 1997). According to this framework when a group of individuals is linked by a joint intellectual task, essentially the group goes through the same processing steps as individuals: it acquires relevant information from the environment, processes it, and combines it with further available information to produce the collective product – a joint response. In Hinsz et al.'s (1997) cognitive architecture which is intended to be equally applicable to both individuals and groups, these steps would correspond to interrelated components of information processing: processing objective, attention, processing work space, memory, and response. Adopting such a framework, of course, does not imply assuming structural or algorithmic identity between the group and individual level. Instead, treating groups as information processors is a theoretical method to address the processes taking place within a group when it performs an intellectual task, using individual information processing as a model. In this sense, the different components postulated at the group level, such as "group perception", or "group attention", are cognitive metaphors which help us to better understand group phenomena in terms of information processing.

The advantage of using a common model for the individual and group level of information processing is that we can use individual behavior in a task as a baseline and compare it to group behavior. In this way, we can investigate "if, when, and how group information processing is similar to, or different from, its individual-level counterpart" (Hinsz et al., 1997, p. 44). Furthermore, the use of a common model allows us to develop a theory on how individual and group levels interact with each other using a conceptually coherent framework in the same sense as investigating the same

psychological phenomenon (i.e., perception) on the neural and behavioral level within the same framework allows us to gain a more complete understanding of that phenomenon.

In the Hinsz et al. (1997) framework the problem of integrating individual judgments that is the focus of the current investigation maps onto two components: information processing and forming a joint response. Thus, to study this phenomenon with rigorous experimental methods, we need to isolate as much as possible these processing components from other components such as memory, processing objective, or attention. In the desired scenario, we should be able to fully control the information the participants in our experiment are processing, their objectives, and their response mode. Separating the components of information integration and forming a joint response can be problematic as research on group decision-making illustrates. For instance, in studies using tasks involving almanac knowledge or problem-solving (Hastie, 1986) the collective product has often been confounded with participants' background knowledge and/or motivational aspects of group performance such as within-group competition or conformity. One prominent method, which has recently been re-discovered to study inter-individual information integration, and which supposedly deals with some of the aforementioned issues, is to use perceptual decision-making tasks.

An additional aspect of perceptual tasks which makes them particularly relevant in the context of this work, is that perception is tightly bound to the environment. In non-human animal species collective decisions are almost exclusively made in relation to perceived environmental states, and are implemented with physical actions performed in and in relation to the environment. In this sense the environment is an inherent part of the decision-making process; it is the primary source of informational inputs and possibly the interface for information exchange (see section 1.4). While modern culturally and technologically developed human civilization requires mostly difficult decisions on issues which are beyond immediate perception, it is conceivable that not such a long

time ago, at the beginning of the *Homo Sapiens* era, most informational inputs for decisions made by groups of our ancestors were perceivable environmental properties. This therefore makes perceptual decision-making tasks a particularly suitable method to investigate the role of the environment in group decision-making, which is one of the main goals of this work.

In terms of information processing the problem of inter-individual integration of perceptual information is analogous to the task of multi-sensory integration faced by individual brains: a process by which individuals combine information from several sensory modalities, i.e. vision and touch, or from several cues, to make a more robust estimate of an object property. Research on multi-sensory integration has developed into a distinct research topic (Calvert, Spence, and Stein, 2004; Ernst & Bülthoff, 2004; Spence & Driver, 2004; Trommershäuser, Körding & Landy, 2011) and by now there is ample experimental evidence on how individual brains solve this problem, as well as elaborate computational models of this process.

In line with Hinsz et al.'s (1997) framework, a considerable number of studies on inter-individual information integration adopted methodological aspects of studies in psychophysics and applied computational models originally developed for individual processing to the group level. The original idea that collective decision-making can be recast as an "information integration" problem similar to multi-sensory integration can be found in Bahrami et al. (2012a, p. 1351). Following the first study which made the direct theoretical analogy between the two processes (Bahrami et al., 2010) this recent line of research has already brought new insights to our understanding of how groups make better decisions or judgments and suggested plausible explanations for collective failure in other circumstances. Placing the empirical studies reported in this manuscript in the context of group psychophysics, the next section reviews the history of how models that were developed for within-individual integration of perceptual information provided new insights for our understanding of inter-individual information integration. Specifically, it reviews a history of how elaborating these

models to the group level provided a better understanding on why certain properties of information processing on the individual level under specific conditions lead to recurrent failures on the group level (Bahrami et al., 2012a).

1.2 Collective Perceptual Judgment and Decision-making

Asking groups of participants to make judgments of perceived object properties was a popular method to study the psychology of group judgment and decision-making in the first half of the 20th century (Lorge et al., 1958). In one famous experiment by Knight (1921) college students estimated the temperature in the classroom. When averaged, the group estimate was 80% more accurate than individual estimates. The experiment was very influential at the time because it was seen as support for the value of group processes in judgment formation (Preston, 1938). Knight's findings were replicated in similar studies that involved estimation of weights (Bruce, 1935; Gordon, 1923), the numerosity of buckshot (Bruce, 1935), and the number of item tokens in a bottle (Klugman, 1945). While these studies demonstrated that the "group" judgment was generally more accurate than individual ones, the judgments did not involve an actual group with interacting group members. Rather, the "statisticized" group judgment was obtained by an experimenter integrating individual judgments. Therefore, the statisticized group judgments provided empirical evidence that a group of individuals possesses latent knowledge, which if integrated properly is generally more accurate than individual estimations. However, the studies were criticized for conveying no information on the psychology of group judgment and the behavioral group process itself (Dashiell, 1935; Preston, 1938; Smith, 1941).

Later studies involving groups with interacting members provided evidence for the collective benefit of group judgment integration that countered these criticisms. In an experiment by Schonbar (1945), reported in Lorge et al. (1958), pairs of individuals estimating the length of a line showed

improvements in accuracy compared to individuals. In Arthur Jeness' experiment (1932) student participants individually estimated the number of beans in a jar, then discussed their judgments in ad hoc groups and formed group judgments. The results of his experiment are somewhat controversial, and the author concluded that discussion does not make estimations more accurate. However, in certain conditions, namely when ad hoc groups were composed of individuals with homogeneous judgments, accuracy of group judgments improved compared to the initial individual judgments.

In the pre-cognitive era in group psychology two computational solutions were commonly used for modeling judgment integration within a group and provided benchmarks for group accuracy: averaging and selecting the most accurate judgment (Einhorn, Hogarth, & Klemptner, 1977; Ferrell, 1985). This simplification of the process reflects a view of individual judgment as a random sample of judgments from the population (Einhorn, Hogarth, & Klemptner, 1977). To illustrate, consider a situation where two judges, Allen and Beatrice, are confronted with a task of estimating the height of the Empire State Building. Assuming that the population of all judges can be represented as a random distribution with mean height estimate $\overline{\mathrm{H}}_0$ and standard deviation σ, then A's individual height judgment J_{A} will be a random sample from that distribution with absolute deviation $D = |J_{\mathrm{A}} - H_{\mathrm{T}}|$, where H_{T} is the true height of the tower, and the same holds for B. The two models predict that the judges can average their judgments or select the judgment of the more accurate judge.

One crucial aspect of both models is that they do not include the accuracy of individual judges as parameters. Instead, judges in a group are hypothesized to assess the accuracy (the error) of the individual *judgments* but the probability of each individual judge being correct is assumed to be the same. Several authors (Einhorn, Hogarth, & Klemptner, 1977; Sniezek & Henry, 1989) considered variable weighing schemes for making joint judgments where individual judgments could have non-equal relative weights. However, such formulations of the problem in terms of weighing

accuracy of individual judgments (not judges!) can tell little more about an optimal weighing policy, than saying that "group judgments will be more accurate than equal weighting if weights are proportionate to accuracy" (Sniezek & Henry, 1989, p. 9).

The non-individuating approach in early studies of group decision making reflects a general focus of small group research on outcome generality and not on process generality (Graesser, 1991, Ed.'s Note 2). Similarly, in his formulation of social decision schemes – a widely accepted framework for research on decision-making in small groups (Hinsz, 1999; Stasser, 1999), - Davis (Davis, 1973) explicitly disallowed for individual differences in analysis, emphasizing a probabilistic approach to group decision processes (see also Graesser, 1991, Ed.'s Note 1). Such methodological constraints limited the progress in understanding the dynamics of group interaction and psychological processes through which individual information is evaluated and integrated into the group response.

An alternative approach based on the information integration theory (Anderson, 1981) was proposed by Cheryl Graesser (1991). Both Graesser and Anderson criticized the social decision scheme approach for the lack of an informational perspective (Anderson & Graesser, 1976; Graesser, 1991) and proposed an alternative process that attempted to model group decision making as compromise seeking contingent on the strength of individual group members' preferences. Even though Graesser's study did not address perceptual decision making, her work is important for the present purpose, because its theoretical aspects and implications are highly relevant for the studies on inter-individual integration of perceptual judgments and were partially accommodated into one the models for group judgment formation tested in Study II.

In her study, Graesser (1991) examined how pairs of individuals integrated conflicting preferences into a compromise solution. Following the principles of information integration theory (Anderson, 1981, 1991), her social averaging theorem predicted that the group response should be a

linear sum of weighted individual positions on the issue. In her model of group compromise Graesser allowed for variation in the relative weights of individual positions, and found that participants provided with a steeper preference function had higher influence on the group decision[2]. The elegant and concise Social Averaging theorem (Graesser, 1991) provided an argument from an information integration theory perspective for the hypothesis that individuals with different expertise are not generally expected to have equal influence on the group decision.

A cognitive turn in research on judgment and decision making in small groups took place starting in the 1980s. New cognitive models of group information processing were developed which viewed groups as a composite comprised of individual decision makers who provide their best answer in a situation of uncertainty. The theoretical and methodological novelty of this approach was to apply signal detection analysis – an analytical tool originally developed in the field of electronical engineering (Sorkin, Luan, & Itzkowitz, 2004), and then adopted and extensively used by cognitive psychologists to model individual perception as a process of statistical decision-making taking place under perceptual uncertainty (Ma, 2010; Maloney, 2002). It is worth discussing the basic principles of the signal detection model of the observer (Sorkin, Luan, & Itzkowitz, 2004) to clarify how the studies on within- and between-individual perceptual information integration using signal detection theory are relevant to the problem of perceptual judgment integration, and why we can compare results and implications across the two lines of research.

In the individual signal detection formulation of the perceptual discrimination problem the decision maker observes a perceptual input X, and in the simplest case estimates one of its continuous properties to compare it against a certain criterion c. The goal is to decide whether the

[2] Although the social averaging theorem has never been applied to the problem of integrating different sources of information to form an integrated judgment, the extension of Graesser's model is straightforward: if one substitutes the individual preference function in her model with a function proportional to the subjective likelihood function, the weight of individual contributions to an integrated judgment will be scaled by the individual judge's precision, making group judgments more reliable, which is the main purpose of sensory integration (Ernst & Bülthoff, 2004; Jacobs, 2002).

perceptual input was generated by a "signal" or "no signal" event. The perceptual input is noisy where the noise is stochastic in nature. Therefore, the true relevant state of the environment is not known but estimated. The amount of noise in the perceptual input is directly related to the amount of uncertainty the decision-maker has about the event which generated the input. The smaller the amount of noise in the perceptual event X, the more reliable is the perceptual input and, consequently, the higher is the expected proportion of correct classification decisions the decision-maker can make.

From this model one can see that the estimation problem is an inherent component of the signal detection task: while the response itself is categorical, the relevant perceptual property to be judged is continuous in nature. In judgment tasks the decision stage is omitted, the estimate of the perceptual input property is the final goal of the task, and the reported estimate is the judgment – a response on a continuous scale. If we return to our judges A and B, suppose that their new task is to decide whether a tower they see can be classified as a skyscraper. For that, they visually inspect the height of i different towers, estimating their respective height $\widehat{H}_{\mathrm{i}}$ and compare them to the standard criterion, for example, 100 m (Emporis Standards, 2009): if the estimated $\widehat{H}_{\mathrm{i}}$ is higher than 100 m, they classify it as a skyscraper. One can see that essentially A and B are doing the same information processing at the stage of height estimation which they did when they were judging the height of the Empire State Building.

From the signal detection model of perception one can derive the prediction that combining signals from multiple sensory modalities within one brain (Ernst & Bülthoff, 2004) or from multiple observers across brains (Bahrami et al., 2010) can reduce noise present in the sensory inputs. Crucially, reduced noise makes perception more reliable: the basic process underlying this effect essentially is cancelling out random components of the perceptual event. As a corollary, estimates of the relevant environmental properties can be made more accurately, and the observer's sensitivity to

small differences in the perceptual event increases. In other words, the same process which allows multiple observers to become more sensitive in a signal-from-noise discrimination task should increase the precision of group judgment.

The problem of multi-sensory integration has been tackled with tasks involving continuous judgments (Beers van et al., 1996, 1998, 1999; Godfroy-Cooper, Sandor, Miller, & Welch, 2015); however, to my best knowledge models for optimal integration have not been applied yet to inter-individual cases with continuous judgments. It is important to point out that applying signal detection to continuous judgments does not imply that group processing and performance are the same for binary decisions and continuous judgements. In fact, there may be systematic differences and this issue will be discussed separately and in greater detail in the discussion of Study II.

Though signal detection theory is sufficiently general to accommodate multiple observers (Green & Swets, 1966), it took until the 1980s until it was formally applied to model group decision-making (Batchelder & Romney, 1986; Metz & Shen, 1992; Pete, Pattipati, & Kleinman, 1993a). The early models for group signal detection (Pete, Pattipati, & Kleinman, 1993a; Pete, Pattipati, & Kleinman, 1993b) were focused on the problems of optimal decision rules. Coming back to our example, a possible question is whether A and B classify Tower i as a skyscraper only when both A and B individually classify it is a skyscraper or whether it is sufficient for either A or B individually to classify it as a skyscraper. These models specified optimal group behavioral decision policy (*decision combinations* in the terminology by Green and Swets, 1966) but did not consider the actual process of information integration when observers exchange the perceptual content they acquired (*observation integration*, also in Green and Swets, 1966).

The first formulation of an optimal group information integration model that goes beyond decision combination rules was provided by Sorkin and Dai (1994). Their model provides a multi-individual extension of a model, originally developed for single-observer integration of multi-

channel acoustic signals (Durlach, Braida, & Ito, 1986). Accordingly, the model specified that optimal weights assigned to individual decisions should be proportional to individual indexes of detectability d' (Sorkin, Hays, & West, 2001, Eq. 6), a statistic characterizing individual performance in the signal-from-noise discrimination task. Sorkin and colleagues (2001) tested the model empirically using a signal-detection task originally developed to study individual discrimination ability with multiple elements in the stimulus (Elvers & Sorkin, 1989). Groups composed of two to seven individuals were exposed to a visual display consisting of nine analogue gauges bearing marks at different gauge intervals. The marks on each gauge were independently sampled from either "signal-plus-noise" or "noise-alone" generative distribution dependent on the nature of the trial. Participants were exposed to displays with individual sets of the stimuli sampled from the same distribution and freely interacted without constraints. Then they were prompted to provide a group decision on whether the stimuli on a given trial were drawn from the "signal-plus-noise" or "noise-alone" distribution. To summarize the findings, groups had higher detection rates compared to individuals; however, the gain of increasing the group size dropped rapidly, and the group performance was substantially below the normative model predictions. The data analysis revealed that participants were weighing their individual contributions quite efficiently. However, at the same time the individual effort invested decreased in the joint condition. The authors attributed this finding to the well-known social loafing effect (Latane, Williams, & Harkins, 1979).

There was one strong assumption in the model by Sorkin and Dai (1994): the individual noise in the participants' perceptual input was assumed to be of equal magnitude across all group members. This is a strict assumption that puts into question the generality of the authors' conclusions regarding the efficiency of adequately weighing relevant information in a group. A decade later, Bahrami et al. (2010) relaxed this assumption by adopting a model originally developed for within-individual multi-sensory integration (Ernst & Banks, 2002). According to the statistical

argument, two (or more) judges or decision-makers should weigh their private estimations proportional to their reliability. To integrate information optimally, individuals would need to exchange not only the perceptual estimates of the relevant input property, but also the variance associated with their estimations. Bahrami and colleagues (2010) hypothesized that individuals are not able to communicate variance inter-individually. Instead they communicate the confidence associated with their estimates, which implies scaling perceptual observation by the standard deviation of random component in the observations. This alternation in the weighing policy has certain negative implications for group performance. Bahrami et al. (2010) tested their model which they coined *weighted confidence sharing* with a group perceptual signal-detection task. This study is reviewed in greater detail in section 3.1.3. It demonstrated the importance of meta-cognitive abilities in group decision-making, and pointed to a potential source of recurrently reported failure of groups to properly integrate information available to them (Bahrami et al., 2012a).

After the publication of the first study using the group perceptual signal-detection paradigm (Bahrami et al., 2010), this technique was further used in follow-up studies on inter-individual integration of perceptual information. Mahmoodi et al. (2015) used the same paradigm to investigate how individuals would make decisions for a group by combining their own decision and their partner's decision accompanied by her respective confidence level. They found that individuals exhibit a strong equality bias and weigh two decisions equally regardless of inter-individual discrepancy in accuracy.

In the next section I review certain limitations of the studies conducted so far in the context of multi-dimensional judgments, and show how the new aspect of complementarity comes into the scene in this new context.

1.3 The Role of Complementarity in Multi-dimensional Information Integration

In contrast to classic one-dimensional perceptual information integration problems, which have been addressed in Section 1.2, the problem of integrating multi-dimensional information across different individuals poses new research questions and requires specifying the conditions under which the collective benefit is maximized. Getting back to our judges Allen and Beatrice, suppose that now they are estimating not only the height of a tower, but also its width and depth. Now, A's judgment can be described with a vector, the entries of which are scalar estimates on the respective dimension: $\mathbf{j}_{\mathrm{A}} = \{\hat{h}_{\mathrm{A}}, \hat{w}_{\mathrm{A}}, \hat{d}_{\mathrm{A}}\}$, and similarly for B. A and B can have different precision on each judgment dimension. In this situation the crucial question is: What are the circumstances in terms of relative accuracies when A and B are expected to benefit most from integrating information between them?

The important role of differences in group members' competences on collective benefits from group interaction was highlighted in reviews of studies on group problem-solving and decision-making (Hill, 1982; Kerr & Tindale, 2004). The problem of complementarity received explicit theoretical consideration in the context of inter-personal dialogue (Fusarolli, Raczaszek-Leonardi, & Tylen, 2014). However, there are very few empirical studies on joint judgment and decision-making addressing the amount of structural overlap in individually available information. The neglect of this aspect of collective judgment making was pointed out by Hastie (1986) on the basis of his extensive review (and perhaps, most comprehensive up to date) of relevant empirical studies: "In the empirical literature the most glaring omission we found was the lack of research on group accuracy under conditions in which nonredundant sources of information are pooled" (p. 157). Ibid, he expressed his personal bewilderment about this neglect:

> The paucity of research on nonredundant information pooling and accuracy is ironic in that probably the most common justifications for the use of group rather than individual

> judgment procedures in natural settings are references to the advantages that accrue when members can share nonredundant information. (Hastie, 1986, p. 152)

One exception are the studies by Budescu and colleagues (Budescu & Rantilla, 2000; Budescu, Rantilla, Yu, & Karelitz, 2003) that investigated how judges making forecasts integrated information from other judges who either shared a high proportion of the same sources of information (redundant condition) or only a low proportion of the same information sources (complementary condition). Judges were more confident in the revised forecasts when they received inputs from advisors who shared a higher proportion of the same sources. Being focused on the decision-maker's confidence in integrated forecasts, their model, however, can tell us little about the relationship between the structure of information distribution across individuals and the expected benefit from integration of multidimensional judgments.

At the same time, redundancy and complementarity are key concepts in the field of data fusion which deals with computational approaches to combining information from different sources (Riccia, Lenz, & Kruse, 2001). Typically, information sources are referred to as redundant when they convey information on the same characteristics or object properties, and thus present information from the same point of view (Durrant-Whyte, 1988). Complementary sources carry information on different properties or aspects of the environment. An example is video cameras with overlapping fields of view. Redundancy of information sources is believed to increase confidence, while complementarity is supposed to increase overall information content. While with multiple human observers the first part of this assertion has been experimentally tested by Budescu et al. (2003), the second part has not yet been addressed in the context of group judgment.

The problem of integrating information from non-redundant sources has received consideration in the context of multi-sensory integration (Ernst & Bülthoff, 2004). As a general rule, the less redundancy there is between signals from different sensory channels, the higher the advantage of using multiple channels for estimating properties of the environment. This

fundamental property of sensory systems can explain why we have evolved sensory channels to complement each other and why there is no duplicating of excessively redundant information within modalities (Stein & Meredith, 1993). It can also explain when and why particular sensory modalities dominate in forming percepts of environmental properties. The most important instance of such a dominance is "visual capture" (Ernst & Bülthoff, 2004), a tendency of visual input to dominate the integrated perceptual representation of an object in a situation of conflict between visual and tactile input. The "modality appropriateness" hypothesis (Welch & Warren, 1986) explains visual capture by the assumption that cross-modal perceptual conflicts are always resolved in favour of the more precise modality. Ernst & Bülthoff (2004) emphasize that the dominance of the visual modality reflects a higher reliability with which the relevant perceptual property is estimated. The fact that the human sensory system exhibits varying dominance relationships between sensory modalities under varying circumstances, can be taken as evidence that different sensory channels are tuned to complementary rather than redundant information.

Several studies have highlighted the complementary interplay of different modalities in multi-sensory integration. One example are the studies by Newell and colleagues (Newell, Ernst, Tjan, & Bülthoff, 2001), who demonstrated that for object recognition vision and touch complement each other with the effect of increasing information content (Ernst & Bülthoff, 2004). They found that accuracy of object recognition differed for the two modalities depending on the object orientation with respect to the observer. Visual recognition of objects was most reliable from the side corresponding to the learned view on the object, which is usually the front surface. In contrast, tactile recognition was most reliable when objects were held from the back side. For graspable objects this side is subject to more extensive tactile exploration (with fingers) during natural exploration behavior and when learning about the surface properties of objects. When both modalities are available, they complement each other and increase information gathered about the

object shape; and the cross-modal integration depends on the spatial correspondence of object surface information across modalities (Newell, Bülthoff, & Ernst, 2003).

Focusing on the complementarity of different sensory channels helped to better explain multimodal processing such as the visual-haptic integration of two-dimensional spatial information. Van Beers and colleagues (1996) asked how individuals would combine perceptions from proprioceptive and visual channels to make an estimate of a stimulus location on a plain. The resulting variability of the data was much lower than was predicted by their model which was supposed to be optimal. However, their model did not include the geometric structure of the uncertainties about the location. It was then discovered (Beers van et al., 1998) that reliability of estimations was anisotropic (i.e., direction-dependent) for proprioceptive and visual modalities: localizations from vision alone were less precise on the depth dimension and more precise on the horizontal dimension, whereas localizations from proprioception were more precise on the depth dimension and less precise on the horizontal dimension. A new model that added the geometric properties of uncertainty in the two sensory modalities explained participants' localization performance when the two modalities were integrated (Beers van et al., 1999).

By taking a formal approach to complementarity one can define a general principle for when there should be maximum collective benefit in providing integrated joint judgments when two individuals have perceptual access of varying reliability to information on two different dimensions. This theorem of Maximum Collective Benefit can be stated as follows:

In a task where the correct answer can be described with a two-dimensional vector, and individuals have anisotropic uncertainty about the true quantity, other things being equal the most beneficial condition for integration of individual judgments is the condition of complementarity, i.e. when the dimension of high uncertainty for one individual is the dimension of low uncertainty for another individual and vice versa.

Where collective benefit refers to increase in overall accuracy of judgments, and the error on both judgment dimensions is assumed to equally matter.

This should hold true if individuals are unbiased and have symmetric uncertainty about the true quantity. While the statement itself is very intuitive and could be taken at face value, providing a formal proof turned out to be rather tricky. In Appendix A the formal proof of the theorem is provided based on certain assumptions from statistical estimation theory (Lehman & Casella, 1998).

In a series of empirical studies reported in this manuscript, specific predictions concerning the outcome of inter-individual interactions with respect to the quality of post-interaction judgments following from the theorem are tested.

1.4 The Environment as a Medium for Information Integration

The previous section specified the condition of information distribution under which a group composed of two individuals, other things being equal, is expected to benefit most from integrating individual information in making collective judgments of a two-dimensional vector quantity. However, the fact that information is distributed across group members in a favorable way does not automatically guarantee that it will be integrated resulting in collective benefit. Whether the latter will take place depends on proper processing of available information. It is commonly held that explicit communication (communication through directed messages) provides the most powerful tool to mediate this processing across individuals, and it is even questionable whether information can be effectively integrated without communication (Bahrami et al., 2010, 2012a). This section provides a theory on how a shared environment can support the process of information exchange across individual agents and puts forward two candidate mechanisms which make possible information integration without explicit communication.

Within the groups as information processors framework (see section 1.1 of this manuscript) the processing part of the information integration problem can be described on two levels: individual and group level. On the individual level it refers to information processing within the minds of individual group members; on the group level it refers to information processing of a group of individuals sharing information (Ickes & Gonzalez, 1994). According to Hinsz et al. (1997), information processing at the group level involves "the degree to which information, ideas, or cognitive processes *are shared* [emphasis added], and *are being shared* [emphasis added], among the group members and how this sharing of information affects individual- and group-level outcomes" (p. 43). The aspect of what and how information is *being shared* corresponds to group interaction – the means by which individually held information is being exchanged and transferred. This exchange is the crucial aspect of group information processing: without it a group is nothing more but a sum of individual members. When we think of information exchange between people, we naturally think of verbal, or more generally, intentional communication where a speaker intends to communicate information to a receiver (Sperber & Wilson, 2001). Indeed, verbal communication plays a huge role in group judgment and decision-making, and it has been extensively studied in this context (Hirokawa & Poole, 1986; Hirokawa, 1990; Innami, 1994).

But is communication the only way information can be transmitted between individuals? Consider the following scenario. You arrive at a desert island and you look for a location to pitch your camp. While inspecting the island you find traces of a previous camp. Would you take this location into consideration when deciding to pitch your camp? Would you perhaps think that the previous traveler had a good reason to pitch his camp exactly at this location? Unless you completely ignore this spotted location in your final decision on where to encamp, then your decision is *influenced* by it. The chain of events involved constitutes a transmission of information – from the previous traveler, who made his private deliberate (not random) decision on where to pitch his

camp, to you. And, maybe, after you leave, the first traveler returns, discovers traces of your camp, and decides to pitch his new camp closer to the location you chose. Now, this process constitutes a reciprocal exchange and integration of information!

Although there was no explicit communication in the example above, there was an exchange of information, and this exchange was mediated by the environment. One may consider these two modes of information exchange not mutually exclusive or even tautological if one equates communication with an act of information transmission. Indeed, in a very broad sense communication is always mediated by the environment unless computations can occur in a Platonian space without physical dimensions. Helpful guidance to address this issue, which captures some common-sense intuitions, was provided by Franklin (1996):

> True, all communications occur when one agent acts on the environment and the other senses the results of that action. But would all such acting and sensing comprise communications? I think not. It's far fetched to call my following the tracks of a bobcat in the snow communication between the bobcat and me. Communication, in the sense of the word I intend, requires the sending and receiving of signals. Is the bobcat signaling when he leaves tracks in the snow? I think not. This situation is quite analogous to me following a river downstream while canoeing. (para. 5)

After a review of some previous work on the role of the environment as a medium for information integration I will provide an explicit definition of communication as a directed transmission of information, and the definition of indirect interaction which would encompass cases similar to the camping example above but exclude Franklin's bobcat/river cases.

The first mechanism for environment-mediated information integration, which corresponds to the mode of information exchange illustrated by the camping example, is *indirect interactions* which is closely related to the notion of *stigmergy* (Bonabeau, Dorigo, & Theraulaz, 1999; Kennedy, Eberhart, & Shi, 2001; Lewis & Marsh, 2016; Theraulaz & Bonabeau, 1999). The term "stigmergy" is composed of two roots from the Greek: *stigma* (sign) and *ergon* (work) (Theraulaz & Bonabeau, 1999, p. 102); which literally means work that stimulates further work. The term was coined by the

French entomologist Grassé (1959) to explain how termites from the genus *Bellicositermes* coordinate their actions in the process of their nest reconstruction (Theraulaz & Bonabeau, 1999).

While there are many definitions of stigmergy (Dipple, Raymond, & Docherty, 2014), there is one common principle lying at the core of all definitions: the actions of one agent being influenced by the effects of prior actions of other agents (Franklin, 1996). In the context of collective perceptual decision-making, one classic example from nature is trail laying to food sources utilized by many ant species (Beckers, Deneubourg, Goss, & Pasteels, 1990; Beckers, Deneubourg, & Goss, 1992; Bonabeau, Dorigo, & Theraulaz, 1999; Deneubourg & Goss, 1989; Detrain & Deneubourg, 2009; Hölldobler & Wilson, 1990; Traniello & Robson, 1995). The trail laying mechanism can be described as follows.

Initially all ants explore the nest surroundings for a food source. Those who discover a source return to the nest and on their way back deposit a pheromone – a chemical substance, which other members of the same species are sensitive to. The liquid trail is surrounded by an elliptic active space proportional to the pheromone concentration. If the concentration of the pheromone exceeds a certain threshold, other individuals who by chance wander within its active space get attracted by the trail and start navigating towards the food source. Once an ant starts to follow the trail, it tends to stay on it, but it may repel from it with a probability inversely proportional to the trail intensity. Then, the exploration behavior starts over until multiple positive and negative feedback loops make certain (as a rule, optimal, see Sumpter & Beekman, 2003) trails stable.

Thus, if we focus our attention on the behavior of individual ants after exposing them to several food sources, we'll see them explore the environment, make their trails (proposals) if they find a food source, follow their own trail or get dragged by others' trails if they find them more intense than their own trail, and occasionally repel and start over exploring the environment or other trails. In this way, all private information is being integrated in the swarm by each individual making

a simple computational operation and depositing the outcome of this operation in the form of a pheromone splash as it walks. As the end result of these local information transmissions, the swarm will ultimately converge on typically the better choice among the food sources, or the choice that optimally trades off food quality and the distance to the source (Hölldobler & Wilson, 1990), with almost no direct (addressed) interaction among the individuals.

Not all ant species rely on specialized pheromones in trail recruitment but may use other indirect cues (Detrain & Deneubourg, 2009). For example, ants from the species *Lasius japonicus* rely on hydrocarbons that their conspecifics deposit from their legs while walking (Akino & Yamaoka, 2005). The *L. japonicus* ants analyze the footprints to distinguish the trails laid by the members of their colony from those belonging to a different colony. The footprints work as a reliable marker for this discernment because the hydrocarbon profiles are colony-specific, while the pheromones are not. In this case, information is integrated across individuals through a passive system of environment-mediated interactions. In contrast to trail pheromones, that can be thought of as a proper signal or sign that requires active emission, footprints are passive by-products of individual decisions, that are exploited due to their property of being informative, observable, and persistent in the environment[3].

While trail recruiting has often been referred to as an epitome instance of stigmergic decision-making (Bonabeau, Dorigo, & Theraulaz, 1999; Deneubourg & Goss, 1989; Holland,

[3] Several authors (Holland, 1996; Marsh & Onof, 2008; Wilson, 1975/2000) have drawn a conceptual distinction between two types of stigmergy: sign-based stigmergy, where one agent is leaving a trace in the environment which has a proper signaling function, and by-product based passive stigmergy. In fact, while the two stigmergy types certainly differ in terms of the actual implementation of the information transmission, and are very important from an evolutionary biology perspective, they are quite the same from a computational point of view. In both cases the observer exploits modifications of the environment resulting from individual information processing by another agent to optimize its own behavioral decision. The difference is that in the case of sign-based stigmergy, evolution has already molded context-specific signs for the species and provided individual agents with sensitivity to those signs as well as with prior knowledge about their validity. Furthermore, in the evolutionary perspective it is likely that most of the sign-based stigmergy evolved as an adaptation from by-product cues which had non-zero validity. For example, in the case of the trail pheromones it is suspected that trail-laying behavior is an adaptation capitalizing on validity of products of defecation: early ants might have been following feces of other working ants returning with their harvest, which led them to the food source (Hölldobler & Wilson, 1990).

1996), the definition of stigmergy based on that of Grassé (1959) turns out to be rather broad and covers diverse mechanisms which are likely to be different in the computations that individual agents perform. One class of these possibly very different mechanisms, which still fall under the classic definition of stigmergy, is directly observable behavior. On-line sampling of others' states and immediate responding to the change in those states can be found in schooling behavior in fish and flocking in birds (Couzin, 2009; Couzin & Krause, 2003; Reynolds, 1987), and in collective decision-making in bees (Passino, Seeley, & Visscher, 2008). The computational principles in these cases are radically different from interaction via persistent changes to the environment: the agents are constrained to be synchronous; the environment has no "memory", and the change in behavior of one individual instantly propagates through the swarm. Due to synchronous properties of behavioral phenomena such as schooling and flocking it is reasonable to assume that information processing taking place in these behaviors is qualitatively different from that taking place in stigmergetic behaviors mediated by persistent traces in the environment. Nevertheless, there is no consensus among theorists on whether directly observable behavior should be counted as stigmergy or not, provoking much confusion in the literature. For example, Castelfranchi (2009) explicitly restricts stigmergy to information transmission through persistent traces in the environment, so that the receiving agent "does not perceive the behavior (during its performance) but perceives other post hoc traces and outcomes of it" (Castelfranchi, 2009, p. 325). At the same time, Huang, Ren, & Jin (2008) include both observable behavior and its traces into stigmergic behavior.

Instead of modifying the notion of stigmergy to restrict it exclusively to communication through persistent traces, I am adopting the definition of *indirect interaction* which can be found in Keil & Goldin (2006). The classic communication, in the sense intended by Franklin (1996) comprises the *direct interaction* mode: interaction through addressed messages where the sender intends to send the message to an identifiable recipient. Accordingly, "sending the message" is an

action which has no other goal than the transfer of information to the recipient. This amounts to explicit communication, including verbal and non-verbal communication.

In contrast, *indirect interaction* is defined as "interaction via persistent observable changes to a common environment; recipients are any agents that will observe these changes" (Keil & Goldin, 2006, def. 9, p. 77). Further in the description they explicitly restrict indirect interaction to asynchronous, possibly not intentional, information exchange:

> <…> acts of making changes (output) and of observing those changes (input) are decoupled in time; persistence of the environment allows the change to endure, allowing for a lapse of time before it is observed. Furthermore, the identity of the observer(s) may not yet be determined when the environment is changed, allowing for anonymous interaction. In fact, the change may not be motivated by the need to communicate, but may occur as a byproduct of the first agent's computation. (p. 77)

From this description, it can be seen how this elegant definition carves the multidimensional space of information exchange into well-defined categories of communication (direct interactions) and environment-mediated asynchronous non-direct interactions. The notion of indirect interactions works as a synonym for stigmergy if one defines the latter narrowly, as in Castelfranchi (2009) or in Dippel et al. (2014), or refers only to exchange through persistent outcomes if stigmergy is defined broadly. In both cases it doesn't include synchronous phenomena such as schooling and flocking as they do not fall into either of the interaction types.

Indirect interactions and stigmergy in particular, have drawn interest among researchers of animal behavior and cognition as an efficient social mechanism of collective-decision making (Conradt & Roper, 2005; Couzin, 2009). Some authors (Kameda, Wisdom, Toyokawa, & Inukai, 2012; Krause, Ruxton, & Krause, 2010) have also suggested indirect interactions in non-human animals as a promising model for some forms of group decision-making in humans. Little is known, however, about how efficient indirect interactions can be as a mechanism for information integration even in the context of simple perceptual judgments. In Study I of this manuscript I

report a series of experiments addressing the question how well two individuals can integrate perceptually acquired location information through indirect interactions in a shared environment.

In addition to indirect interactions which do not require communication, the environment can also play a role as a medium for conveying information intentionally. If such a communication does not involve specific codified actions aimed only at communication, this mode of interaction can be related to the notion of *behavioral implicit communication*, introduced by Castelfranchi (2003). Castelfranchi (2003) used the term to describe a situation where an agent intends (therefore, this interaction qualifies as communication) to inform a recipient, but does it through an action (therefore, it is behavioral), which has (or can have) another practical, or instrumental goal. And because the vehicle for message transmission is not specialized, and the message itself is not codified, this communication is implicit[4]. According to Castelfranchi (Tummolini, Castelfranchi, Ricci, Viroli, & Omicini, 2004), behavioral implicit communication is a "parasitical" form of communication which exploits the environment property of making behavior or its outcomes visible to other agents, and the capacity of others to interpret and respond to the observed behavior. Tummolini et al. (2004) provide the following definition of behavioral implicit communication:

> the agent (source) is performing a usual practical action *a* but he also knows and lets or makes the other agent (addressee) to observe and understand such a behaviour, i.e. to capture some meaning μ from that 'message', because this is part of his (motivating or non motivating) goals in performing *a.* (p. 220)

Coordination via behavior observable in a shared environment can help individuals to avoid a specific negotiation process in the course of a task allocation and can establish an implicit agreement on what to do (Tummolini et al., 2004, p. 219). If you are supposed to clean out the garden and see your farther taking a hand barrow, you don't need an additional explicit invitation

[4] Castelfranchi (2003) also makes a point, that the implicit character of such communication means that it is "unmarked, undisclosed, not manifest, and thus deniable" (p. 1). These latter properties are not of particular relevance for the current purpose.

"Could you help me, son?" – the action speaks for itself (under the assumption that you observe and interpret it). But communication via observable action in a shared environment can also support the process of collective decision-making. To get some intuition on how this may function in the decision-making context, consider the following scenario (the example is actually inspired by some real-life hiking experience). You and your friend come across a fork on your path through the terrain, and you have to make a decision on where to go. Your friend believes that you should go right, and you believe that you should go left. Verbal communication here does not help much because what arguments could you provide other than that you remember one thing, but your friend remembers the other thing, and you both are pretty confident in your memory? After you almost escalate the conflict, your friend just starts walking to the right (where he believes you both should go). This action is something more than just a practical action: it conveys information about your friend's uncompromising stand and his commitment to it, and perhaps also about his confidence.

Environment-mediated interactions, possibly involving behavioral implicit communication, have been studied in humans as a mechanism of joint action coordination mainly in dynamic tasks (Ganesh et al., 2014; Glynn & Henning, 2000; Harrison & Richardson, 2009; Wahn, Schmitz, König, and Knoblich, 2016; van der Wel, Knoblich, & Sebanz, 2011), but also in cooperative tasks requiring coordination on a larger time-scale (McGraw, Wallot, Mitkidis, & Roepstorff, 2014; Vesper, Morisseau, Knoblich, & Sperber, 2016). Little is known, however, how well environment-mediated interactions can support coordination in the context of collective judgment and decision-making, in contrast to verbal and non-verbal communication (Bahrami et al., 2012a) or restricted symbolic communications (Rowe & Wright, 1999).

In Study II of this manuscript I report a series of experiments that addressed how well environment-mediated interactions can support the process of coordinating agreement on collective judgment, and how efficient it can be as a mechanism for interactive inter-individual information

integration. Specifically, it addressed the question whether additional opportunities for verbal communication can add anything to the process beyond and above behavioral interactions via a shared environment, and if so, under what conditions.

1.5 Two-stage Approach to Information Integration: Judgment Revision and Joint Judgment Formation

When a group of individuals integrates individually perceived information to make a joint judgment, complex and inter-related processes take place within individual members' minds and between individuals in the form of information exchange. The two levels, group and individual, interact. "The shared and sharing aspects of group information processing are interdependent of each other. In addition, group-level processing is dependent on various aspects of individual-level processing, and individual-level processing is also affected by group-level processes" (Hinsz et al., 1997, p. 44). Taking into account obvious differences in the organization of computational units between individual brains and groups of individuals it seems reasonable to assume qualitative differences in information processing on the two levels. A proper analytic method would require isolating the two levels as much as possible to better understand the individual contributions of the processes on each level to the joint product of information integration, and then understand the interactions between the two levels. This section provides a two-stage model for inter-individual information integration proposed by several authors. Each stage is alleged to be more appropriate to investigate the group processes on one of the two levels.

In its most explicit form the two-stage model of inter-individual information integration can be found in Sniezek & Henry (1989, 1990). Sniezek and Henry (1989) were faced with the fact that neither of their models could predict the empirical results of their experiment where groups of subjects were given a task to estimate risks associated with different death causes. The models could

not accommodate group judgments which were outside of the boundaries of the initial distribution of private judgments. They suggested conceptualizing the group judgment as a two-stage process: "First, information and judgments are exchanged, with ongoing revision of individual judgments. Then, there is the process of weighting of postinteraction revised judgments. While revision is largely an individual process, weighting occurs at the group level" (Sniezek & Henry, 1989, p. 24). Accordingly, at the first stage group members exert reciprocal social influence on each other. As a result of this influence, individuals may voluntarily change their subjective judgments and the confidence associated with them. At the second stage, unless group members converge, they need to combine the distribution of revised judgments into a single group judgment. In the follow-up study, Sniezek & Henry (1990) replicated their first findings, and found that the two-stage model can well explain the group information processing of private judgments, with most of the integration taking place at the second stage: individual judgment revisions were relatively small.

A schematic representation of the two-stage model for the process of inter-individual information integration into the join judgment based on the Revision and Weighing model from Sniezek and Henry (1990) is provided in Figure 1.

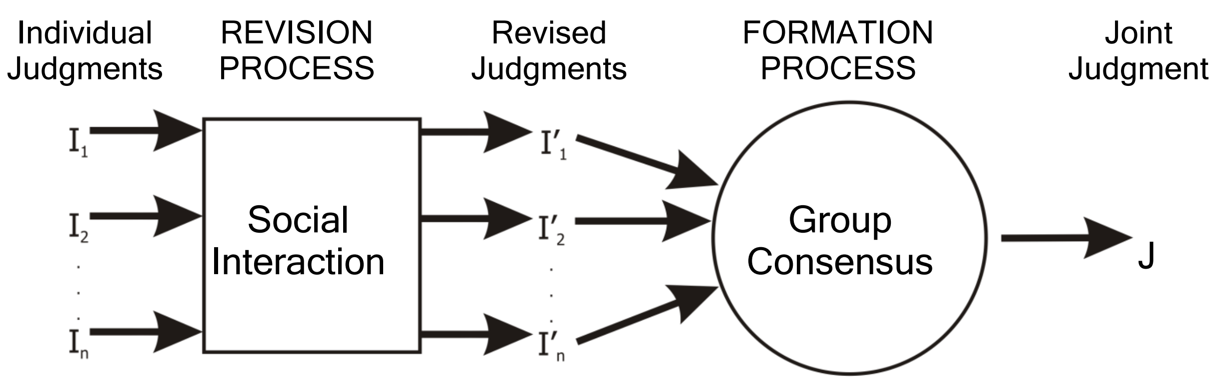


Figure 1. Representation of the two-stage model for group judgment sequence.

Similarly, Anderson & Graesser (1976) proposed a two-stage model for group attitude processing. The first stage is one of attitude formation and revision by the group members. The second stage is one of compromise-seeking among the group members. Anderson and Graesser (1976) believe that the compromise requirement introduces forces beyond information and persuasive influence on individual members. One could refer to these forces as motivations other than improving the joint judgment quality such as a desire to resolve the conflict in opinions (Einhorn & Hogarth, 1981) or to contribute to the collective decision despite personal incompetence (Bahrami et al., 2012a). Thus, while the joint response requirement is important to study the performance of a group as a system, it "largely confounds" the analysis of reciprocal influence on individual judgments. Anderson and Graesser (1976) sum up that "the two sets of forces [related to influence and to compromise-seeking respectively] seem to be qualitatively different, and they deserve study in their pure form" (Anderson & Graesser, 1976, p. 221).

Accordingly, the two studies placed in this manuscript correspond to different stages of information integration. In Study I the first stage of informational influence is addressed in isolation, and the experiments are focused on judgment revision. The main question of this series of experiments is whether individuals can improve their perceptual judgments by means of indirect [environment-mediated] interaction with another individual performing the same task.

In Study II the requirement of a joint judgment is introduced, therefore it includes the inseparable stages I and II, and is focused on joint judgment formation. The main question of this series of experiments is whether, and under which conditions, verbal communication can add to the quality of joint judgments above and beyond that achieved through interactions via a shared environment in the context of a simple perceptual judgment task.

In both series of experiments the impact of complementarity in individually held information on the interaction outcome is addressed.

2. Study I. Judgment Revision

2.1 Introduction

2.1.1 Social influence on perceptual judgments

According to the two-stage model presented in section 1.5, at the first stage of inter-individual information integration individuals reciprocally influence each other and revise their private judgments accordingly. Exerting reciprocal inter-individual influence can be thought of as an elementary cognitive operation at the group level. At the individual level, this operation is realized through complex processes of information encoding, valuation, and integration. At the group level, these processes, unfolding in the social context, behaviorally manifest in the shift in individual position following social interaction. Social influence encapsulates these processes, and its strength (measured with the size of the shift) and direction provide their simple yet exact summary[5]. The idea that social influence is the basic element of interactive social processes can already be found in works by the pioneer in the theory of self-organized animal behavior, Pierre-Paul Grassé (1963, p. 8):

> Social groups are above all characterized by the fact that any individual taken separately produces a specific stimulus upon its fellows, while the group (that can be reduced to a single fellow) produces in turn a specific stimulus that will influence the behavior of that animal. (translation from Theraulaz & Bonabeau, 1999, p. 100).

In human social psychology, social influence was a central topic since the beginning of the 20th century (Sherif, 1936); and it was brought into the focus of experimental research nearly half a century before the works of Grassé. Since those early studies tasks involving simple perceptual judgment and decision-making have been established as a widespread and by now recognized method to investigate the phenomenon (Graham, 1962; Mojzisch & Krug, 2008). Perhaps the earliest (to my best knowledge) report of a proper psychological experiment on social influence

---

[5] A similar argument is made by Graesser (1991) in respect to the concept of *social weight*, which is another way to represent social influence.

employing this kind of task can be found in Münsterberg's *Psychology and Social Sanity* (1914/2014). Participants in that experiment were instructed to decide which of a pair of cardboard sheets contained more overlaid dots. It was experimentally established that during discussion individuals often change their private judgments and adopt the majority's response.

There was a general consensus among the scholars of that time that in a situation of uncertainty about the issue in question, or where the object of perception is ambiguous, people rely more on information from others, which was quite intuitive. It was the experiment conducted by Solomon Asch (1952) that yielded results which challenged that obvious view and engendered extensive revision of existing theory on how information available from others can influence individual decisions and possibly even low-level perception. In Asch's (1952, 1956) experiments, participants were exposed to lines of different lengths, and were instructed to find the line that exactly matched the standard comparison line. The task was cognitively very simple and supposedly effortless: the line lengths were manifestly different, and when doing the task in isolation, participants had very low uncertainty about which of the lines matched the standard line. The task was made easy on purpose: Asch's initial hypothesis was that when confronted with unambiguous stimuli, people would not be sensitive to others' choices. To test this hypothesis a collective version of the task was introduced: participants performed the task in a group, where each of the individuals announced his or her private choice in turn. What was not known to the participants was that they were the only actual participant among the group. All others were confederates who unanimously gave an obviously incorrect answer on 2/3 of the trials. Contrary to Asch's initial prediction, the average proportion of real participants' incorrect answers drastically increased from 0.7% in the control condition to 37% in the experimental condition. These results provided unequivocal evidence that people conform to a majority choice even when this choice sharply contradicts unambiguous perceptual information they perceive individually.

In the ensuing discussion of Asch's experiments two factors were identified that could have led participants to adopt the majority's incorrect choice in Asch's experiments (1952, 1956): normative influence and informational influence (Deutsch & Gerard, 1955). Normative influence accounts for people's intrinsic motivation to be accepted and liked by the group, and their concern that public commitment to an outlying opinion can lead to a rejection by the group. On this account, people are aware that they are adopting a wrong choice, but they are doing it strategically not to stand out in the group. Some of Asch's participants explicitly reported such a motivation for adopting an admittedly wrong choice (Asch, 1956).

Informational influence accounts for people's motivation to increase their accuracy and for their uncertainty about being correct. Under this account, an individual believes that the majority is more likely to be correct than he or she is and resolves the informational conflict by adopting the choice made by the majority. Interestingly, the theoretical argument of Deutsch & Gerard (1955) included an assumption that others' perceptual judgments generally tend to be valid, and relying on them can be beneficial for an individual:

> From birth on, we learn that the perceptions and judgments of others are frequently reliable sources of evidence about reality. Hence, it is to be expected that if the perceptions by two or more people of the same objective situation are discrepant, each will tend to re-examine his own view and that of others to see if they can be reconciled. (p. 635).

Some of the participants in Asch's experiments (1956) explicitly reported that they *believed* that the others were correct, favoring the informational influence explanation.

There was a third category of participants, whose response behavior was more difficult to explain: these participants reported that they actually *saw* that the incorrect choice was the one matching the standard stimulus. Although Asch (1951) believed that some individuals can experience perceptual distortions due to social influence, he interpreted the results of his experiments as "judgment distortions" – a tendency to comply with the majority's response due to decreasing certainty in one's own perception.

The “perception distortion” hypothesis was further developed by Moscovici (1980), who was interested in the mechanism by which a minority can exert social influence on a majority (Moscovici, 1976; Moscovici , Lage, & Naffrechoux, 1969; Moscovici & Personnaz, 1980), and who suggested that “social conversion” (Moscovici, 1980) can work through altering the underlying perception processes under social influence. To test this hypothesis, Moscovici & Personnaz (1980) developed an experiment where they implemented what nowadays would be called an implicit measure of change in individual perception.

In their experiment participants first judged the color of uniform projected slides which varied in light intensity but were always unambiguously blue. Then they judged the afterimage reporting the perceived color of a white screen presented after the colored stimulus presentation. When individuals perform this task, they perceive the afterimage to have a color that is complementary to the color of the preceding image, that is, yellow for a blue image. The experimental manipulation introduced a confederate, whose answer allegedly represented either a majority (the legend was that over 80% of previous participants in this experiment gave the same answer as the confederate) or a minority (the legend was that less than 20% of previous participants in this experiment gave the same answer as the confederate). The confederate consistently provided a wrong answer on the image color saying publicly it was green (instead of blue). In the next block of trials the afterimage judgments were made privately. If the confederate’s image judgment influenced participants’ perceptual processes, then the judgment of the afterimage color would shift towards the red spectrum, which is the complementary of green.

Moscovici & Personnaz (1980) found that when the confederate represented the majority, participants more frequently judged the stimulus image color to be green. However, their judgments of the color of the afterimage remained unchanged. In contrast, when the confederate was representing a minority, participants’ judgments of afterimage color significantly shifted towards the

red spectrum. This led Moscovici & Personnaz (1980) to make the intriguing conclusion that while the majority exerts influence through response change and verbal compliance, a minority can exert influence via altering genuine perceptual processes. The authors' interpretation was that people seek to resolve an "unconscious conflict" when confronted with conflicting information. When the conflicting information is represented by one individual, people think that this individual's perception is miscalibrated and ignore it. When the same information is represented by the majority people think that now it is their own perception that is totally miscalibrated and adopt the majority's response. Finally, when the same information is represented by a minor proportion of others, they recalibrate their own perception to resolve the conflict.

The results obtained by Moscovici & Personnaz (1980) had some far-reaching implications but turned out to be unreliable: several authors failed to replicate their findings using the same method and stimuli (Doms & Van Avermaet, 1980; Martin, 1998; Sorrentino, King, & Leo, 1980). The original study (Moscovici & Personnaz, 1980) was also criticized regarding methodology: the utilized afterimage response scale was designed in such a way that any increase in judgment uncertainty was interpreted as a shift in participants' perception in the predicted direction (Sorrentino, King, & Leo, 1980). Finally, several alternative explanations were suggested. Doms & Van Avermaet (1980) suggested that social influence modulates perception via attention mechanisms: exposure to a conflicting judgment made participants inspect the stimulus more carefully and intensively, paying more attention to the green component of the stimulus color spectrum. Sorrentino, King, & Leo (1980) reported empirical evidence for their claim that "perceptual distortion" is related to suspicion. According this claim, initially, more suspicious participants are more attentive than less suspicious participants to all aspects of the experiments. Confronted with a conflicting judgment, more suspicious participants may "conceivably stare more intensely at the color stimuli than do unsuspicious subjects, and the shift in their afterimages over

time may be a result of this greater attention" (p. 298). Altogether, it seems safe to conclude that the original findings by Moscovici & Personnaz (1980) are suggestive but not compelling. Still, some theorists consider perception alternation as a viable mechanism for social influence, and propose coupling more fine-grained behavioral and neurological methods to resolve this question due to its theoretical relevance in an evolutionary perspective (Mojzisch & Krug, 2008).

The focus on informational influence in Deutsch & Gerard's (1955) account implies that social influence might be advantageous for individuals, as it increases their chances of making a better decision or judgment. Evidence for positive effects of social influence on individual judgments has been sparse in studies on small groups. A positive effect of judgment revision was reported in the earliest study of Münsterberg (1914/2013): the proportion of correct judgments shifted from 52% to impressive 75%. Another report of this effect can be found in a study conducted by Bechterev & Lange (1924) reviewed in Lorge et al. (1958). In their experiment, Bechterev & Lange (1924) gave their participants a task of making judgments in various domains including judgments of perceived physical properties. Participants were making their judgments privately, and then the private judgments were presented for a discussion. After the discussion the new private judgments were made. The interaction had a positive effect: post-discussion judgments were more accurate than pre-discussion judgments. All individuals benefited from this interaction, although less competent individuals had more benefit.

In the experiment by Jenness (1932) participants were estimating the number of beans in a bottle (see also section 1.2). Participants made private judgments, and then discussed their opinions in a face-to-face setting to provide a group judgment, and finally provided an individual post-discussion judgment. Jenness found that when ad hoc groups were composed of individuals with heterogeneous initial judgments, individual post-discussion judgments were more accurate than individual pre-discussion judgments. In this case the fact that the effect was reliably observed only in

heterogeneous groups can be explained by a statistical property of heterogeneous judgments (judgments spaced at a larger distance from each other) to have a higher rate of bracketing – the expected proportion of instances when the true quantity value lies between individual judgments (Larrick, Mannes, & Soll, 2012; see also section 3.1.2 for detailed description of how bracketing works).

For early research on conformity, however, empirically investigating the relations between social influence and the quality of revised judgments was of little concern. The reason for neglecting informational aspects of social influence and the focus on normative and motivational factors related to social influence was that the dominant view on conformity at the time was that it was a maladaptive phenomenon leading individuals to irrational fallible decisions. The main question of interest was why individuals cannot resist social influence even when it is in their best interest to do so (Soll & Larrick, 2009; Tajfel, 1969). A shift and expansion of the research focus took place when researchers started to realize that adopting others' decisions can be adaptive if viewed from an evolutionary perspective or viewed from a social science perspective (Kameda & Tindale, 2006). The statistical argument was that an aggregate of diverse inaccurate but at least somewhat valid decisions or judgments can result in a more accurate response than independent responses (Einhorn, Hogarth, & Klemptner, 1977; Hastie & Kameda, 2005; Hogarth, 1978; Yaniv, 2004a, 2004b).

This gave start to another thriving line of research, which focused on how individuals utilize information available from others (often referred to as "advice") estimating the same quantity in a situation where the correct judgment is not obvious, and where all individuals are expected to make randomly distributed errors in their best estimate. In a certain sense, this line of research took the opposite extreme compared to Asch's experiments. In its attempt to factor normative influences out, it mostly eliminated the core of the group process – inter-individual interactions, allowing only for communicated messages containing symbolically encoded individual judgments. Nevertheless,

this research has yielded some genuine insights on how individuals integrate information available from others with information that is available to them. The next section will provide an overview of research performed in the Judge-Advisor-System approach addressing characteristic methodological tools and key findings relevant for the present purpose.

2.1.2 Judge-Advisor System perspective

The problem of decision or judgment revision in the light of information available from others, and, in particular, the question of how judgement revision affects the quality of revised judgments, has been extensively studied in a research field studying the process of advice utilization also known as Judge-Advisor-System, or, JAS (for a review, see Bonaccio & Dalal, 2006; Yaniv, 2004b). The first study in social psychology which gave start to the new topic and also contoured an experimental procedure that would become the golden standard in the field was a study by Brehmer and Hagafors (1986). In their task participants were instructed to combine multiple cues in the format of medical test results with advice from multiple experts based on the same cues to make a quantitative judgment about the severity of a disease. The cues (medical test results) were correlated with severity of the disease. The innovation by Brehmer and Hagafors (1986) was that in the motivation for their study they explicitly emphasized the shift from normative aspects of social influence to information processing.

Accordingly, the experimental procedure proposed by Brehmer and Hagafors (1986) was tailored to address exclusively the informational aspect of advice utilization prefiguring the key methodological features of future research in JAS: "The general aim of this experiment was to assess the role of cognitive factors as clearly as possible. We therefore decided not to confront the decision-maker with live experts. Instead, <…> the experts appeared only in the form of messages on pieces of paper" (p.185). Their analysis included two aspects of advice utilization: achievement

(accuracy of judgments), and the amount of influence the experts' advice exerted on the final judgment of the participants. Brehmer and& Hagafors (1986) also introduced the concept of *weight* to capture the amount of influence advisors had on the judge's final decision. Since then the key question in the JAS research has been generally formulated in terms of how participants *weigh* information provided by an advisor or multiple advisors under specific conditions.

In studies on advice-taking a participant (the judge) is asked to provide a decision or judgment based on information available to her and information coming from the advisor(s). Participants usually have a choice to either utilize or ignore information coming from the advisor(s). The key question addressed in these studies concerns the degree of utilization of the advice and its influence relative to the judge's initial response or relative to the influence of other sources (or advisors). The exact methodology employed in these studies varies, but in studies addressing judgments on a continuous scale, the events to be judged are dominantly drawn from abstract domains such as dates of historical events (Yaniv, 2004b) or estimates of product sales (Fischer & Harvey, 1997). The judgments are typically provided in format of a numeric estimate. Interactions between the judge and the advisor(s) are typically highly restricted: normally it is a one-way transmission of a piece of information, minimally a numeric estimate provided by an advisor to the judge in written form. This precludes any influence of the advisor on the judge beyond the advice itself. Often, experimental set-ups presume physical isolation of participants. In this case the experimenter would transmit messages from the advisor(s) to the judge (Bonaccio & Dalal, 2006).

A general finding in this field is that judges utilize the advice provided at least to some extent (Bonaccio & Dalal, 2006). At the same time, judges tend to weigh the advisor's judgments less than they should to gain maximum benefit from information integration. In other words, judges tend to discount the advisor's judgment relative to their own– a robust effect that is known as "egocentric advice discounting" (Bonaccio & Dalal, 2006; Harvey & Fischer, 1997; Yaniv, 2004b; Yaniv &

Kleinberger, 2000). Nonetheless, in most cases advice utilization leads to an improvement in accuracy: revised judgments are generally more accurate than an individual's initial judgments, but the achieved level of accuracy is lower than predicted by normative models of information integration (Bonaccio & Dalal, 2006; Sniezek, Schrah, & Dalal, 2004; Yaniv, 2004a).

Several experiments have also demonstrated that people are sensitive to the expertise of the advisor and to their own expertise (Bonaccio & Dalal, 2006). Weaker egocentric discounting has been observed in judges who are less competent in the domain of judgment relative to their advisor(s) (Harvey & Fischer, 1997; Sniezek, Schrah, & Dalal, 2004) or relative to other judges (Harvey & Fischer, 1997; Yaniv, 2004b; Yaniv & Kleinberger, 2000). Also, judges discriminate advice of high and poor quality, and discount inaccurate advice more radically than accurate advice (Yaniv & Kleinberger, 2000). Yaniv and Milyavsky (2007) found that in their assessment of the advice quality judges rely on the difference between the advisor's judgment and their own judgment, and give less weight to judgments that are further away from their individual judgment. Yaniv and Kleinberger (2000) investigated the dynamics of judge-advisor interactions and discovered several patterns of how judges track the credibility of the advisor and his or her advice. They found that judges can learn credibility from trial-by-trial feedback. Advisors who have been providing accurate judgments for a certain period slowly gain reputation, and the weights given by the judge to their advice will gradually increase. However, reputation is easy to lose: judges are very sensitive to inaccurate advice. If an advisor after a long run of accurate judgments starts to provide inaccurate judgments the magnitude of weights the judge will assign to further judgments from the advisor will drop rapidly.

Sorkin, Luan, and Itzkowitz (2001) conducted an experiment (reported in Sorkin, Luan, & Itzkowitz, 2004) which stands out in the literature in that it combined elements of JAS research and collective perceptual decision making. In this experiment participants performed a game-like version

of a joint signal-from-noise discrimination task introduced by Sorkin et al. (2001; see section 1.2 for more detailed description of this study). Participants were deciding whether tick-marks on analogue gauges presented on a graphical display were generated by a signal-plus-noise or noise-alone event. In the spirit of Brehmer and Hagafors (1986) and other than in the original study (Sorkin et al., 2001), participants were performing the task individually and were then exposed to likelihood estimates generated by virtual group members. The question of interest was how participants would utilize these estimates to arrive at their final decision. The result were different from the ones obtained in an interacting group setting (Sorkin et al., 2001): in contrast to the original study, participants optimally integrated advisors' estimates and exhibited no reduction in their efforts relative to individual trials. Sorkin et al. (2004) conclude that under some circumstances, people are able to integrate perceptual information from others in a statistically optimal fashion. However, generalizations from the results of this experiment should be taken with caution. Virtual experts, who as a rule are generated parametrically, would normally behave more internally consistent than real human participants and, therefore, possibly gain more credibility for the judge.

The theoretical models developed to explain advice utilization specify the cognitive mechanisms that humans (and supposedly other animals, see Mojzisch & Krug, 2008) use when they revise their initial estimations and integrate external information coming from others. For a scenario where two individuals interact, three strategies are commonly cited as models for the psychological process of advice utilization: averaging, selecting one judgment of the two, and differential weighing of the two judgments (Sniezek & Henry, 1989; Soll & Larrick, 2009). All three strategies can be thought of as different variants of linear information integration (Sniezek & Henry, 1990). Humans have been shown to calibrate their advice-utilization strategy to environmental[6] properties in the

[6] It should be noted that the concept of *environment* in the JAS research is used in a different sense from that used in the context of EMI. While in the context of EMI *environment* refers to surroundings upon which an agent acts

context of uni-dimensional estimation tasks (Soll & Larrick, 2009). A review on when certain strategies should be favored over the others can be found in Soll and Larrick (2009) as well as in section 3.1.2 of this manuscript. At the individual level, the three strategies (averaging, selecting one, differential weighing) correspond to individual mechanisms through which individuals integrate the outcome of others' information processing into their own representation. Reciprocity in these processes allows information to be integrated on the group level, and if there is no consensus requirement, collective decision is the distributed product of applying integration strategies individually (Kameda et al., 2012).

This raises the question whether the findings obtained in JAS settings can be generalized to a more interactive situation, where individual roles in a group are undifferentiated, and at the same time individuals do not necessarily perceive themselves as members of a group. These situations occur when individuals do not share a joint goal, the individual outcome does not necessarily depend on the joint product, and therefore individuals utilize information publicly available from each other for their own purpose. It is also crucial, though, that individuals have no incentive to hide behavioral outcomes of their private decisions making them public.

The second question is how the environment (meant in the context of environment-mediated interactions, see section 1.4) can support information integration processes, and whether findings from JAS research generalize to situations where the information exchanged involves publicly perceivable changes of the physical environment. This type of information exchange directly relates to research on stigmergy (see section 1.4 for an overview). Examples of how stigmergy can aid individual decision-making are reviewed in the next section.

---

and from which he or she receives perceptual input, in the context of JAS *environment* refers to informational conditions in which an agent is making his or her decision. For example, the amount of correlation between advisors' judgments would refer to environmental factors in JAS terminology.

2.1.3 Environment-mediated interactions (EMI) as a mechanism for joint perception

In studies of group decisions and group judgments that are conducted on other species than humans the roles of "advisor" and "judge" are rarely differentiated. Instead, collective decision-making is viewed as emerging from local interactions with horizontal decentralized information exchange. In this sense, decision-making in non-human species sometimes bears the key egalitarian properties of decision-making procedures prevalent in human societies (Kameda & Tindale, 2006). In many cases the environment plays an important role in this process as a medium and the depository for the exchange of the products of individual information processing. Extensive reviews on environment-mediated decision-making in non-human animal species and discussions of its relevance for studies in human cognition can be found in Conradt and Roper (2005), Kameda et al. (2012), Krause et al. (2010), and in Couzin (2009).

One of the mechanisms for information exchange is producing publicly observable changes in the environment, also known as *stigmergy* (Franklin, 1996; Theraulaz & Bonabeau, 1999). There is some ambiguity in the way different authors use the term "stigmergy" (see section 1.4) and depending on the definition qualitatively different mechanisms can be covered by the concept. This section will focus on a specific type of stigmergy where information is exchanged by leaving persistent traces in the environment. I will refer to this scenario as *indirect interactions* following Keil and Goldin's (2006) terminology. Research on indirect interactions in biological systems provides many examples of how such interactions can aid individual organisms to make better decisions. The general principle is that an organism observes traces that were left as the result of another conspecific's actions. Because the traces were produced as a result of another organism's behavioral decision they convey some information about this organism's information processing. Relying on these traces can help the observer to improve its own decision, and in this sense it is very similar to advice utilization processes (see section 2.1.2).

Perhaps the example most commonly used as an illustration of indirect information exchange is pheromone trail-laying by ants (Hölldobler and Wilson, 1990; Deneubourg & Goss, 1989; see also section 1.4 of this manuscript). When an ant scout locates food it deposits a pheromone on its way to the nest and back. Its foraging behavior is influenced by sensing the food source and by sensing pheromones deposited by itself or deposited by other ants. In being influenced by another ant's pheromone individual ants integrate the information collected and processed by their fellows. This way they can be attracted to the food source they themselves have not explored or they can be attracted back to a source they rejected in favor of another source.

Crucially, to maximize efficiency in relying on observed pheromones, certain cognitive mechanisms must be (and, in fact, are) in place on the level of individual information processing. These mechanisms must ensure that the impact of information source on the individual behavior is contingent on the informative value of the source. In various ant species several such mechanisms have been identified. The most important ones are sensitivity to the quality of perceived food source (or another target object) and to the strength of the source sensation, and sensitivity to the reliability of information available from other ants.

The fact that this contingency exists is indicated by probabilistic response that a pheromone evokes in the ant observing it, where the probability of the observer turning to the trail is contingent on a number of factors. The pheromone concentration threshold of provoking a turning response in those who sense it is hypothesized to depend on the condition of an individual, its history, and other contextual factors (Calenbuh & Deneubourg, 1990). Thus, an individual who has located a high-quality food source would require a stronger pheromone concentration on a trail to get attracted to it compared to an individual who has not located a food source itself.

To calibrate the informative value of the deposited pheromones, trail-laying behavior in ants is modulated by the quality of located food source: there is a positive dependence between the

amount of pheromone deposited by individual ants and calorific value of the source (Beckers, Deneubourg, & Goss, 1993; De Biseau, Deneubourg, & Pasteels, 1991; Jackson & Châline, 2007; Verhaeghe, 1982). Thus, the strength of the pheromone reflects its reliability. To exploit this information contained in deposited pheromones, ant foragers are endowed with sensitivity to the strength of the pheromone. Their individual turning response exhibits a non-linear reaction to the concentration of the pheromone trail, with stronger signals evoking a disproportionally more reliable response in the observer (Sumpter & Beekman, 2003). The non-linearity of the response coupled with environmentally supported positive and negative feedback loops (Hölldobler & Wilson, 1990), works as an amplifier of small differences between the sources, and allows the colony to rapidly lock on the trail to the richer source. This combination of individual information processing and information exchange via the environment allows ant colonies to make an optimal choice between two or more food sources of different quality, which has been demonstrated in a number of laboratory experiments (Beckers, Deneubourg, & Goss, 1993; Beckers, Deneubourg, Goss, & Pasteels, 1990; De Biseau, Deneubourg, & Pasteels, 1991; Pasteels, Deneubourg, & Goss, 1987).

Collective perceptual decision making via indirect interactions is not limited to ants. Certain species of gregarious caterpillars employ chemical trail-laying to recruit their fellows to located feeding sites (Roessingh, 1990; Fitzgerald & Peterson, 1988; Peterson, 1988). Larvae from the genus *Yponomeuta cagnagellus* combine chemical information from the recruitment and exploratory trails with tactile perception of the silk strands which these caterpillars trace as they roam in search for food (Roessingh, 1989; Roessingh, Peterson, & Fitzgerald, 1988). Following slime trails left as the result of conspecifics' locomotion allows some intertidal mollusks to locate better protected rest sites (Focardi, Deneubourg, & Chelazzi, 1985). Gregarious intracortical-feeding larvae of a bark beetle, *Dendroctonus micans* use a passive mechanism of chemical indirect communication (Deneubourg, Grégoire, & Le Fort, 1989): these larvae emit aggregation pheromones at a constant rate and get

attracted to the pheromones of others. This simple mechanism, coupled with random search for a food source, allows the feeding group to converge on the feeding site located by single fellows.

There is much less knowledge on the role of indirect interactions in mammals. One obvious example of information exchange in an indirect way is communication via odors which many species trace in the environment with their urine or exocrine gland secretions (Halpin, 1986; Gorman & Trowbridge, 1989). The evolutionary functions attributed to specialized communicative odors are believed to aid individual animals to solve social coordination problems, such as mating, clumping, or avoiding competing conspecific individuals or groups (Brown, 1979; Halpin, 1986). It has been speculated, though, that as a side effect these odors can aid individuals to make better perceptual decisions (Theraulaz & Bonabeau, 1999). For example, social following of urine odors of a conspecific who had succeeded in food search can lead to a feeding site discovered by the previous animal. Thus, Gorman and Trowbridge (1986) provide an example of European badgers who squart-mark onto the grass as they forage for earthworms. This behavior is hypothesized to help individuals to locate patches where they had been previously foraging. Equally, one could hypothesize that these squart-marks could provide other badges with cues of spots rich with earthworms.

Distributed collective decision-making originally discovered in social insects has been suggested as a possible model for inter-individual interactions emerging in small-groups (Kameda et al., 2012). Several studies have begun to investigate "herding" behavior in humans, where a group is making a navigational decision through local online interactions without direct communication (Raafat, Chater, & Frith, 2009). At the same time, to my best knowledge, the question whether individuals can make better (more accurate) perceptual judgments by means of simple indirect interactions has not yet been addressed in empirical studies. In the following, I will propose a new experimental task that can be used to study how individuals making judgments under perceptual

uncertainty would integrate information by means of indirect interactions. Then, I will report a series of experiments which tested several questions related to previous research on the Judge-Advisor-System (Bonaccio & Dalal, 2006; Yaniv, 2004b) and new questions that address the role of complementarity of information in this process.

2.1.4 Method overview and hypotheses

According to Soll and Larrick (2009, p. 780), the fundamental but unanswered question in the literature on social influence is whether people use the most appropriate strategies to revise their judgments in realistic environments (Tajfel, 1969). To address this question, we developed a new experimental task which adopts certain methodological elements from research on social influence (section 2.1.1), research on Judge-Advisor-Systems (section 2.1.2), and collective decision-making through indirect interactions (section 2.1.3). The latter assumes the following characteristics which distinguish indirect interactions from communication (Keil & Goldin, 2006; p. 81):

- *time decoupling* (asynchrony): due to persistence of the environment, state changes in the environment can be observed not at the time of their execution;
- *space decoupling* (locality): one agent may leave after making a state change, and the second agent may arrive later to observe it;
- *non-intentionality*: indirect interaction does not require an intent to communicate;
- *analog form*: the medium of an indirect interaction may be the real world, for example, for embedded or situated agents;
- *mutual causality* (Def. 5, p. 75): the outputs of each agent may causally influence that agent's later inputs;

The last feature distinguishes indirect interactions from the one-way communication commonly employed in JAS research (Bonaccio & Dalal, 2006). Keil and Goldin (2006) emphasize

reciprocal influence as a key feature of true interaction. Hence, introducing behavioral feedback is a necessary component which needs to be included in the design of an experiment if behavioral processes observed in the experiment are assumed to be an accurate model of a group process.

Therefore, the new task did not only enable individuals to make private location judgments under perceptual uncertainty but they were also given an opportunity to revise their judgments after observing judgments made by another individual. Judgments were made and persisted until revised in a common virtual environment. To address the role of complementarity (section 1.3) in the process of information integration, the judgments and the subjective uncertainty associated with the judgments were made two-dimensional. For this purpose, geometric properties of subjective 2D uncertainty were reliably manipulated in the new task. Crucially, we manipulated the spatial orientations of individual uncertainty, while keeping the total amount of perceptual uncertainty fixed across orientations for each individual. This allowed us to address the question whether people are sensitive to these higher-order structural properties and whether the way they utilize information available from another reflects the difference in information available individually.

We asked pairs of participants to take turns in locating a target at the bottom of a 3D container displayed as a 2D projection on a monitor. This was achieved by placing a pointer on the top plane of the container exactly above the target (so that a perpendicular would connect its apex with the centre of the target, see Figure 2A on p. 59). Pilot studies showed that individual performance in this task corresponds to van Beers et al.'s (1998) previous findings on visual object localization: errors of judgments were scattered in elliptic distributions. Participants made larger (more variable) errors along their virtual line of sight (the dimension of their high uncertainty, see Figure 4 on p. 67). The dimension at the right angle to the line of sight was the dimension of participants' low uncertainty (errors had lower variance on this dimension). The orientation of the ellipses describing the error distribution with respect to the container's intrinsic coordinate system is

determined by the generated viewpoint on the container. One can think of the viewpoint manipulation as of the participant being rotated around the container while keeping the center of the container and the participant's "cyclopean eye" on one line. If the container is centered at the origin of the coordinate system, this line will coincide with participant's line of sight generated by the viewpoint (Figure 2C, p. 59), and its orientation in respect to the origin will determine the orientation of the error distribution. This bivariate distribution of errors is supposed to reflect participants' subjective uncertainty about the true target location.

We manipulated the difference between two participants' uncertainties with respect to the container's intrinsic coordinates by varying participants' relative viewpoints on the container. In the first experiment there were three levels: in the orthogonal J90° condition the angular difference between the two views was 90° so that the dimension of low precision (high uncertainty) for one participant corresponded to the dimension of high precision (low uncertainty) for the other participant and vice versa, thus full complementarity of information between two perceptions was implemented. In the J45° condition participants' viewpoints still implied different reliabilities but to a lesser degree than in the 90°condition. In this condition the two individual uncertainty distributions formed an "X" shape (see Figure 2B, p. 58). In the J0° condition both participants had the same viewpoint on the object implying overlapping (fully redundant) uncertainty distributions.

We also obtained an individual baseline (IS) reflecting the accuracy of locating targets from one viewpoint. This corresponded to performing one's own part of the task in the joint condition without observing the partner's judgements. Thus the IS condition provided an estimate of an individual's accuracy without additional information and, therefore, a lower bound for accuracy in the joint conditions.

Hypotheses

The main aim of this study was to investigate how individuals would integrate their partner's judgments under different conditions of complementarity. Following the tradition established by Brehmer and Hagafors (1986), two aspects of participants' behavior were assessed: their achievement - (accuracy) improvement resulting from the judgment revision, and the relative influence that another's judgment had on the participant's revised judgment. This influence can be captured with the concept of relative importance, or *weight* (Anderson & Graesser, 1976). More specifically, the revised judgment can be represented as a weighted sum of the participant's initial judgment and another's observed judgment. The main question was whether participants can adjust their weighing strategy to informational properties arising from the complementarity of information access between the two participants and the structure of their own and another's uncertainty about the target location. The following hypotheses were tested:

Hypothesis I. According to the theorem of Maximum Collective Benefit (section 1.3), information integration is expected to produce more benefit when the amount of complementarity of two individuals' accuracies on two orthogonal spatial dimensions is increased. The benefit should be maximal when two individuals' perceptions of the same situation differ maximally in their accuracy on two complementary dimensions so that one individual's accurate dimension corresponds to the other's inaccurate dimension and vice versa. Accordingly, the improvement in accuracy relative to purely individual performance was expected to be higher in the J45° condition than in the J0° condition, and to be higher in the J90° condition than in the J45° condition.

Hypothesis II. The individual accuracy gain resulting from observing a partner's judgments should be achieved through a decrease in the variability of revised judgments (Stewart, 2001; Yaniv, 2004a, 2004b). Therefore, the reduction in variability on a participant's inaccurate dimension (Ego-Y) was expected to be higher in the J45° condition than in the J0° condition, and to be highest in the

J90° condition (in the J90° condition integrating another's judgments implies the largest benefit). For the accurate Ego-X dimension the opposite pattern was expected (in the J90° condition integrating another's judgments implies the smallest benefit).

Hypothesis III. In analogy to findings in the domain of visual-proprioceptive multi-sensory integration within individual brain (van Beers et al., 1999), we expected individuals to be sensitive to the geometrical structure of their localization precision. Participants were expected to apply a dimension-selective weighing strategy to observed judgements provided by another individual. Because higher uncertainty implies larger susceptibility to external influence (Deutsch & Gerard, 1955), the more an individual is uncertain about his judgments, the more his revised judgments are expected to be influenced by another's judgment. Accordingly, we predicted the weights on the inaccurate Ego-Y dimension to be overall larger than on the accurate Ego-X dimension.

Hypothesis IV. Individuals were expected to take into account the geometric properties of the reliability of another's judgments. This implies that weights on the inaccurate Ego-Y dimension and on the accurate Ego-X dimension were expected to be different and to show a different pattern of change under different complementarity conditions. In the J0° condition for each pair of participants and for each of the two dimensions two scenarios are possible: both participants are equally accurate, or one participant is more accurate than the other participant. In the latter scenario the stronger influence of judgments of the more accurate participant on revised judgments of the less accurate participant is expected to be averaged out by weaker influence of judgments of the less accurate participant on revised judgments of the more accurate participant from that dyad. Therefore, in the J0° condition we predicted the average of weights across participants to be at the 0.5 value unless there were any general self- or other- prioritizing bias in individual weighing policy. As the amount of complementarity is increased, another's judgments will become more reliable on participant's Ego-Y dimension. Therefore, Ego-Y weights were expected to be higher in the J45°

condition than in the J0° condition, and to be higher in the J90° than in the J45° condition. At the same time, as the amount of complementarity is increased, another's judgments will become less reliable on participant's Ego-X dimension. Consequently, Ego-X weights were expected to be lower in the J45° condition than in the J0° condition and lower in the J90° condition than in the J45° condition.

2.2 Experiment 1

2.2.1 Method

Participants

Twenty-four students (16 females, 8 males) aged between 19 and 26 years ($M = 21.3$) were tested in pairs. One pair was replaced because one of its participants performed very poorly (three SDs below average), and two additional pairs were replaced because of a program or experimenter error. Participants in pairs were always of the same gender and were paid for their participation. Participants provided informed consent on their participation in the experiment[7].

Material and apparatus

Experimental stimuli were presented using Apple iMac computers (2.5 GHz Intel Core i5 iMac desktopss with 21.4" Display and AMD Radeon HD 6750M 512 MB onboard graphics) and an additional external monitor (BenQ RL2240H 21.5, connected to one of the iMacs). The screen resolution was always set to 1600 X 900 pixels. The two iMacs and the external monitor were calibrated to have matched colour output. Logitech Attack3 joysticks were used as input devices with which participants gave their judgments. The virtual environment for the location task was

[7] Participants provided informed consent and were remunerated with a fixed payment per hour in all experiments reported in the rest of the thesis.

generated using the perspective mode of the MatLab (The Mathworks, Inc.) software. The main element was a two-dimensional projection of a three-dimensional square-based rectangular cuboid (see Figure 2A) with a length × width × height ratio of 2×2×1. The cuboid was centered on the screen area (263×204 mm, expressed as dimensions of the MatLab axes rectangle) and simulated a real-world cuboid (256×256×128 mm). The actual visible length of the sides of the cuboid varied as a function of the virtual camera orientation relative to the cuboid according to the laws of perspective projection. The filling of upper and lower facets of the cuboid had the following proportions of color components in RGB code: [.1 .8 .8] with .5 transparency value. Its side facets were fully transparent. To provide participants with cues about their current spatial orientation, edges of one of the cuboid sides were colored green and had line width thickened to 1.5 points. All other edges were colored black and had standard line width – 1 point. The target appeared in an inner area of the bottom plane of the cuboid with the constraint that it could not be closer to the edges than 1/5 of each side length.

Different viewpoints on the cuboid were generated by varying two parameters of the camera orientation: azimuth (polar angle in the x-y plane) and elevation (angle above and below the x-y plane), see Figure 2C. Throughout the experiment, elevation was kept constant at 14°. The azimuth values were varied to generate different viewpoints on the layout. Four viewpoint sets were used in the experiment, each of which included three viewpoints, which could comprise perspective pairs with angular difference 45° and 90° (see Table 1). The camera view angle was set to a constant value equal to 6.3° in an attempt to minimize angular distortions in the scene (the angle of an average telephoto lens ranges from 10° to 35°). The distance from the cuboid center to the camera was

constant. The geometrical center of the cuboid was always scene-centered and, hence, screen-centered.

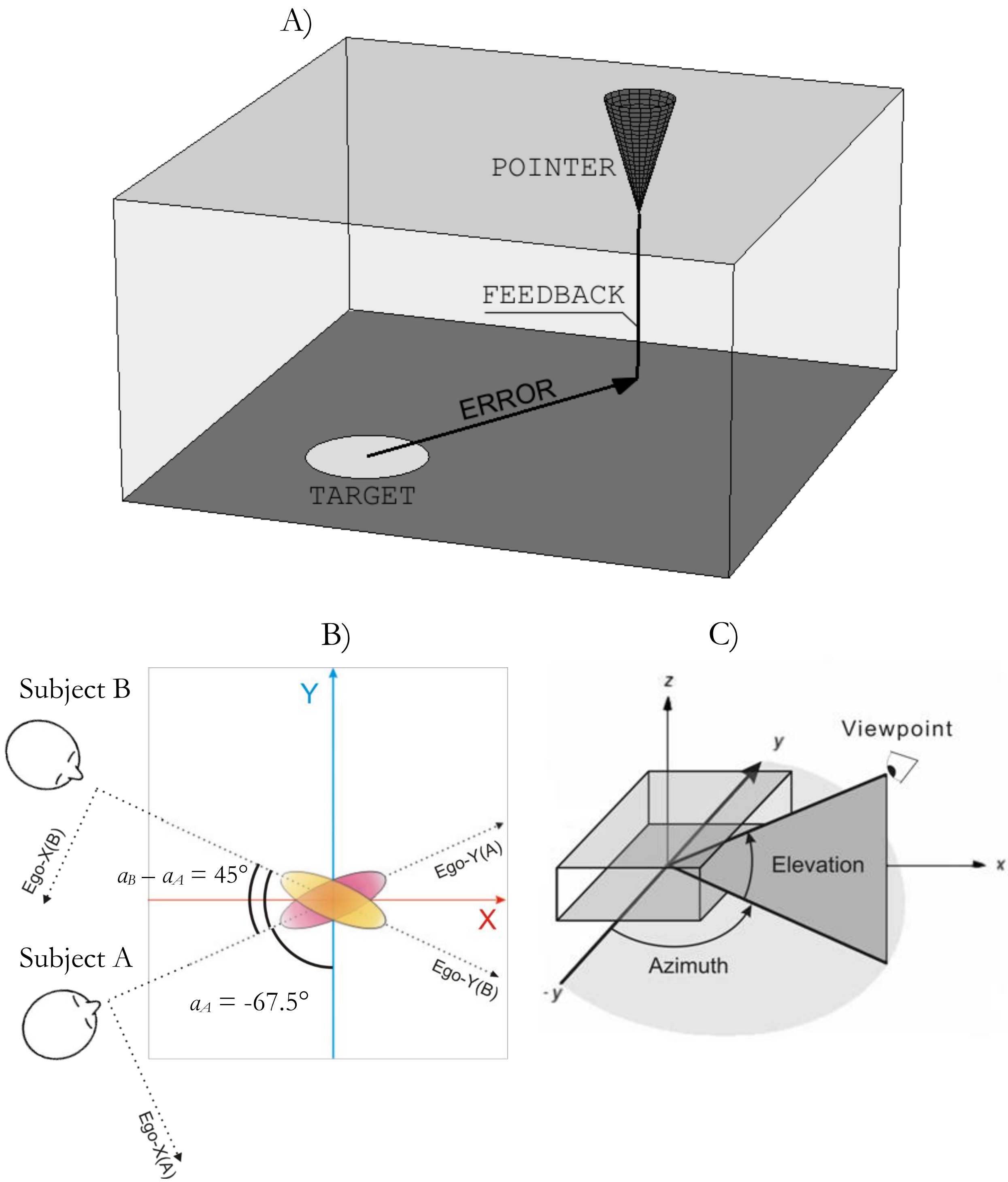


Figure 2. Virtual environment used for the location task. A) Main elements of the virtual

layout from participants' view. Participants were asked to locate the pointer above the target so that it would point exactly to its centre. FEEDBACK was only provided once at the end of a trial (i.e., after the first judgment in the IS condition and after the sixth judgment in the joint conditions, see Figure 3 on p. 64). It consisted of a line that projected the position of the pointer on the upper plane to the lower plane, visualizing the distance between the pointer and the target on the lower plane (ERROR). The size of the pointer and the target were much smaller during the actual experiment. B) A schematic representation of the top-down view on the virtual layout. The colored solid lines represent the intrinsic coordinates of the cuboid; the dashed lines represent participants' respective egocentric coordinate systems. The two ellipses illustrate idealized uncertainty distributions for the two participants with respect to the correct pointer's location. Because perception is much more inaccurate on the depth dimension (Ego-Y) than on the dimension orthogonal to the depth dimension (Ego-X) uncertainty distributions are expected to generally take the form of elongated ellipses with high variability on the Ego-Y dimension and low variability on the Ego-X dimension (van Beers et al., 1998, 1999). In this illustration the azimuth value of participant A's viewpoint is -67.5°. The angle between the two participants' viewpoints is 45°. Note, that in order to depict accurate virtual distances, participants' heads would need to be drawn more than ten times farther away than on the picture. C) A schematic illustration of parameters that define different viewpoints on the cuboid in the virtual environment (virtual camera orientation). Elevation (the angle relative to the x-y plane) was constant throughout the experiment. Azimuth (polar angle in the x-y plane) was controlled to vary viewpoints and angular differences between the two participants' viewpoints in the Joint condition.

The viewpoint sets were chosen so that all four azimuth values (-112.5; -67.5; -22.5; and 22.5) occurred with the same frequency. One of the participants in a pair (P1) had the same viewpoint in all three joint conditions, the other participant had a different viewpoint in each of the three joint conditions (J0°, J45°, and J90°). Initial analyses showed that there were no effects of whether participants had the same perspective or changing perspectives in different conditions of the experiment and that there were no interactions with this factor.

Table 1. Viewpoint sets and their corresponding azimuth values used in Experiment 1 (Study I). P1 and P2 are respectively participant 1 and participant 2 constituting a dyad.

Viewpoint Difference Conditions	Set 1		Set 2		Set 3		Set 4	
	P1	P2	P1	P2	P1	P2	P1	P2
0°	-22.5	-22.5	22.5	22.5	-67.5	-67.5	-112.5	-112.5
45°	-22.5	22.5	22.5	-22.5	-67.5	-112.5	-112.5	-67.5
90°	-22.5	-112.5	22.5	-67.5	-67.5	22.5	-112.5	-22.5

Procedure

Participants performed a task that required locating a pointer in the upper plane of the cuboid precisely above a target placed on the lower plane of the cuboid, so that a perpendicular drawn from the pointer's apex on the upper plane would connect the apex with the center of the target on the lower plane (Figure 2A, p. 58). They were instructed to be as precise as possible in their location judgment. They were seated at an approximate distance of 60 cm from the screen throughout the experiment.

There were three different joint conditions where two participants took turns in providing locations judgments from one particular viewpoint. The three conditions varied the difference in viewpoints between the two participants: orthogonal viewpoints in the J90° condition, same viewpoints in the J0° condition. The J45° condition served as an intermediate condition (see Figure 2B, p. 58). A given pair was randomly assigned to one of the four possible sets of viewpoints (Table 1). During the trials the only aspect of the other participant's perception of the target location that was accessible to participants was the partner's pointer position seen from the individual's own viewpoint. The Individual Single condition (further: IS) served as a baseline for participants' individual task performance from a particular viewpoint. All participants performed all conditions.

The experiment was divided into three blocks. The angular difference between two participants' viewpoints was fixed in each block (e.g., J45° in the first block, J0° in the second block, and J90° in the third block). The order of conditions was counter-balanced across participants. Each block had three phases. It started with instruction and familiarization followed by the joint condition (12 trials). It ended with collecting an individual baseline (IS, 4 trials) from the same viewpoint that each individual in the pair had been exposed to in the preceding joint condition. Participants were offered a ten-minute break between blocks. In the second and third experimental block participants

received only short instructions informing them about the condition type they were going to perform next. In total, the duration of the experiment was 60-90 minutes.

The consecutive phases in each experimental block were as follows: At the beginning of the familiarization participants received detailed instructions about the task on a computer screen. Then they initiated a sequence of four individual familiarization trials pressing a joystick button. The first two trials were performed from the participant's viewpoint in the ensuing joint condition. Each trial started with displaying the pointer in its initial position in the cuboid's upper corner together with a new randomly positioned target on the cuboid's bottom; after that participants had 45 s (a warning beep sounded after 35 s) to locate the target by moving the pointer with the joystick. A judgment could be comfortably made within this time. Once participants believed to have positioned the pointer precisely above the target, they pressed the response button on the joystick and immediately received feedback about their judgment error, displayed as a straight line projecting the pointer's position on the upper level to the lower level. This feedback remained available for 5 s. After the second familiarization trial the camera smoothly rotated around the cuboid to the partner's viewpoint in the ensuing joint condition. The remaining two trials were performed from the partner's viewpoint in the ensuing joint condition. This ensured that participants had information about the perceptual reliability of the two viewpoints and their relation to each other. In the J0° block all four familiarization trials were performed from the same viewpoint and no camera rotation occurred.

The joint phase of each condition started with further instructions on the screen. Each partner was seated in a separate room with their own screen and joystick. The instruction stressed that participants would take turns in providing judgements about target location and that it was important to be as precise as possible for intermediate and final location judgments. Once both participants had indicated that they were ready they visited each other's rooms to look at each

other's setup to remind them about the two different viewpoints they had on the same virtual environment. Then the first joint trial started. It was randomly determined which participant provided the first judgment and a sound signal indicated to the pair who would start on the first trial. After the first trial, assignment of the first judgment alternated between the two partners.

In each trial in the joint condition, participants took turns to provide their judgment of the target location. Each trial comprised three cycles, so that after the first judgment was given by each of the two participants, they could adjust their judgment two more times. Altogether, a trial thus consisted of six turns where one participant provided judgments on odd-numbered turns, and the other participant provided judgments on even-numbered turns. Participants had 15 s to provide a judgment (a warning beep sounded after 10 s). Participants saw only the movements of their own pointer. The other's pointer was visible at the location which he or she indicated as his or her response, or where the pointer was registered at the time limit if the response had not been provided by that time (or in the starting position during the first turn). While active, the pointers had different colors matching the pointer color during the individual familiarization phase. While pointers were dysfunctional they were colored in grey. Only after the final judgment in a trial was provided (third cycle, sixth turn) both participants received feedback about the location error of both pointers relative to the target by drawing a straight line projection for each pointer's positioning on the upper plane to the lower plane on which the target was located (Figure 2A, p. 58).

In the final phase of each block an individual baseline was collected from the same single viewpoint participants had used in the joint condition. The course of each trial was the same as in the initial individual familiarization phase. Participants were reminded that they would perform the task alone, with one judgment per trial and a 45 s time limit corresponding to the overall time participants had for their three turns in the joint condition.

2.2.2 Data preparation

Location judgments were coded as two-dimensional coordinates in the cuboid's *x-y* plain. These coordinates were transformed into egocentric coordinates using the viewpoint azimuth angle. Accordingly, the resulting egocentric Y–axis was parallel to participants' line of sight (depth dimension), and the egocentric X-axis was parallel to the fronto-parallel plane. The error on a specified axis simply coded the distance between the target coordinate and the judgment coordinate. We excluded trials where the error on the egocentric X-axis (further Ego-X error) exceeded three standard deviations of the egocentric X-axis error averaged for all conditions. The same criterion was applied in the joint condition: If at least one of the two individuals in a pair provided an initial judgment where the error exceeded three standard deviations the whole trial was removed for both. After this step 97% of the data were preserved for the ensuing analyses.

From these data for each judgment we computed Absolute Error (A.E.), - the Euclidean distance between the judgment coordinates and the true location coordinates, translated into millimeters. We took judgments from the last cycle in the joint condition when comparisons with the IS baseline were made. Because Absolute Error confounds systematic errors (bias) and variability of judgments, we computed standard deviation of the judgments as a measure of uncertainty associated with location estimations (Gigone & Hastie, 1997).

In addition, we derived the *weight* measure to determine how much participants' own judgments were influenced by another's observed judgments in the joint condition. In joint trials we computed *serial weights* for the two consecutive opportunities provided to participants to revise their previous judgment, or iterations of revision. This process is further illustrated in Figure 3. Serial weights reflect how much weight the participant providing the odd judgments assigned to another participant's even judgment immediately preceding the revised judgment. In particular, we assessed how much the first participant's judgment in cycle 2 was affected by the second participant's

preceding judgment in cycle 1 at the first iteration of revision (w_{R1}) and how much the first participant's judgment in cycle 3 was affected by the second participant's preceding judgment in cycle 2 at the second iteration of revision (w_{R2}). In Figure 3A and 3B the w_{R1} serial weight would be 1 if J3 were equal to J2; and it would be equal to 0 if J3 were equal to J1.

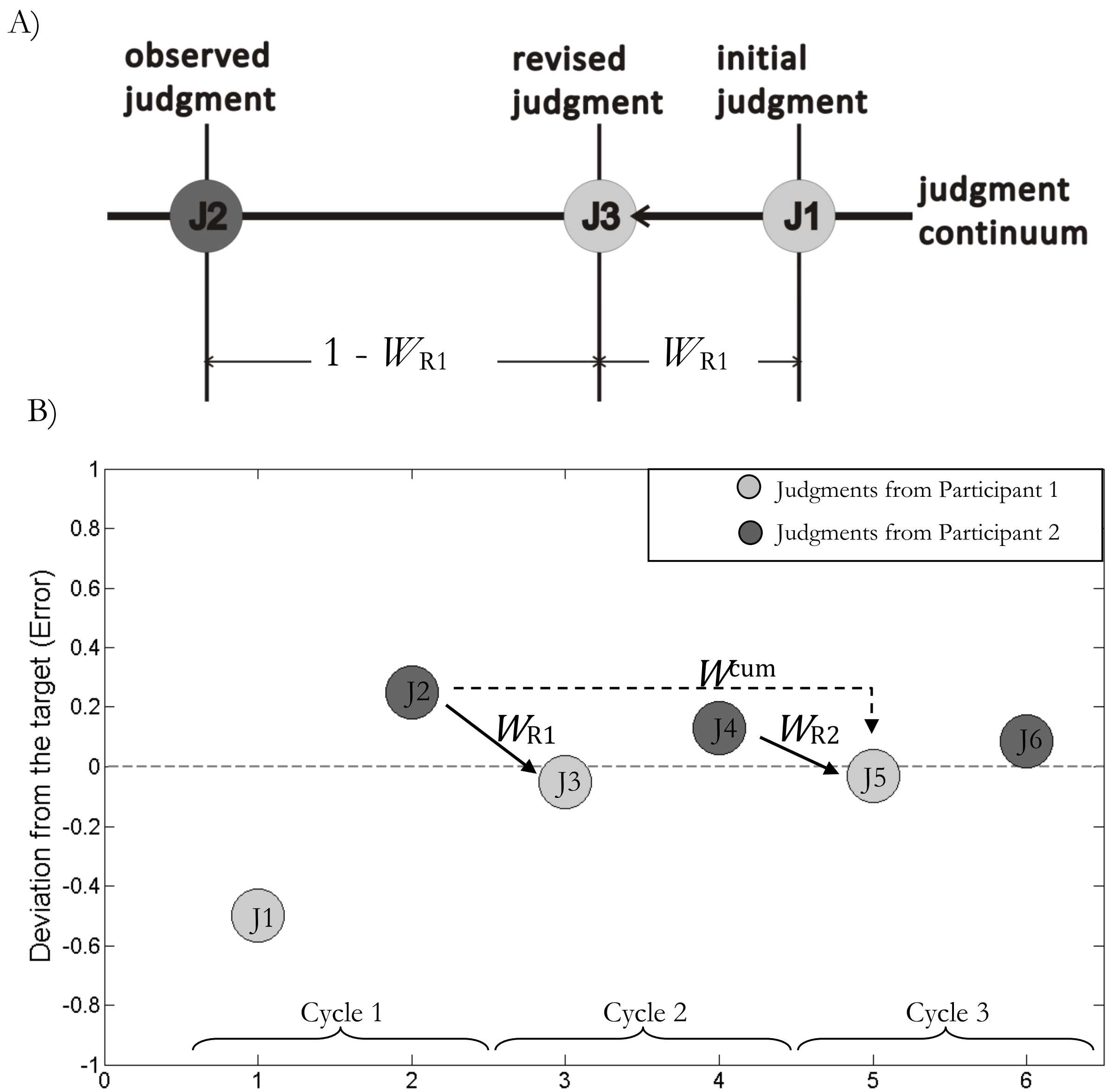


Figure 3. A schematic unidimensional representation of the course of a trial. A) Representation of judgment revision at the first iteration of revision. Participant 1 changes his judgment from location J1 to location J3 in response to the observed judgment at J2. Judgment continuum corresponds to the Y axis in Figure 3B. B) Progression of judgments in the course of the trial. The X-axis displays the judgment's serial number and roughly corresponds to the time axis. The

Y-axis displays spatial error relative to the true target location. Two participants took three turns corresponding to three judgment cycles and six judgments overall. Solid arrows depict immediate influence of the observed judgment on the revised judgment and correspond to serial weights at each iteration of revision (R1 and R2). Revised judgment becomes an initial judgment for the next iteration of revision. The dashed arrow depicts overall influence of the initial judgment of Participant 2 on the final judgment of Participant 1 (cumulative weight).

To put it more formally, serial weights were computed as a normalized shift in location from the participant's preceding judgment to the revised judgment relative to the distance between the participant's preceding judgment and the observed judgment (Harvey & Fischer, 1997). The exact formula for this calculation is:

$$w_{\mathrm{Ri}}^{\mathrm{serial}} = \frac{J_{\mathrm{odd,c}} - J_{\mathrm{odd,c-1}}}{J_{\mathrm{even,c-1}} - J_{\mathrm{odd,c-1}}}, \qquad (2.1)$$

where c is the current cycle for which a serial weight was calculated, J is judgment coordinates and $i = c - 1; 1 < c \in \mathbb{Z}$.

Additionally we calculated the cumulative weight $w^{\mathrm{cumulative}}$ as a measure of the overall influence that the initial judgment of the participant providing the even judgments had on the final judgment of the participant providing the odd judgments. In Figure 3B $w^{\mathrm{cumulative}}$ would be 1 if J5 were equal to J2; and $w^{\mathrm{cumulative}}$ would be 0 if J5 were equal to J1. Formally, cumulative weight was computed with the following formula:

$$w^{\mathrm{cumulative}} = \frac{J_{\mathrm{odd,n}} - J_{\mathrm{odd,1}}}{J_{\mathrm{even,1}} - J_{\mathrm{odd,1}}}, \qquad (2.2)$$

where n is the total number of cycles in the trial.

Both serial and cumulative weights were computed separately for the two dimensions in participant's egocentric coordinate system (Ego-Y and Ego-X dimension).

Weights were computed only from trials where a participant was the first to make a judgment. This is because the initial judgments of participants responsible for even turns were

already influenced by another's observed judgment and thus did not provide a valid baseline for an individual's first attempt to locate the target.

2.2.3 Results

The alpha-level for all statistical analyses was set to .05. All reported statistical tests are two-tailed where applicable, and unless stated otherwise. Before reporting the results for the main dependent variables we will first characterize the distributions of the raw data obtained in the present experiment. Figure 4 displays all raw judgments obtained in the experiment when individuals judged target location from a particular viewpoint in isolation. The figure illustrates that there were four clearly separate distributions with elliptical shapes corresponding to the perceptual limitations on accuracy that the four different viewpoints imposed on the participants. The different orientations of the distribution ellipses follow the azimuth value associated with a given viewpoint and thus show that different perspectives systematically influenced the pattern of uncertainty about location estimations (higher uncertainty on the depth dimension).

Using these data we tested two assumptions that are a pre-requisite for the ensuing analyses of performance in the joint condition. The first assumption is that the pattern of uncertainty can be described by two orthogonal egocentric dimensions. To test this we computed the Pearson correlation between the errors on the two dimensions obtained in the IS condition separately for each viewpoint used. The results of the test for the four perspectives are provided in Table 2. None of the correlations were significant (all ps $> .2$). Thus it seems safe to assume that the pattern of uncertainty is best captured by two orthogonal dimensions coinciding with the principle axes of the egocentric coordinate system.

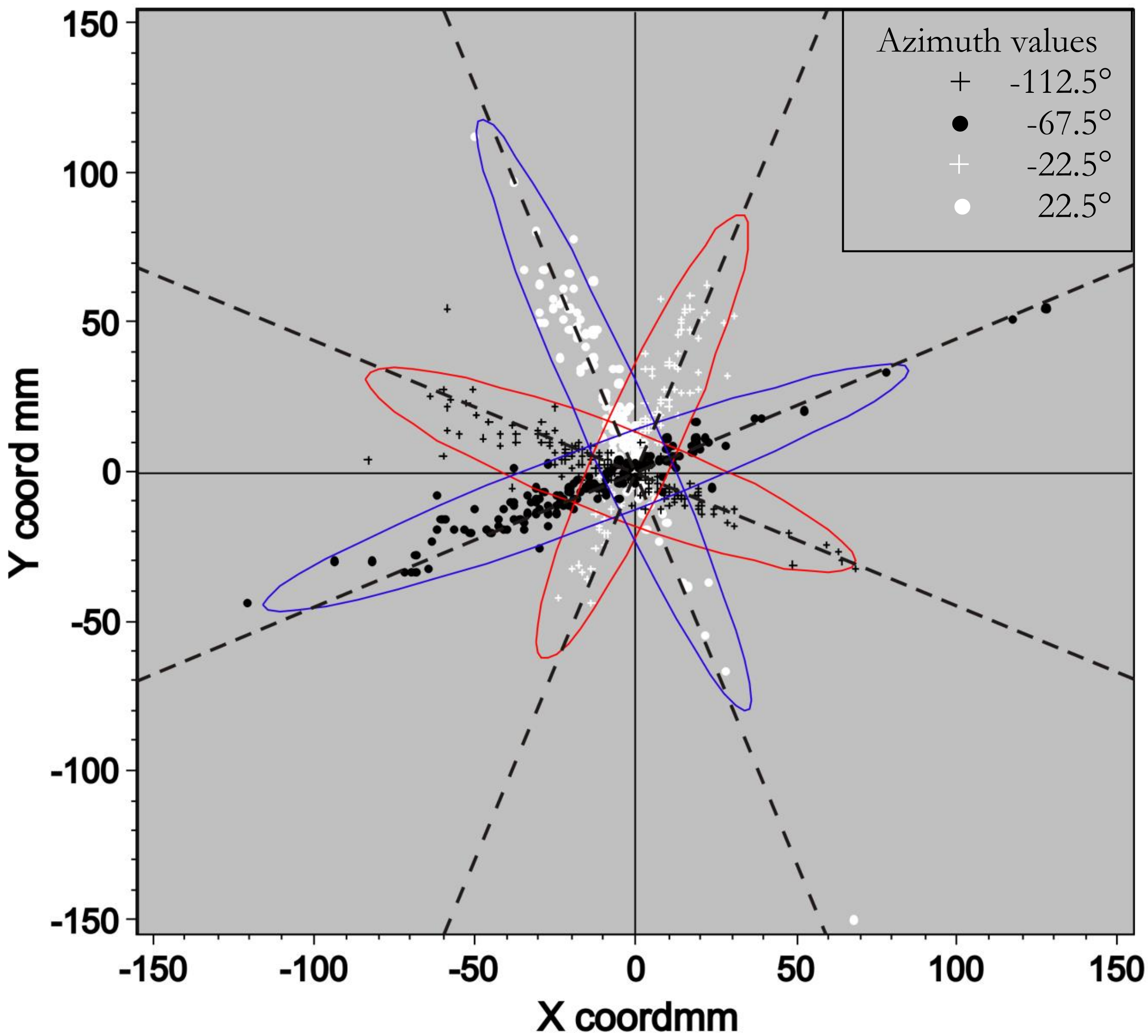


Figure 4. Raw distributions of target location judgments in the IS condition of Experiment 1 where participants provided single judgments from particular viewpoints. Coordinates are centered on the true target position. The units are in millimeters. The legend provides azimuth values corresponding to particular viewpoints. Ellipses are bivariate normal fits, where the perimeter of an ellipsoid corresponds to the 99% confidence curve. Blue and red ellipses correspond to the two different pairs of orthogonal viewpoints. Confirming the reliability of the manipulation, participants' variability was much higher on the Ego-Y (depth) dimension than on the orthogonal Ego-X dimension as indicated by ellipses' eccentricity approximately along the virtual lines of sight (dashed lines).

Table 2. Pearson correlations between signed errors on the two dimensions after rotating judgment coordinates into egocentric coordinate system using the viewpoint azimuth value.

Viewpoint Azimuth value	Pearson correlation (two-tailed)	
	Correlation *r*	Significance value *p*
-112.5°(*N*=125)	-.019	.836
-67.5°(*N*=136)	-.004	.965
-22.5°(*N*=177)	-.036	.638
22.5°(*N*=155)	.101	.212

We also tested whether it is safe to assume that uncertainty on the Ego-Y dimension is substantially larger than on the Ego-X dimension. We tested this by comparing variability of errors on the two egocentric dimensions obtained in IS baseline trials. Standard deviations of the Ego-X and Ego-Y errors were computed for each participant and then compared by means of a paired-sample *t*-test. The analysis confirmed that SDs of the Ego-Y errors (M = 21.2 mm; SD = 7.75 mm) were much larger than SDs of the Ego-X errors (M = 2.92 mm; SD = 2.24 mm), $t(23) = 13.6$, $p <$ .001 confirming that participants experienced higher uncertainty in judging target location on the depth dimension.

Absolute Error

Figure 5 shows the results for absolute error that were entered into a 2 × 3 repeated-measures ANOVA with the factors Condition (J vs IS) and Viewpoint Difference (0° vs. 45° vs. 90°). There was a significant effect of Condition, $F(1, 23) = 6.01$, $p = .022$, partial $\eta 2 = .207$. Absolute error was lower in the joint conditions (M = 18.3 mm; SD = 16.9 mm) than in the individual baseline condition (M = 21.1 mm; SD = 18.3 mm). The main effect of Perspective Difference and the interaction between Perspective Difference and Condition were not significant.

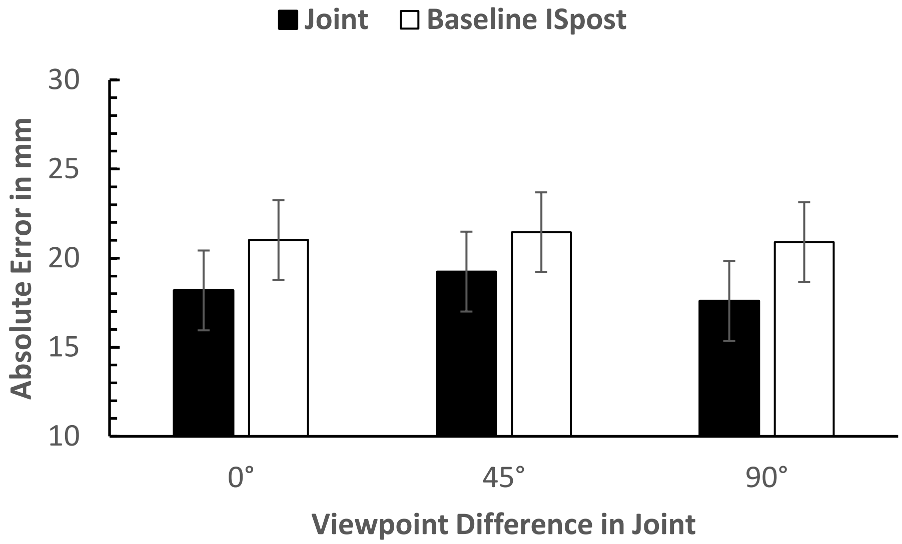


Figure 5. Results of Experiment 1 (Study I): absolute error. Absolute error in the Joint condition and ISpost baseline for blocks with different viewpoint differences in the Joint condition. Error bars represent within-subject 95% CIs from the two-way ANOVA(Loftus & Masson, 1994).

To assess whether performance improved over time in the Joint conditions we computed a 3 × 3 repeated-measurement ANOVA with the factors Viewpoint difference (0° vs. 45° vs. 90°) and Cycle (First vs. Second vs. Third). This analysis revealed a significant main effect of Cycle, $F(2, 46) = 30.7$, $p < .001$ partial $\eta 2 = .572$. The main effect of Condition and the interaction between the two factors were not significant. Our planned contrasts showed that absolute error on the second cycle ($M = 20.0$ mm; $SD = 8.35$ mm) was significantly lower than on the first cycle ($M = 25.1$ mm; $SD = 7.93$ mm), $F(1, 23) = 22.6$, $p < .001$ partial $\eta 2 = .495$, and significantly lower on the third cycle ($M = 18.3$ mm; $SD = 7.16$ mm) than on the second cycle, $F(1, 23) = 18.2$, $p < .001$ partial $\eta 2 = .441$. The parallel trajectories of error reduction over the cycles, are illustrated in Figure 6. Small gaps between the lines indicate absence of the advantage of conditions with higher information complementarity over the informationally redundant condition for the quality of revised judgments.

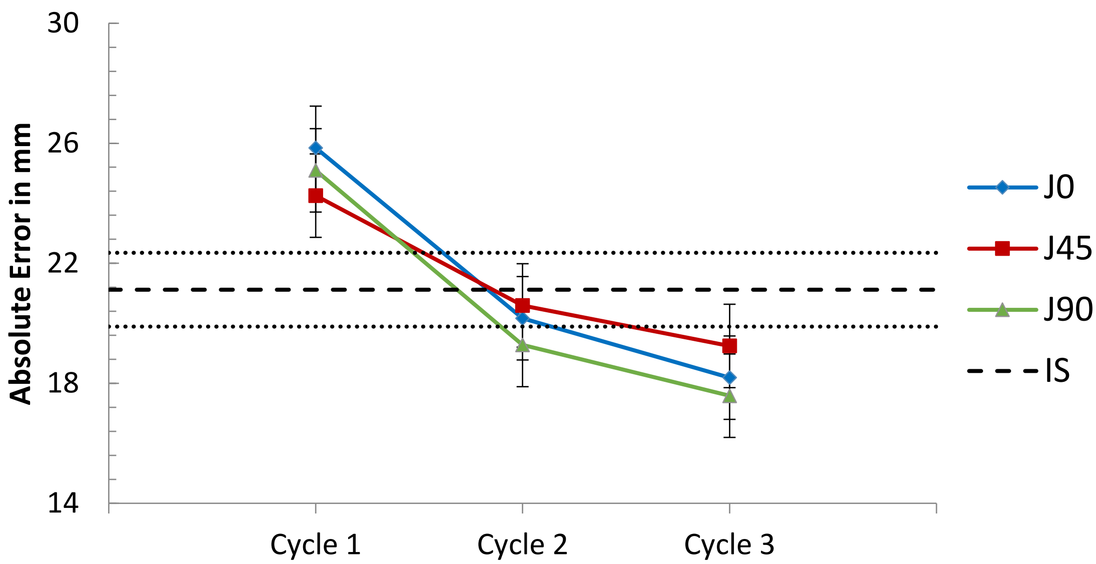
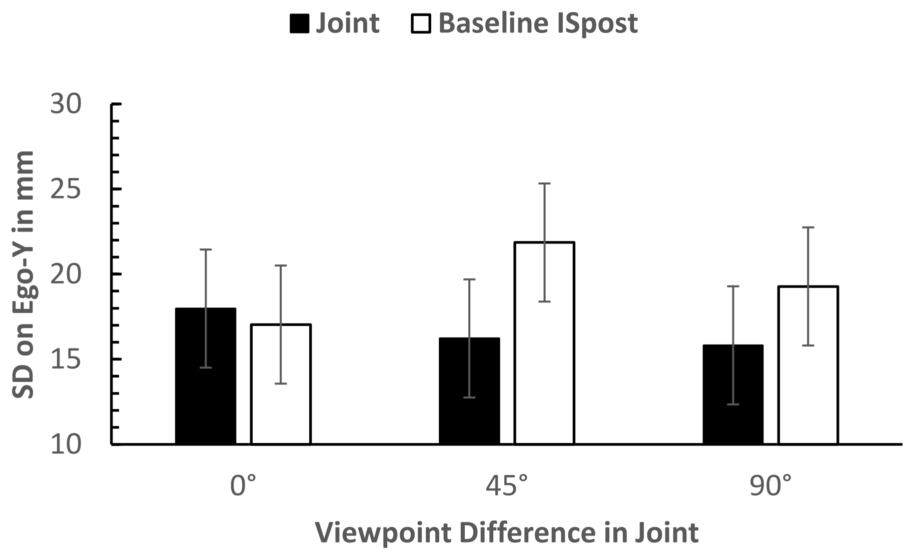


Figure 6. Results of Experiment 1 (Study I): accuracy across cycles. Colored lines stand for different viewpoint difference conditions. Error-bars stand for within-subject 95% CIs form the Condition × Cycle ANOVA (see main text). Black dashed line is average IS accuracy collapsed across three blocks of the experiment. Black dotted lines are upper and lower SE of the means from the averaged IS baseline.

Variability of location judgments

Figure 7A and 7B show the standard deviations of errors of location judgments separately for the Ego-X and Ego-Y dimension. We performed a 2 × 3 repeated-measures ANOVA with the factors Condition (J vs IS) and Viewpoint Difference (0° vs. 45° vs. 90°) for each dimension. For the Ego-Y (depth) dimension the effect of Condition did not reach significance, $F(1, 23) = 3.95$, $p = .059$ (Figure 7A), partial $\eta 2 = .146$. There was also no significant effect of Viewpoint Difference. However, there was a significant interaction between the two factors, $F(2, 46) = 3.58$, $p = .036$, partial $\eta 2 = .135$. Planned pair-wise comparisons revealed a significant difference between the Individual and Joint conditions for the J45° condition, $t(23) = 2{,}53$, $p = .019$ and the J90° condition, $t(23) = 2.21$, $p = .037$, but not for the J0° block ($p = .661$). Unexpectedly, the analysis of variability on the EgoX-axis revealed a main effect of Condition, $F(1, 23) = 14.2$, $p = .001$, partial $\eta 2 = .381$

(Figure 7B). Variability of errors was significantly higher in the joint condition than in the individual baseline. The main effect of Viewpoint Difference and the interaction were not significant.



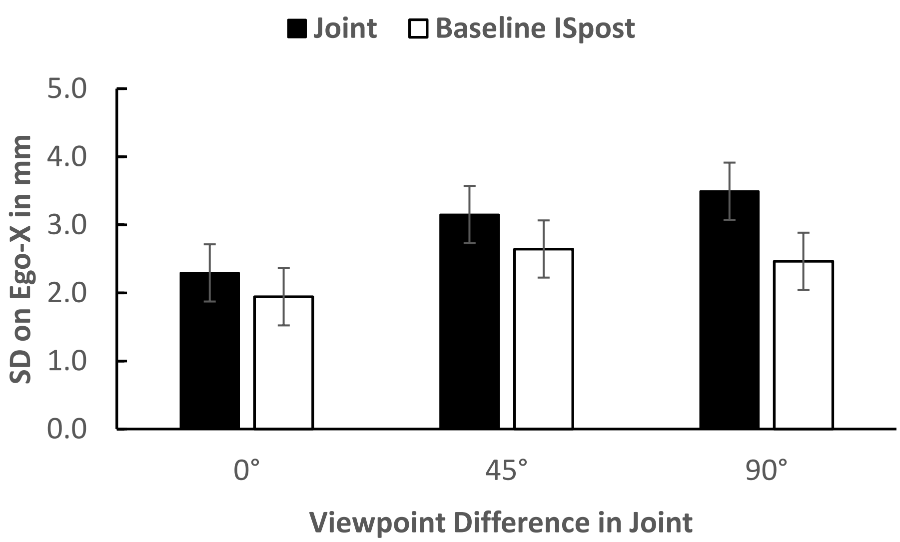


Figure 7. Results of Experiment 1 (Study I): variability of judgment errors. A) SD of errors on the Ego-Y axis. C) SD of errors on the Ego-X axis. Error bars represent within-subject 95% CIs from the two-way ANOVA (see main text).

Weight Analysis

Figure 8 on p. 74 shows the serial and cumulative weights - measures of how much influence the other's previous judgment had on a participant's current judgment from one cycle to another and how much influence the other's initial judgment had on the participant's final judgment. On this plot each data point corresponds to averaged weights on two egocentric dimensions. Ego-Y weights correspond to the value on the Y-axis of the plot, and Ego-X weights correspond to the X-axis of the plot.

The first weights analysis addressed Hypothesis III: participants were expected to provide, in general, lower weights on the dimension where they had lower uncertainty, or the Ego-X dimension, than on the dimension where they had higher uncertainty, or the Ego-Y dimension. To this end, the two dimensions were treated as the two levels of one factor – reliability of Participant's Own judgments (High vs. Low). The correspondence between the cells in the factorial ANOVA table and the actual experimental conditions are provided in Table 3.

Table 3. Correspondence of experimental conditions to the cells of the 2-by-3 ANOVA table with the factors reliability of Own judgments (High vs. Low) and reliability of the Other's judgments (High vs. Intermediate vs. Low). Each cell codes Own – Other's reliability, where H – High, I – intermediate, L – Low.

Spatial dimension in participant's egocentric coordinate system	J0	J45	J90
Ego-Y	L - L	L - I	L - H
Ego-X	H - H	H – I	H - L

The weights were then analyzed by means of a 2 × 3 repeated-measures ANOVA with factors reliability of Own judgments (High vs. Low) and reliability of the Other's judgments (High vs. Intermediate vs. Low) with w_{R1} (the ones corresponding to the first iteration of revision) serial weights as the dependent measure. This analysis revealed a main effect of reliability of Own judgments: $F(1, 23) = 4.72$, $p = .040$, partial $\eta^2 = .170$. Following the prediction, participants overall

gave larger weight to the other's judgments on the Ego-Y dimension (M = 0.44, SD = 0.30) than on the Ego-X dimension (M = 0.27; SD = 0.24) which means that participants were more influenced by the other's judgments when they had more uncertainty. The effect of reliability of the Other's judgments was also significant: $F(2, 46) = 5.13$, $p = .010$, partial $\eta^2 = .182$, - participants gave larger weight to the other's judgments when the latter were more reliable. Finally, there was a significant interaction between the two factors: $F(2, 46) = 7.98$, $p = .001$, partial $\eta^2 = .258$. This interaction can be seen in Figure 8: on the Ego-X dimension the R1 weights data points are ordered in the predicted pattern: they are closer to zero in the J90° condition relative to the J0° condition, and in the J45° condition they are in between. This reflects the fact that on this dimension from participant's perspective reliability of the other's judgments decreased in the J45° condition relative to J0° condition, and decreased in J90° condition relative to J45° condition. At the same time, on the Ego-Y dimension R1 weights data points are numerically higher in the J45° condition than in the J0 condition, but unexpectedly fall below J45° and J0° in the J90° condition. This goes against our prediction that the serial Ego-Y weights should be larger in the J90° condition than in the J45° and larger than in the J0° condition.

The same analysis was conducted with cumulative weights as the dependent measure. This analysis also revealed a significant main effect of reliability of Own judgments: $F(1, 23) = 28.8$, $p < .001$, partial $\eta^2 = .556$. Overall cumulative weights were larger on the Ego-Y dimension (M = 0.57, SD = 0.27) than on the Ego-X dimension (M = 0.21, SD = 0.30). The main effect of reliability of the Other's judgments was also significant: $F(2, 46) = 4.89$, $p = .013$ $\eta^2 = .173$. Cumulative weights were higher when the other's judgments were more reliable. However, the interaction between the two factors was not significant ($p = 0.750$). This analysis corroborates our conclusions made on the serial weights data, and adds that sensitivity to accuracy of one's own and to that of the other's

judgments is reflected not only in the immediate individual revision policy but also in the outcome of iterative reciprocal interactions.

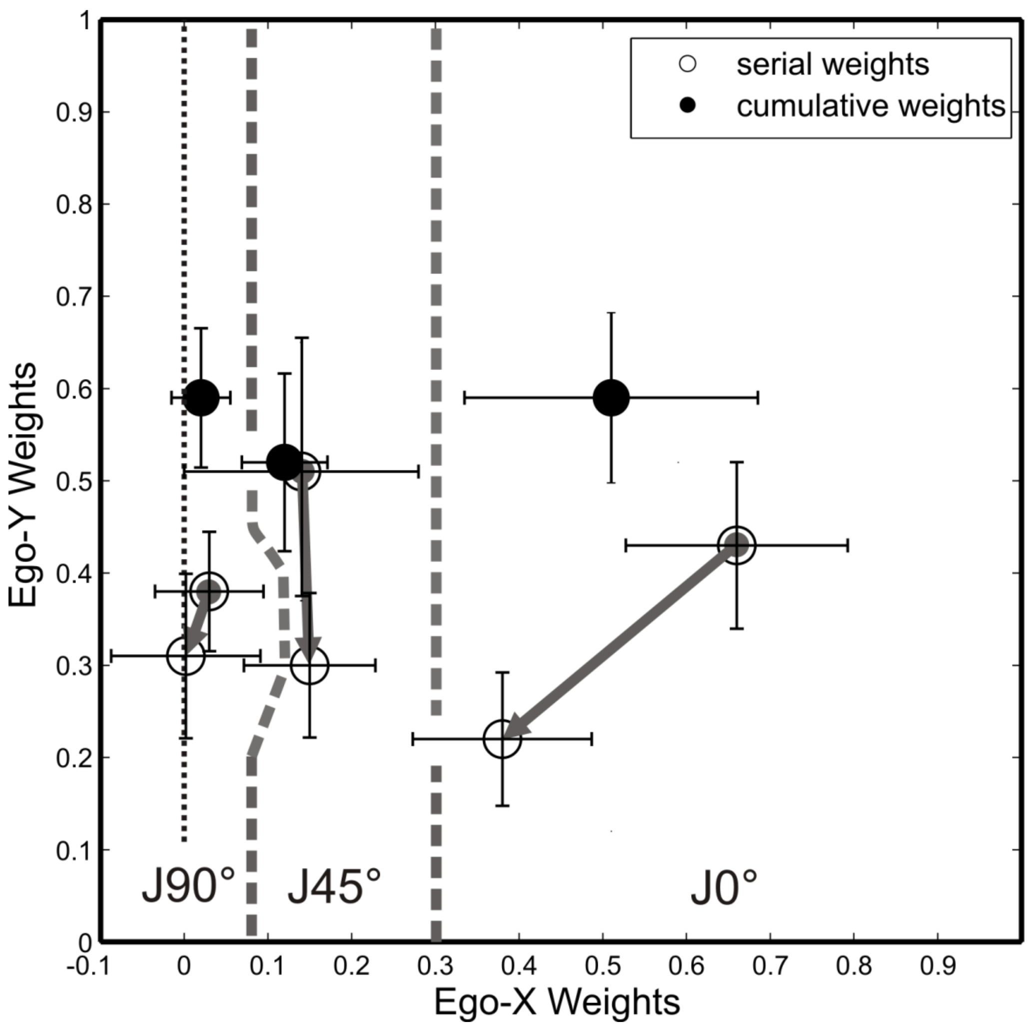


Figure 8. Weights analysis for Experiment 1 (Study I). Data points corresponding to the J0°, J45°, and J90° condition are separated by dashed lines. Serial weights indicate how much another's judgment in the previous cycle influenced the participant's judgment on the current cycle. A value of 0 means no influence, a value of 1 means that another's judgment fully determined the participant's judgment. Cumulative weights (filled circles, one per condition) characterize the influence of another's initial judgment on a participant's final judgment. Grey arrowed lines go through serial weights data points in the order of revision iteration: from Revision1 to Revision2. Data points exhibit horizontal spatial clustering by condition indicating a different weighing policy in each condition on the Ego-X dimension. Absence of clear vertical clustering by condition means that subjects' weighing policy on the Ego-Y dimension was not sensitive to manipulations of partner's reliability on this dimension. Error bars stand for SE of the means.

To get further insights into whether influence of the other's judgments changed over the trial course and to understand the difference between the strategies participants used on their accurate and inaccurate dimensions, a 2 × 3 repeated-measure ANOVA with the factors Revision Iteration (First vs. Second) and Viewpoint Difference (J0° vs. J45° vs. J90°) was conducted separately for each spatial dimension. For the Ego-Y weights there was only a main effect of Revision Iteration, $F(1, 23) = 5.35$, $p = .030$, partial $\eta 2 = .189$. They became smaller as the trial progressed, implying that less weight was given to the other's judgments on the next cycle. The analysis of the Ego-X weights revealed a main effect of Condition, $F(2, 46) = 20.7$, $p < .001$, partial $\eta 2 = .474$. The weights in the J0° condition ($M = 0.52$; $SD = 0.42$) were significantly higher than in the J45° condition ($M = 0.14$; $SD = 0.17$), $F(1, 23) = 15.3$, $p = .001$, partial $\eta 2 = .399$; and higher than in the J90° condition ($M = 0.02$; $SD = 0.20$), $F(1, 23) = 31.1$, $p < .001$, partial $\eta 2 = .575$. Thus participants weighed observed judgments nearly as much as their own ones when accuracies of the two individuals were comparable, and the given weight decreased with decreasing accuracy of the other's judgments. This contingency, however, was observed only on the dimension where participant's own judgments were highly accurate.

Cumulative weights were analyzed with one-way repeated-measures ANOVAs with the factor Condition (J0° vs. J45° vs. J90°), separately for the two dimensions. The effect of Condition was significant for the Ego-X dimension $F(2, 46) = 5.66$, $p = .006$, partial $\eta 2 = .197$, but not for the Ego-Y dimension. The cumulative Ego-X weight in the J0° condition ($M = 0.51$, $SD = 0.86$) was significantly larger than in the J90° condition ($M = 0.02$, $SD = 0.17$), $p = .012$, partial $\eta 2 = .244$, and in the J45° condition ($M = 0.12$, $SD = 0.25$). This result demonstrates that in the outcome of the interaction, the participants' judgments were more strongly influenced by the other's judgments when accuracies of the two individuals were comparable. The overall influence decreased with

decreasing accuracy of the other's judgments, but again, only when participant's own judgments were highly accurate.

2.2.4 Discussion

The results of the first experiment provided evidence that interpersonal integration of location information can improve performance when individuals can only indirectly interact by observing each other's judgments in a shared environment. Mean absolute error was significantly lower in the joint condition than in the individual baseline (Figure 5, p. 69). The error distributions derived from individual judgments (Figure 4, p. 67) confirmed that uncertainty about stimulus localization was well captured by two orthogonal dimensions. This validated our basic assumptions regarding the task to be performed and our manipulation of viewpoints and viewpoint differences (Table 1).

Although joint performance was better than individual performance, viewpoint differences between the two participants in the joint condition did not significantly alter the absolute error of location judgments. This result was unexpected because the potential benefit of integrating the other's judgment differed greatly depending on viewpoint difference. According to Hypothesis I, the potential benefit was highest in the J90° condition, lower in the J45° condition, and lowest in the J0° condition and was mainly expected to improve location judgments on a participant's dimension of low accuracy (Ego-Y). In line with this expectation, and in line with the prediction from Hypothesis II, participants' variability on their dimension of low accuracy was more reduced in the J45° condition compared to the baseline than in the J0° condition, but, contrary to our predictions, there was no further reduction of variability in the J90°condition (Figure 7A, p. 71). Variability on the high accuracy dimension was consistently higher when participants located the position of the target in interactive mode than when they located its position in isolation (Figure 7B, p. 71). Two

explanations can account for this pattern. First, participants may have relied too much on unreliable information from their partner. Alternatively, they may have ignored their partners, and then lower baseline variability could reflect just an effect of learning, since assessment of the baseline accuracy took place after an experimental condition. The analysis of weights supported the latter explanation for the J45° and J90° conditions where the Ego-X weights were marginally small. At the same time, the considerable influence of the partner's judgments in the J0° condition suggests that participants engaged in a non-efficient case-wise weighing policy that had detrimental effects on the aggregate level.

The analysis of weights given to the partner's judgment provided further insights on how participants integrated the observed judgments of another person into their own judgment. In line with the prediction from Hypothesis III, participants' weighing policy was dimension-selective: participants weighed their partner's judgment more on the dimension where they had higher uncertainty. However, we did not get full support for Hypothesis IV. Weights given to another's judgments were only partially sensitive to another's dynamically changing reliability on a particular dimension. More specifically, participants discounted another's judgments more radically when they became less precise on the Ego-X dimension, where they themselves had low uncertainty. They mostly avoided any influence by inaccurate judgments. The small increase in variability on the Ego-X dimension did not outbalance the decrease in variability on the Ego-Y dimension, leading to smaller absolute errors overall.

Ego-Y weights did not show the predicted pattern. The absence of a difference across serial weights for different viewpoint difference conditions means that participants were not sensitive to the changing variability of another's judgments on this dimension. Specifically, on the Ego-Y dimension in the J90° condition the other participant was providing judgments that were much more accurate than participant's own judgments. While participants were expected to give larger

weight to the other's judgments than to their own, the observed Ego-Y weights given to the other's judgments in the participants' revised judgments were around 0.5 implying that they gave equal weight to another's accurate judgment and to their own inaccurate judgment.

Cumulative weights indicate that over several rounds of indirect interaction, another's more accurate judgments did not have a significantly larger impact on the participants' final judgment than the participant's own less accurate initial judgments on the Ego-Y dimension. In the process of reciprocal interaction within a trial the overall influence of another's initial judgment on the participant's final judgment was somewhat increased compared to direct influence exerted by observed judgments in the process of immediate revision by the participants. However, the immediate influence of another's judgment diminished with each iteration of revision: participants were making progressively smaller adjustments to their most recent judgment. As a consequence, the overall influence of another's initial judgment on the participant's final judgment stabilized in the course of the interaction. We cannot conclude that reciprocal interactions had a significant effect on the participants' final judgments after the first revision. The accumulated influence of observing another's judgments in the process of interaction was not sufficient to produce a difference in the overall influence of another's initial judgment on the participant's final judgment across different levels of reliability of another's judgments on the dimension where participants had high perceptual uncertainty about the target location.

The weights pattern on the Ego-Y dimension could be taken to provide support for a hypothesis (Hypothesis VI) put forward by Deutsch & Gerard (1955): "If an individual perceives that a situation is objectively difficult to judge — that others as well as he experience the situation in the same way he will not trust their judgments any more than he trusts his own" (footnote 5, p. 630). Apparently in the current task, neither trial-by-trial feedback nor the opportunity to see the layout from another's viewpoint in the beginning of each joint block, provided participants with sufficient

evidence that their partner was in a better position than they themselves to judge the location of an object on their Ego-Y dimension. It seems that poor judgments are easy to identify for people only when they have low uncertainty about their own judgments. High uncertainty about the estimated quantity leads also to high uncertainty about the quality of another's judgment. An interesting question for future research is what evidence can increase individual confidence in reliable information available from others.

2.3 Experiment 2

The most unexpected outcome of Experiment 1 was that participants benefitted less than expected from location information provided by a more accurate partner, especially in the J90° condition. One possible reason for this finding is that participants encountered different viewpoint differences between their own view and their partner's view in the three different blocks of the experiment. Furthermore, the order of conditions in the fully counterbalanced design may have had unintended effects. It was shown previously (Yaniv & Kleinberger, 2000) that gaining confidence in the reliability of one's own judgments is a slow and gradual process which is characterized by high inertia. This predicts that participants will need time to recover confidence in the reliability of another's judgments after having perceived them to be unreliable on the same dimension in the previous condition. For instance, participants starting with the J0° degree condition (same viewpoint) implying a relatively low benefit from the partner's judgment may have continued to assign relatively little weight to their partner's judgment in the orthogonal J90° condition. Unfortunately, the sample size in Experiment 1 was too small to analyze such effects in more detail. Therefore, the first aim of Experiment 2 was to investigate whether pairs of participants are better able to integrate location information provided by each other's judgments in a shared environment when there is a more stable viewpoint difference between them. This question can best be asked in a

situation where participants can greatly benefit from observing each other's judgments. Therefore, we focused on the J90° condition where participants had orthogonal viewpoints on the target.

The second aim of Experiment 2 was to disentangle two possible causes for why participants did not realize the full potential of the indirect interaction and for why they failed to assign appropriate weights to observed judgments. This could be due to a specific problem with integrating information from a partner because the reliability of this information cannot be assessed. Alternatively, it could be due to a more general failure to integrate location information from two viewpoints to form an integrated location judgment (Avraamides, Adamou, Galati, & Kelly, 2012). To test these two alternatives, we introduced an additional individual double (ID) condition where individual participants had full sequential access to both orthogonal viewpoints onto the target, and thus could integrate the judgments using all task-relevant information from the two viewpoints intra-individually. The ID90° condition was the same as the J90° condition but participants provided judgements from both viewpoints rather than only from one viewpoint. If accuracy was higher in the ID90° condition than in the J90° condition, this would favor the explanation that participants had problems assessing the reliability of information provided by another individual.

Finally, the results of Experiment 1 showed that in the joint condition location accuracy still improved from the second to the third judgment cycle. Therefore, we added a fourth judgment cycle to both the J90° and ID90° condition to further increase the chances that the last judgment in a trial would reflect the maximum accuracy with which participants could locate a given target.

2.3.1 Method

Participants

Thirty-two students (20 females, 12 males) aged between 19 and 26 years were tested in pairs. There were 7 female, 3 male, and 6 mixed-gender pairs. One additional pair was replaced because of an experimenter error.

Stimuli and Apparatus

The experimental stimuli and the set-up were identical to Experiment 1. We used two orthogonal sets of viewpoints: with -70°/20° and -110°/-20° azimuth values.

Procedure

The task and procedure were the same as in Experiment 1 with the following exceptions. The angular difference between viewpoints was kept constant at 90° throughout the experiment. An Individual Double condition (ID90°) was added. In this condition individual participants performed both parts of the same judgment task that was used in the J90° condition. They had two pointers, each controlled from one particular viewpoint. The two pointers were always visible, and the position of one pointer could be seen while positioning the other pointer. In the course of an ID trial participants switched back and forth between the two perspectives for each turn. Thus in the ID90° condition individuals performed both parts of the task they performed with another person in the J90° condition.

The experimental session was divided into two blocks. Before the start of each block four trials familiarized participants with the two particular viewpoints used in the ensuing block. Each block started and ended with 12 IS trials (IS-pre and IS-post) where individual participants provided location judgments from one viewpoint. The 12 trials in the middle of the block always involved judgments from two viewpoints, either by one individual (ID90°) or by two different individuals (J90°). There was a 10 minute break between the two blocks. The order of J90° and ID90° sessions

and the viewpoint set used in each condition was counter-balanced across participants. The course of each trial was the same as in Experiment 1 with one exception: a fourth cycle was added to the J90° condition and to the ID90° condition. This resulted in longer decision times. To equate decision times in the IS condition with the other conditions, we increased the time limit for IS trials to 60 s even though in Experiment 1 participants had responded much faster on average.

2.3.2 Data preparation

For the individual single and the joint condition data preparation was the same as in Experiment 1. In the new ID condition all error measures were computed as an average of two pointers' errors for a given cycle, thus on each cycle an individual contributed judgments from two turns (i.e. 7th and 8th turn on the fourth cycle). Note, that this is comparable to the J condition, because in the J condition individuals contributed to their grand average performance with an equal number of odd and even turns. Excluding trials where at least one participant had produced a judgment that was further than 3SD away from the mean accuracy for either dimension removed 165 trials in total so that 97.5% of the data were preserved for the ensuing analysis.

2.3.3 Results

Initial analyses showed that there were no main effects or interactions depending on the different viewpoint sets used. Therefore, we collapsed the data across viewpoint sets.

Absolute error

We compared absolute error between single viewpoint and dual viewpoint trials. For the ID90° condition a one-way repeated-measures ANOVA with the factor Condition (IS-pre vs. ID90° vs. IS-post), resulted in a significant main effect of Condition, $F(2, 62) = 20.2$, $p < .001$, partial $\eta 2 =$

.395 (see Figure 9A). A planned contrast revealed that absolute error in the ID90° condition (M = 16.6 mm; SD = 16.0 mm) was significantly lower than in the IS-post condition (M = 25.1 mm; SD = 13.6 mm), $F(1, 31) = 36.6$, $p < .001$, partial $\eta2 = .542$. The difference between IS-pre and IS-post was not significant ($p = .474$).

The same analyses were conducted for the J90° condition. There was a significant effect of Condition, $F(2, 62) = 13.9$, $p < .001$, partial $\eta2 = .310$ (see Figure 9A). A planned contrast revealed that absolute error in the J90° condition (M = 15.3 mm; SD = 15.5 mm) was significantly lower than in the IS-post condition (M = 22.8 mm; SD = 14.0 mm), $F(1, 31) = 49.9$, $p < .001$, partial $\eta2 = .617$. The difference between IS-pre and IS-post was not significant ($p = .444$). A repeated-sample t-test showed no significant difference in absolute error between the ID90° and J90° conditions, $p = .552$.

To assess whether overall accuracy improved across cycles we computed a 4 × 2 repeated-measures ANOVA with the factors Cycle (First vs. Second vs. Third vs. Fourth) and Condition (ID vs. J). There was a significant main effect of Cycle, $F(3, 93) = 42.1$, $p < .001$, partial $\eta2 = .576$. Planned contrasts showed that participants were more accurate on the second cycle (M = 17.4 mm, SD = 14.9 mm) than on the first cycle (M = 23.0 mm, SD = 14.1 mm), $F(1, 31) = 47.4$, $p < .001$, partial $\eta2 = .604$, and more accurate on the third cycle (M = 16.3 mm, SD = 15.3 mm) than on the second cycle, $F(1, 31) = 8.95$, $p = .005$, partial $\eta2 = .224$. The difference between the fourth (M = 16.0 mm, SD = 14.5 mm) and the third cycle, the main effect of Condition, and the interaction between Cycle and Condition were not significant.

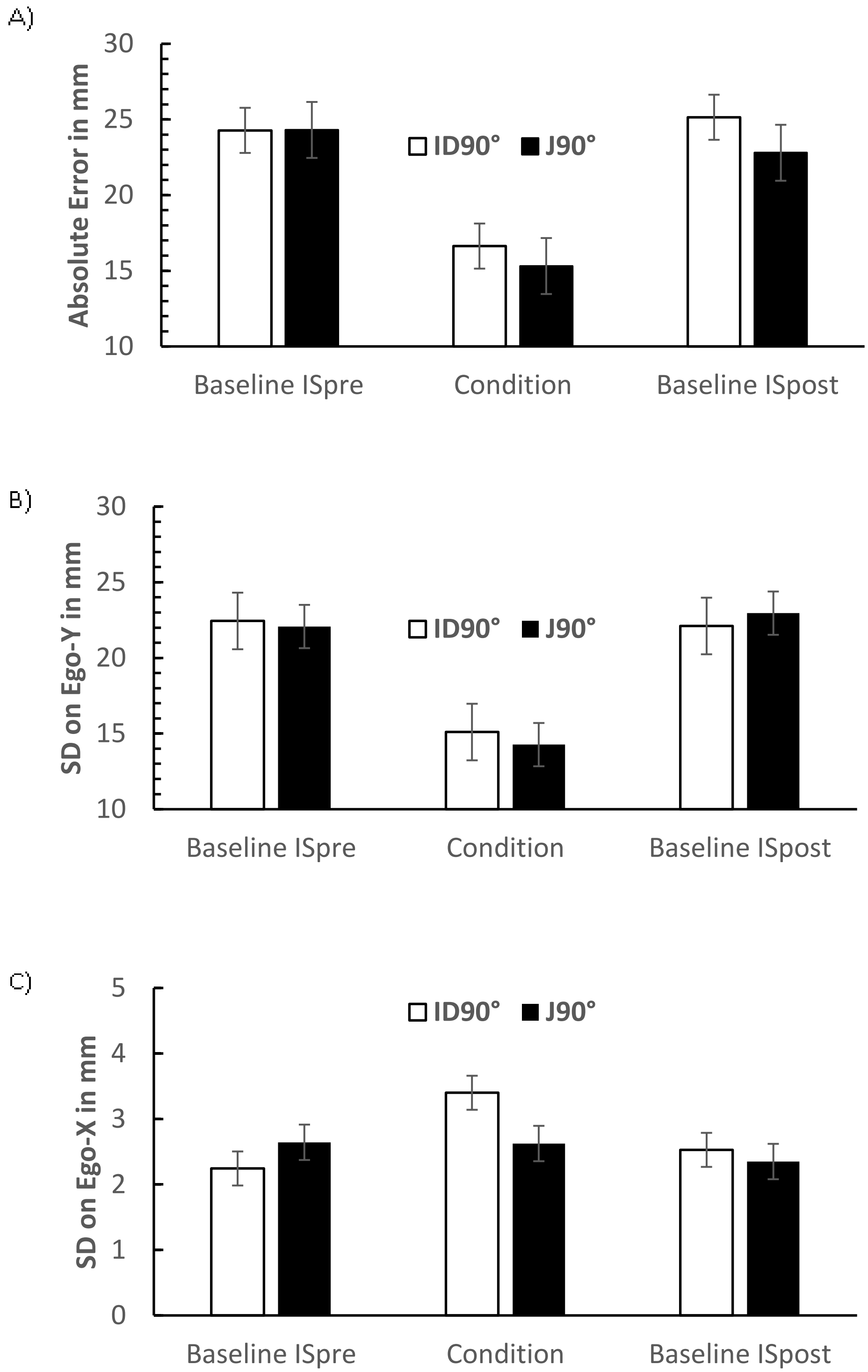


Figure 9. Results of Experiment 2 (Study I). A) Absolute Error (mm) in the experimental conditions ID90° and J90° and in the corresponding IS-pre and IS-post baselines. B) Variability of judgment errors on the Ego-Y axis. C) Variability of judgment errors on the Ego-X axis. Error bars stand for within-subject 95% CIs from the corresponding one-way ANOVAs (see main text).

Variability of location judgments

We analysed variability of errors in individual judgments separately for the Ego-Y and Ego-X dimension and separately for the ID90° and the J90° condition. We start with ID90°. The standard deviations of errors on the Ego-Y dimension were entered into a one-way repeated-measures ANOVA with the factor Condition (IS-pre vs. ID vs. IS-post). There was a significant main effect, $F(2, 62) = 10.1$, $p < .001$, partial $\eta 2 = .246$ (see Figure 9B). Planned contrasts revealed that variability of errors in the ID condition ($M = 15.1$ mm; $SD = 12.6$ mm) was significantly lower than in the IS-post condition ($M = 25.1$ mm; $SD = 9.30$ mm), $F(1, 31) = 14.6$, $p = .001$, partial $\eta 2 = .320$. The difference between IS-pre and IS-post was not significant ($p = .805$). The same ANOVA for the Ego-X dimension revealed a significant effect of Condition, $F(2, 62) = 10.8$, $p < .001$, partial $\eta 2 = .265$ (Figure 9C). Variability of participants' judgment errors was higher in the ID condition than in the IS conditions. Planned comparisons showed that variability was higher in the ID condition ($M = 3.40$ mm; $SD = 1.58$ mm) than in the IS-post condition ($M = 2.53$ mm; $SD = 1.40$ mm), $F(1, 31) = 8.20$, $p = .007$, partial $\eta 2 = .209$. The difference between IS-pre and IS-post was not significant ($p = .128$).

The same ANOVAs were conducted for the J90° condition. For the Ego-Y dimension there was a significant effect of Condition $F(2, 62) = 22.9$, $p < .001$, partial $\eta 2 = .425$ (see Figure 9B). Planned contrasts showed that variability of participants' judgment errors was significantly lower in the J90° condition ($M = 14.3$ mm; $SD = 9.24$ mm) than in the IS-post condition ($M = 22.0$ mm; $SD = 9.44$ mm), $F(1, 31) = 48.8$, $p < .001$, partial $\eta 2 = .611$. The difference between IS-pre and IS-post was not significant ($p = .525$). For the Ego-X dimension, there was no significant difference between the conditions (see Figure 9C).

Paired-samples t-tests were used to analyse whether standard deviations of errors of location judgments on the Ego-Y and Ego-X dimensions differed between the ID90° condition and the J90°

condition. For the Ego-Y dimension the difference between the two conditions was not significant ($p = .661$). For the Ego-X the difference was significant, $t(31) = -2.11$, $p = .043$: standard deviations of errors were smaller in the J90° condition ($M = 2.62$ mm; $SD = 1.76$ mm) than in the ID90° condition.

Weights analysis

Serial weights were analyzed by means of a 2 × 3 repeated-measures ANOVA with the factors Condition (J90° vs. ID90°) and Revision Iteration (First vs. Second vs. Third). For the Ego-Y dimension there was a significant main effect of Revision Iteration, $F(2, 62) = 20.8$, $p < .001$, partial $\eta 2 = .402$, but no significant main effect of Condition ($p = .659$) or interaction between the two factors ($p = .745$). Figure 12A on p. 94 illustrates the drop in Ego-Y weights in both conditions as the trial went on. For the Ego-X dimension there was a significant effect of Condition, $F(1, 31) = 5.14$, $p < .001$, partial $\eta 2 = .142$, the weights being smaller in the ID condition than in the J90° condition. There was no significant effect of Revision Iteration ($p = .054$) and no interaction between the two factors ($p = .247$). Cumulative Ego-X weights were also smaller in the ID90° condition ($M = 0.02$, $SD = 0.14$) than in the J90° condition ($M = 0.18$, $SD = 0.22$), $t(31) = - 3.32$, $p = .002$. There was no difference between Ego-Y cumulative weights ($p = .742$).

2.3.4 Discussion

The second experiment compared how viewing a target from orthogonal viewpoints helped individuals (ID90° condition) and pairs (J90° condition) to locate its position compared to individually viewing the target from a single viewpoint (IS condition). Interestingly, there was no additional benefit when individual participants had full sequential access to both viewpoints in the ID90° condition suggesting that intrapersonal integration was at least as effective as interpersonal

integration. This conclusion was further supported by the finding that variability of location judgments on the low accuracy dimension (Ego-Y) was reduced by the same amount in the ID90° and the J90° condition compared to the IS condition.

Other than in Experiment 1, variability of participants' judgments on the high-accuracy dimension (Ego-X) in the J90° condition was not higher than in the IS condition. This indicates that participants better succeeded in ignoring unreliable information provided by another individual and that the observed differences in Experiment 1 may have been due to the need to handle different configurations of viewpoints. Surprisingly, variability on the high-accuracy dimension (Ego-X) was significantly higher in the ID90° condition than in the J90° condition and then in the IS condition. One potential explanation for this unexpected finding is that individuals switching between two perspectives in the ID90° condition were not able to maintain a viewpoint-neutral spatial representation. This may have created a need for recalibration of viewpoints after each turn that was not required in the J90° condition.

The analysis of cumulative weights revealed that on the dimension of low accuracy (Ego-Y) participants' final judgments were almost equally influenced by their own and another's initial judgments (J90° condition), and by their own judgments provided from a more appropriate viewpoint (ID90° condition). Both were close to 0.7, which is numerically larger than weights people give on average to external judgments in most JAS tasks (Bonaccio & Dalal, 2006). This result suggests that in the course of indirect interaction with multiple rounds of judgment and revision an individual judgment can be asymptotically attracted to a state where it is closer to another's more accurate judgment than to the individual's initial judgment.

In line with the results from Experiment 1 serial weights dropped across consecutive iterations of revision. Thus information available in another's and participants' own revised judgments was increasingly ignored in later phases of the trial.

As expected, cumulative weights on the dimension of high accuracy (Ego-X) were substantially lower than on the dimension of low accuracy (Ego-Y). This replicates the findings of Experiment 1 where participants were also more successful in rejecting low-quality information from another individual than incorporating high-quality information. Unexpectedly, Ego-X weights were significantly higher in the J90° condition than in the ID90° condition where they were very close to 0 implying that participants ignored their own previous judgments but gave more weight to another individual's judgments.

2.4 Experiment 3

The third experiment was conducted to further investigate whether interpersonal integration of location information may have benefits compared to intrapersonal integration. In both the ID90° and J90° conditions of Experiment 2 participants could have gained accuracy benefit merely from the opportunity to revise their judgments (Vul & Pashler, 2008). We attempted to rule out this explanation comparing individual and joint performance with the same reciprocal judgment-revision-judgment procedure using a viewpoint difference of 0° implying that neither in the J0° condition nor in the ID0° condition a change in viewpoint took place. Furthermore, Experiment 3 aimed to replicate the finding obtained in Experiment 1 that increased weight is given to another individual's judgment even on the dimension of high accuracy when it is shared by both participants as is the case in the J0° condition. This would provide further evidence that effective weighing of information provided by another person can take place for the more accurate Ego-X dimension.

2.4.1 Method

Participants

Twenty-four paid students (20 females, 4 males) aged between 18 and 26 years (M = 21.5) were tested in same-gender pairs. One pair was replaced because of a program error, and one pair was replaced because of its poor performance on the task (3 SD worse than the average in at least one experimental condition).

Stimuli and Apparatus

The experimental stimuli and set-up were identical to Experiment 1 and 2. However, we only used two different viewpoints with azimuth values of -110° and -70°.

Procedure and Design

The procedure was identical to Experiment 2. In the J0° condition two participants were asked to make location judgments from the same viewpoint using one pointer each and in the ID0° condition one participant was asked to make two judgments from the same viewpoint using two different pointers. The order of JP and ID blocks and the viewpoints used in each block were counter-balanced across participants.

2.4.2 Results

Absolute error

A one-way repeated-measures ANOVA with the factor Condition (IS-pre vs. J0° vs. IS-post) showed a significant effect, $F(2, 46) = 13.1$, $p < .001$, partial $\eta 2 = .363$ (see Figure 10A, p. 91). The crucial contrast between the J0° (M = 11.7 mm, SD = 3.88 mm) and the IS-post (M = 13.4 mm, SD = 5.85 mm) condition was also significant, $F(1, 23) = 4.80$, $p = .039$, partial $\eta 2 = .173$

demonstrating that two participants looking from one viewpoint were more accurate than one participant looking from one viewpoint. A similar analysis for the ID0° condition (ID-pre vs. ID0° vs. ID-post) also revealed a significant effect of condition, $F(2, 46) = 8.47$, $p < .001$, partial $\eta 2$ = .269. The crucial contrast between the ID0° and IS-post condition was not significant though ($p > .5$). This result along with the data pattern (see Figure 10A) indicates that the difference in this condition is simply due to increasing familiarity with the task. A comparison of mean accuracy in the fourth cycle of the J0° and ID0° condition with a paired-sampled t-test confirmed that the absolute error in the J0° condition was significantly lower than in the ID0° condition (M = 14.8 mm, SD = 5.72 mm), $t(23) = -2.62$, $p = .015$.

A 2 × 3 repeated-measurement ANOVA with the factors Condition (ID0° vs. J0°) and Cycle (first vs. second vs. third) showed a significant main effect of Condition, $F(1, 23) = 5.30$, $p = .031$, partial $\eta 2$ = .187 and Cycle, $F(3, 69) = 16.3$, $p < .001$, partial $\eta 2$ = . 415. Absolute error was lower in the J0° (M = 13.2 mm, SD = 4.21 mm) condition than in the ID0° (M = 16.3 mm, SD = 6.53 mm) condition. It was also generally lower on the second cycle (M = 14.6 mm, SD = 4.70 mm) than on the first cycle (M = 17.7 mm, SD = 6.32 mm), $F(1, 23) = 20.1$, $p < .001$, partial $\eta 2$ = .467, and was lower on the third cycle (M = 13.6 mm, SD = 4.17 mm) than on the second cycle, $F(1, 23) = 4.68$, $p = .041$, partial $\eta 2$ = .169. The difference between the third and the fourth cycle was not significant. There was no interaction between the two factors.

Figure 11B on p. 93 shows how absolute error changed over the trial cycles in the J and ID conditions, and allows us to make a comparison between Experiment 2 and Experiment 3. In Experiment 2 (90° difference) the error drops below IS averages (dashed lines) from the third cycle both in the ID condition and in the J condition. In Experiment 3 (0° difference) the error was significantly reduced only in the J condition.

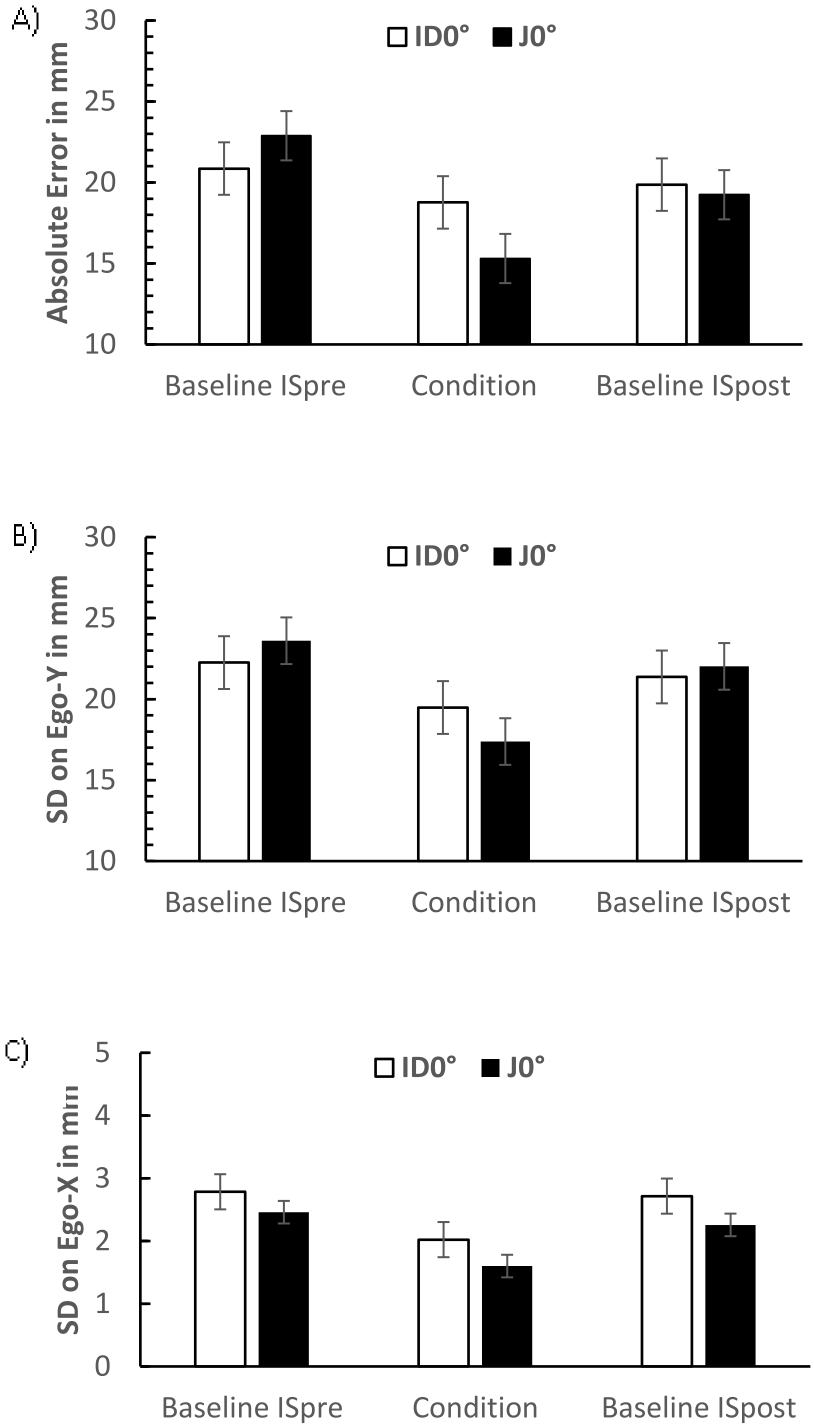


Figure 10. Results of Experiment 3 (Study I). A) Absolute Error (mm) in the experimental conditions ID0° and J0° and in the corresponding IS-pre and IS-post baselines. B) Variability of errors on the Ego-Y axis. C) Variability of errors on the Ego-X axis. Error bars stand for within-subject 95% CIs from the corresponding one-way ANOVAs (see main text).

Variability of location judgments

We first analysed variability of errors in location judgments for the J0° condition by means of a one-way repeated-measures ANOVA with the factor Condition (IS-pre vs. J0° vs. IS-post). For the Ego-Y dimension there was a significant effect of Condition, $F(2, 46) = 14.0$, $p < .001$, partial $\eta 2 = .394$. However, the main target contrast between the J0° and IS-post condition was not significant ($p = .086$) and the main effect can be attributed to increasing familiarity with the task (see Figure 10B). For the Ego-X dimension the effect of Condition was also significant, $F(2, 46) = 9.83$, $p < .001$, partial $\eta 2 = .299$. The difference between the J0° condition ($M = 1.32$ mm, $SD = 0.51$ mm) and the IS-post condition ($M = 1.96$ mm, $SD = 1.13$ mm) was significant, $F(1, 23) = 9.92$, $p = .014$, partial $\eta 2 = .234$ (see Figure 10C).

The same one-way ANOVA conducted for the ID0° condition did not show a significant main effect neither for variability of errors on the Ego-Y dimension nor for variability of errors on the Ego-X dimension. Variability of judgment errors in the J0° condition was significantly lower than in the ID0° condition both on the Ego-X dimension, $t(23) = -3.70$, $p = .001$ and on the Ego-Y dimension $t(23) = -2.40$, $p = .025$.

Weights analysis

Serial weights were analyzed with a 2 × 3 repeated-measures ANOVA with the factors Condition (J0° vs. ID0°) and Revision Iteration (First vs. Second vs. Third). For the Ego-Y dimension the main effect of Revision Iteration was significant, $F(2, 46) = 14.1$, $p < .001$, partial $\eta 2 = .379$. There was no main effect of Condition ($p = .329$) or interaction ($p = .156$). There was a similar pattern for the Ego-X weights: a significant main effect of Revision Iteration $F(2, 46) = 5.07$, $p = .010$, partial $\eta 2 = .181$ and no effect of Condition ($p = .564$), and no interaction ($p = .98$). This decreasing pattern of weights (see Figure 12B, p. 94) indicates that with each iteration of revision

participants made progressively smaller adjustments to their most recent judgment. This seems to be a robust pattern observed in all three conducted experiments.

Cumulative Ego-X weights were generally lower in the J0° (*M* = 0.35, *SD* = 0.56) condition than in the ID0° condition (*M* = 0.63, *SD* = 0.55), $t(23) = -2.32$, $p = .030$. This means that participants' own judgments converged more closely to their own judgments made from the same viewpoint than to another individual's judgments.

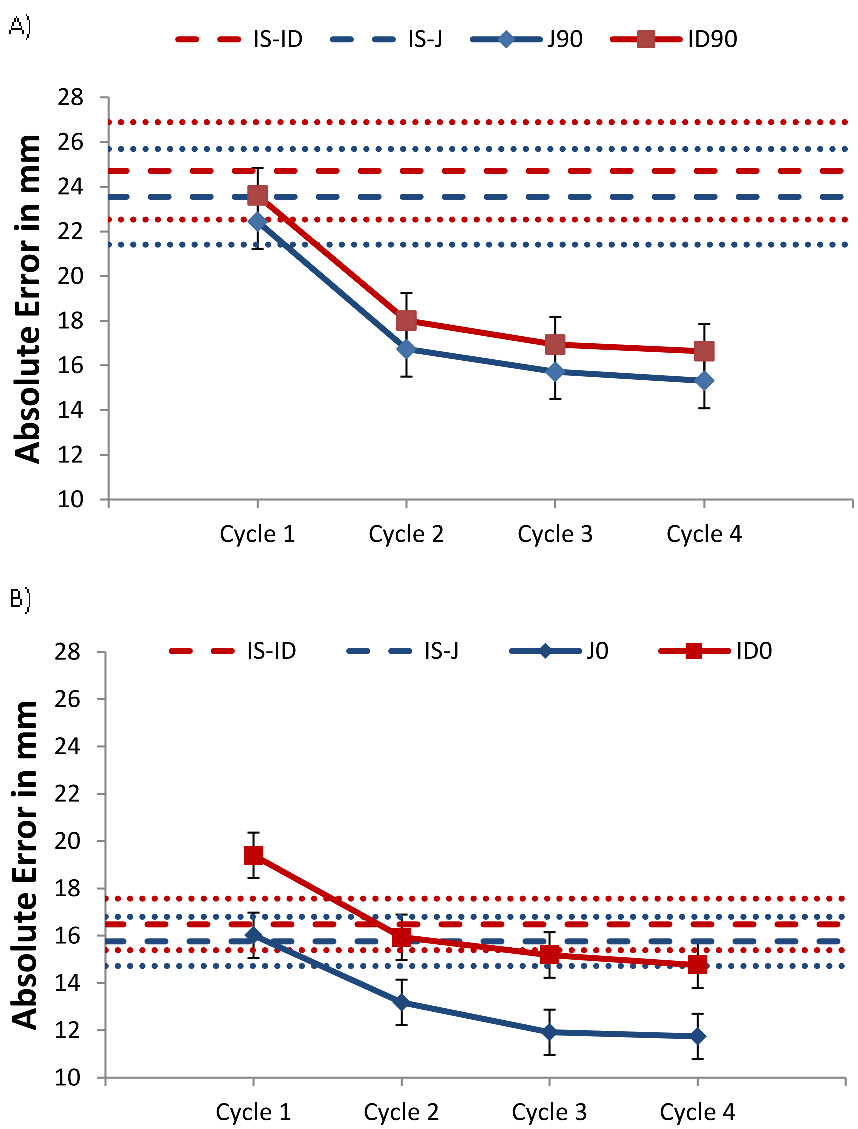


Figure 11. Absolute error across trial cycles: Experiment 2 and Experiment 3 (Study I). Error-bars stand for within-subject 95% CIs from the Condition × Cycle ANOVA (see main text). Dotted lines stand for the upper and lower SE of the means from the corresponding (color-coded) IS baseline.

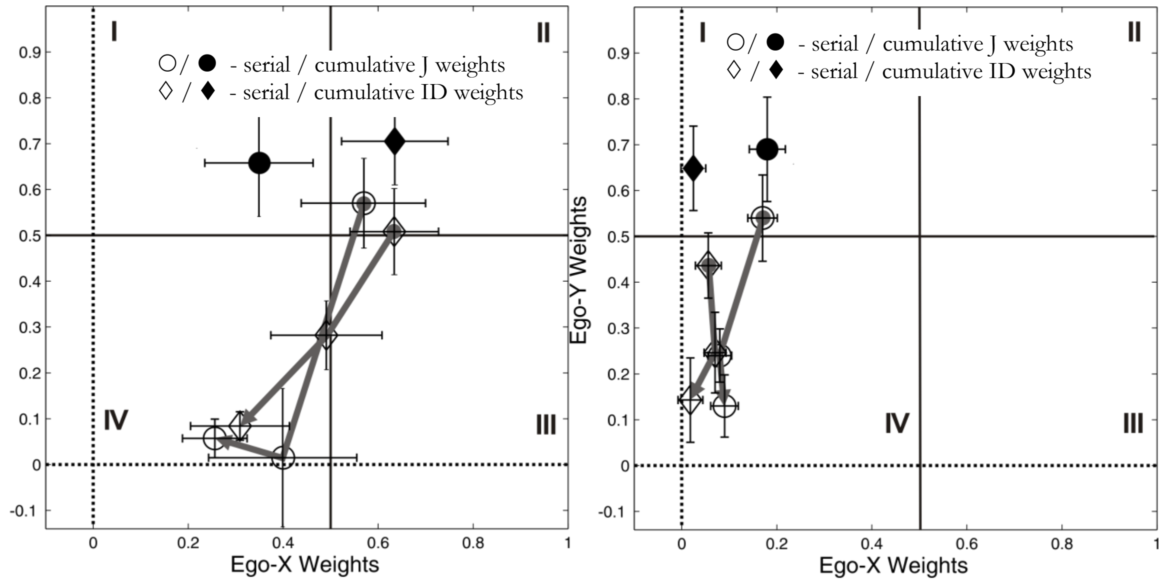


Figure 12. Weights analysis for Experiment 2 and Experiment 3 (Study I). A) Data from Experiment 2; B) Data from Experiment 3. Empty circles refer to serial weights; filled circles refer to the cumulative weights Diamonds code the same weight types for the ID condition. Grey arrowed lines go through serial weights data points in the order of revision iteration: from Revision1 through Revision2 to Revision3. Black solid lines intersect at the .5 value and divide the graph into four quadrants. Data points in quadrants I and II indicate that revised judgments were more influenced by observed judgments than by initial perception on the Ego-Y dimension. Data points in quadrants II and III indicate that revised judgments were more influenced by observed judgments than by initial perception on the Ego-X dimension. Error bars stand for the SE of the means.

2.4.3 Discussion

The results of Experiment 3 demonstrated that two individuals looking from the same viewpoint in the J condition were more accurate in localizing a target than when localizing the target in isolation. Surprisingly, reduction of variability on the difficult Ego-Y dimension was not statistically significant. Variability was significantly reduced for the easy Ego-X dimension, and this was also reflected in the significantly reduced absolute error rate. The benefit was not present in the ID condition where one individual had twice the number of opportunities to revise her own judgments provided from one viewpoint. Also, participants had smaller absolute error and lower variability of errors in the Joint condition than in the ID condition. This result rules out an alternative explanation that the benefit of incorporating another's judgments found in Experiments

1 and 2 can be attributed merely to the opportunity to revise one's own judgments (Vul & Pashler, 2008). We retain the conclusion that this benefit comes from integrating another's judgments.

Cumulative and serial weights on the dimension of high accuracy (Ego-X) had the same pattern as in the J0° condition of Experiment 1. Furthermore, cumulative and serial weights on this dimension in Experiment 2 had the same pattern as in the J90° condition of Experiment 1. Together, results from Experiment 2 and Experiment 3 replicate our initial findings: on the dimension where their own judgments were highly accurate participants gave near equal weight to another's judgments when those were as reliable as their own because the dimension of high accuracy was shared; and they gave less weight to another's judgments when those were, respectively, less reliable because for another individual this was the dimension of low accuracy. On the dimension of low accuracy (Ego-Y) participants trusted another's judgments approximately as much as their own, regardless of the other's accuracy.

Numerically, cumulative and serial weights were not exactly the same (see Figure 12B). This difference, however, was not significant. We therefore cannot conclude that reciprocal interactions had any significant effect on the participant's final judgment after the first judgment revision. Additional influence exerted by another's revised judgments possibly was not naught, but it was not substantive relative to the immediate influence of another's initial judgment.

In line with one part of our Hypothesis IV, both Ego-X and Ego-Y cumulative weights were close to 0.5 value. The same was true for serial weights corresponding to the first iteration of judgment revision. We can conclude that when two individuals are comparable in accuracy of their judgments, they would trust each other's judgments nearly as much as their own.

An unexpected result was that on their Ego-X dimension participants' own second judgment in the ID condition had higher influence on their final judgment than another's judgment in the J condition. A likely interpretation for this result is that a proportion of participants in the ID

condition only on the second turn provided the location judgment which was close to their perceived correct location and then adjusted their final judgment to this location.

2.5 General Discussion of Study I

The first aim of the present study was to investigate interpersonal integration of multidimensional perceptual information by means of indirect interactions (Keil & Goldin, 2006). Individuals were performing an object localization task and were observing judgments made by another individual left in the shared environment. They were provided with a mutual opportunity to sequentially revise their judgment. Two individuals were provided with either the same or different visual perspectives on the layout. Regardless of the difference in perspectives, judgments made after several rounds of judgment and revision in the interactive mode were systematically more accurate and less variable than judgments made by the same individuals in isolation and looking from one perspective. Experiment 3 confirmed that increases in judgments' accuracy observed during interaction results from the process of inter-individual information integration.

The second aim was to investigate inter-individual information integration processes under different conditions of complementarity, which we operationalized as structural overlap between dimensions of high and low uncertainty in individual perceptions. Our first hypothesis predicted that individuals would obtain more benefit from an opportunity to integrate another's judgment when the amount of complementarity was increased. This prediction was not supported by the experimental results: the increase in individuals' accuracy was not higher in conditions with more complementary visual information relative to conditions with more redundant visual content. We link this absence of an effect of complementarity on the accuracy gain to the difference in efficiency with which people can integrate information from another individual when they have high and low uncertainty associated with their own judgments (see below).

According to Hypothesis III, people were expected to be sensitive to the structure of their own uncertainty. To test this hypothesis, we induced our participants with anisotropic subjective uncertainty about a location in 2D space. This was achieved by providing participants with visual information which had two dimensions: the unreliable dimension where they had high uncertainty about the location and where their judgments had low precision (Ego-Y, or depth dimension); and the reliable dimension where they had low uncertainty about the location and where their judgments had high precision (Ego-X dimension, the dimension orthogonal to the depth dimension on the horizontal plain). In contrast to some previous findings (Zhang, Daw, & Maloney, 2013) we found compelling evidence that people are sensitive to the anisotropy of their own spatial uncertainty. This sensitivity was manifested in an unequal propensity to change one's own judgment on the two dimensions. In line with the prediction from Hypothesis III, the reliable component of participants' judgments was less influenced by another's judgments, i.e. the participants' final judgments were less shifted in the direction of another's judgments on the more reliable dimension of their own visual percept. This provides further evidence for the hypothesis originally proposed by Deutsch and Gerard (1955), which claims that people are more susceptible to the influence of observed judgments when they have more uncertainty about the task.

According to Hypothesis IV people were expected to be sensitive to the structure of reliability of another's judgment. To test this hypothesis, we varied the difference in viewpoints provided to two interacting individuals. This allowed us to vary the structural overlap between the subjective uncertainties of two individuals concerning the location, and this way to examine how participants would integrate another's judgment into their own revised judgment under different conditions of complementarity. On the reliable Ego-X dimension they were expected to weigh another's judgments less than their own initial judgments when the observed judgments were less reliable than their own; on the unreliable Ego-Y dimension people were expected to weigh another's

judgments more than their own initial judgments when the observed judgments were more reliable than their own. The results of three experiments demonstrated that whether participants exhibited sensitivity to reliability of a particular component of another's judgment depended on their uncertainty about this component in their own judgment.

On the reliable Ego-X dimension participant's response to another's observed judgment followed the prediction from Hypothesis IV: people relied on another's judgment less when the latter was less reliable. On this dimension we did not observe rigid egocentric discounting frequently reported in JAS studies (Bonaccio & Dalal, 2006; Yaniv & Kleinberger, 2000). As predicted, when the other's judgments were approximately as reliable as participant's own, the average observed weights were approximately at the 0.5 value, which means that participants trusted observed judgments nearly as much as their own perception. When the observed judgments made by another individual had larger error on participants' Ego-X dimension, the Ego-X component of participants' own judgments was less influenced by the observed judgments. This increase of resistance to influence by the observed judgments was inversely proportional to the reliability of the Ego-X component of the observed judgments. Consequently, the pattern of variability of participants' judgments on the Ego-X dimension followed the prediction from Hypothesis II: variability on the Ego-X dimension was decreased most when two individuals' visual input was fully redundant (Experiment 3), and it either was not decreased at all (Experiment 2) or slightly increased (J45° and J90° conditions in Experiment 1) when the amount of complementarity between the two visual inputs was increased. Thus, resistance to influence by unreliable information available in observed judgments allowed participants to mostly avoid potential negative effects of integrating these judgments into their revised judgments.

For the Ego-Y dimension predictions from our hypotheses were not fully confirmed. Similar to what we found on the Ego-X dimension, and in line with our predictions, when the other's

judgments were approximately as reliable as participant's own, participants trusted observed judgments nearly as much as their own perception (0.5 weights on average). However, in contrast to our predictions, the Ego-Y component of participants' judgments was not subject to larger influence by the observed judgments when the Ego-Y error of observed judgments was smaller than participants' Ego-Y error. In other words, on the Ego-Y dimension participants always trusted another's judgments nearly as much as their own despite the difference in reliability of judgments on that dimension. In this pattern we observe the egocentric discounting effect: a general failure to weigh external judgment ("advice" in JAS terminology) more than one's own judgment (Bonaccio & Dalal, 2006; Yaniv & Kleinberger, 2000). As a consequence, the decrease in variability of error on the Ego-Y dimension did not exactly follow the pattern predicted in Hypothesis II: while variability of Ego-Y errors was decreased in the complementary conditions relative to the fully redundant condition (J0° in Experiment 1), there was no further variability decrease in the fully complementary condition (J90° condition) relative to the partially complementary condition (J45°).

While participants in our experiments were supposed to benefit most from integrating another's judgment on the dimension where their own judgments had low precision, this was exactly the dimension where participants' information integration strategy was insensitive to the quality of the other's judgments. This lack of sensitivity led participants to a recurrent failure to utilize the other's judgments more efficiently in the more profitable conditions.

A possibility we addressed in the second experiment was that availability of more information about the structure of reliability of judgments performed from each viewpoint may lead to a more efficient weighing strategy. To test this, we provided participants with sequential access to two orthogonal viewpoints and let them solve the task interacting with themselves by making consecutive judgments from the two viewpoints. Thus, participants could intra-individually relate the Ego-Y dimension of the percept available from the first viewpoint to the Ego-X dimension of

the percept available from the second viewpoint. In this case, egocentric discounting should not occur because both judgments belonged to the same individual. However, there was no indication that on average individuals could integrate information from two viewpoints more efficiently when it was an intra-individual process. Crucially, they did not decrease the Ego-Y error in this condition relative to the Joint condition.

The similarity between the ID90° and the J90° conditions in Experiment 2 indicates that intra- and inter-individual integration of information through indirect interactions was compromised in the same manner. Therefore, it is possible that there is a common cause that prevented participants both from relying on another's judgments more than on their own judgment and from relying more on their own complementary, second judgment than on their first judgment. This common cause might be general dominance of evidence immediately available over imaginary evidence and consequent anchoring to immediate evidence in the process of judgment revision.

We can generally conclude that the human ability to identify the reliability of information from others publicly available in a shared environment crucially depends on the individual's uncertainty about his or her own judgment. When their expertise is high, an egocentric discounting strategy (Yaniv & Kleinberger, 2000) helps people to identify unreliable information and to ignore it without harming their own accuracy. At the same time, they will integrate information if it is reliable. When their expertise is low, people seem not to be able to discriminate external judgments of higher quality from those of mediocre or lower quality without additional information such as confidence (Bahrami et al. 2010) or related markers of judgment quality (Sniezek & Henry, 1989; Yaniv, 1997).

Several aspects of the process of judgment revision, not directly addressed in this study, are still poorly understood and will comprise an interesting topic for future research. The rest of the discussion is organized as follows. First, I discuss the plausible cues that participants could have relied on to infer the reliability of the partner's judgments in the task we used. Then I suggest a

Bayesian perspective on the problem of judgment revision and briefly discuss how models of a Bayesian observer could contribute to our understandings of the computational processes unfolding within the mind of an individual confronted with both perceptual and social (by origin) information. Finally, I discuss the phenomenon of progressively decreasing influence of the observed judgments over the consecutive iterations of judgment revision, that we observed in all three experiments.

Cues to the reliability of another's judgments

One such question is about the cues which experts use to identify judgments of poor quality just as participants did on the Ego-X dimension in our experiment. First, participants could rely on the feedback immediately available from the partner's judgments. Several studies have demonstrated that people are sensitive to this kind of feedback and utilize it to infer the expertise of the advisor (Budescu, Rantilla, Yu, & Karelitz, 2003; Harvey, Harries, & Fischer, 2000; Henry, Strickland, Yorges, & Ladd, 1996, Yaniv & Kleinberger, 2000). Furthermore, decision-makers can learn to rely on an outlying advisor whose judgments they initially discard if they receive consistent feedback indicating high accuracy of her judgments (Harries, Yaniv, and Harvey, 2004). Yet, we deem the feedback to be at least not the primary cue to reliability of another's judgments in our task. First, participants had relatively few trials (as few as 16 in the first experiment) to learn from the feedback reliability of another's judgments and to noticeably update their weighing strategy within a block of trials. Second, even if participants learnt from feedback about the reliability of another's judgments, by itself this does not explain why they rapidly learnt to discard the unreliable component of another's judgments but did not learn to sufficiently rely on the reliable component of another's judgments.

The second possible cue is visual information which participants acquired when they were given an opportunity to familiarize themselves with the other participant's viewpoint. To exploit this

information, participants would have to construct a mental model of how particular viewpoints are related to subjective 2D uncertainty and to do a mental remapping when revising their judgments. Yaniv and Choshen-Hillel (2012) found that perspective-taking can improve the efficiency of judgment revision as it reduces egocentric bias in advice-taking. However, we suspect that participants were not exploiting this potential source of information, at least not to a sufficient degree. The evidence for this doubt comes from the ID90° condition, in which participants effectively failed to build an allo-centric model of the task and to map the egocentric spatial dimensions from one viewpoint to another, even though they frequently switched between the viewpoints. We hypothesize that this failure is related to a general cognitive difficulty of sequential information integration across consecutive viewpoints (Avraamides, Adamou, Galati, & Kelly, 2012).

The third potential cue is the distance of another's judgments from the participant's own judgment. Davis (1996) proposed that the influence another's judgment will exert on the individual is inversely proportional to the distance between the two judgments. Consequently, judgments which are farther away from the observer's judgment will receive smaller weight in his revised judgment (Davis, 1996; Yaniv, 2004b). Several studies have found empirical support for this theory and established that indeed individuals discount more those judgments which are farther away from their own judgment (Davis et al., 1997; Ohtsubo, Masuchi, & Nakanishi, 2002; Yaniv, 2004b; Yaniv & Milyavsky, 2007). Reliance on the absolute distance between one's own and another's judgment may best explain the data pattern we observe in the current experiments. On the Ego-X dimension, where participants are very accurate, large distances between two judgments would provide strong evidence that the partner's uncertainty is high on this dimension. That would lead to discounting the partner's judgment and this is what we observed. However, on the Ego-Y (low-accuracy) dimension, both accurate and inaccurate partner's judgments are equally likely to occur at a large distance from the participant's own estimate. Hence, the distance will no longer work as a reliable cue for another's

accuracy. If this heuristic is the dominant one, accurate and inaccurate partners will be indistinguishable and will receive equal weight – just what we observed concerning how our participants revised their judgments on their Ego-Y dimension.

Bayesian interpretation

Another important task for cognitive scientists interested in small group decision making is to understand judgment revision at the algorithmic level of the process. One prospective framework is to model individuals as rational Bayesian decision-makers operating under conditions of uncertainty. The process of judgment revision then can be modeled as a process of probabilistic inference where by applying the Bayes rule the observer (the judge) is combining his prior knowledge with new evidence acquired through perception from the environment to arrive at the posterior probability distribution of the environmental state parameter values (Knill & Richards, 1996).

One example of such an approach applied to inter-individual information transmission can be found in Beppu and Griffith (2009). Beppu and Griffith (2009) considered several models which differed with respect to what information was available to individuals from the environment and information they received from the previous agent. One of their models, the *mixed data* model, was particularly designed to capture the situation of indirect interaction where individuals receive evidence about the state of the environment from their own perception and acquire evidence data generated by the previous individual. Beppu and Griffith (2009) claim that if information is transmitted in this way in a chain it will not be very profitable in the long run, and asymptotically the last individual in the chain is expected to perform no better than if he relied solely on his own perception. The question then arises whether this type of interaction can still lead to a better

judgment if two individuals interact reciprocally and exchange samples from iteratively recalculated posterior distributions.

Data from the reported experiments show that cumulative weights numerically differed from immediate serial weights. This implies that taking turns in providing judgments can create an outcome that is different from a one-shot act of judgment revision. In both the J90° condition of Experiment 1 and in Experiment 2 the cumulative Ego-Y weights tended to be higher than the serial weights and closer to what would be expected from the normative weighing strategy. Thus, it is possible that reciprocal interactions have an additional advantage over the one-shot learning considered by Beppu and Griffith (2009). Better understanding the interactive aspects in the process of information exchange requires elaborating existing informational models of judgment revision to accommodate the factor of reciprocal influence and to explain how outcomes on the group level depend on the probabilistic inference processes on the individual level. Indirect interactions can serve as a good starting model for the process due to their relative computational simplicity and tractability of transmitted pieces of information.

Iterative revision process

One aspect that has been largely neglected in research on social influence and small group decision-making is the progressively decreasing amount of judgment shift after revision (Graesser, 1991; Sniezek & Henry, 1990). In the present experiments this effect was manifest in systematically decreasing serial weights across consecutive iterations. The rate of this decrease suggests that at each iteration of revision the most recent judgment made by another was weighed less than his previous judgment. The observed pattern of decaying weights is similar to the pattern of decreasing concessions that are found in dyads approaching a compromise in regard to their arbitrary preferences (Graesser, 1991). Graesser interprets her results in the light of information integration

theory (Anderson, 1981). The decrease in distance at each concession reduces "social forces"[8] operating between the two members. Using the Bayesian framework, this pattern can be interpreted as reflecting individual Bayesian belief updating, where individuals raise confidence in their own revised judgments more than confidence in their partner's revised judgments.

Conclusions

Altogether, our findings extend existing research on judgment revision in the context of inter-individual information integration on several lines. First, we focus on environment-mediated mechanisms, which allow inter-individual integration of quantitative analog judgments without verbal communication and put low demands on meta-cognitive abilities. The results demonstrate that simple indirect interactions are sufficient for individuals to integrate information distributed across them with individual benefit. Second, we demonstrated that people are sensitive to the multi-dimensional structure of their own uncertainty and that their weighing policy is dimension-selective. People also calibrate their weighing policy according to changes in reliability of different components of others' judgments. This ability is restricted though: while people are good at discarding unreliable information, they seem to be much worse at identifying reliable information and at trusting publicly available judgments more than their own perception.

These results correspond to Deutsch and Gerard's (1955) credo that an individual won't trust another's judgment any more than he trusts his own unless his confidence in another's judgments increases. The key question for future research is to identify the conditions under which individuals would justifiably raise their confidence in the reliability of another person to understand how individuals can learn to trust those who have higher expertise to make better judgments.

[8] Graesser uses the term "force" in a heuristic way, implying by it "valuation and integration of informational stimuli in the mind of each group member" (Graesser, 1991, Note 2, p. 30).

Although many authors have emphasized the role of meta-cognition in this process (Bahrami et al., 2010, 2012a, Yeung & Summerfield, 2012), the role of meta-cognition might be overestimated for success in simple behavioral decisions. Just as in the case of chemical pheromones (Hölldobler & Wilson, 1990), much of the information which is supposed to be communicated in meta-cognitive format, can be implicitly and non-intentionally transmitted and perceived in a shared environment. It is an open question how the efficiency of indirect interaction and explicit communication compare but it is not impossible to imagine instances where indirect interaction is more efficient than explicit communication. For example, when choosing which movie to watch, don't we trust more the overall viewer ratings and visiting stats at the IMDb[9] than explicit reviews? And don't we often follow robot-tracked choices of others in the "others who like this also liked…" window?

Further judgment and decision-making experiments with humans and non-human species which utilize more natural environment-mediated modes of interactions might shed further light on the cognitive machinery for human collective intelligence, its evolutionary origins and its role in the social phenomenon of inter-individual accumulation of knowledge.

[9] The Internet Movie Database - an online database of information related to films, television programs and video games.

3. Study II: Joint Judgment Formation

3.1 Introduction

3.1.1 Theoretical background

According to the two-stage model presented in section 1.5, at the stage of joint judgment formation, individuals integrate their judgments into a group judgment. Different individuals can no longer apply independent integration procedures. Instead, they need to agree on a common integration algorithm, or *scheme,* that reflects how individual judgments will be combined into a group judgment. Such an integration scheme does not need to be explicitly stated. It can also emerge dynamically in the course of social interaction. Group judgments that integrate different individual judgments are typically described as a compromise (Hinsz, 1999). The key questions at this stage are: what is the best integration scheme to reach a compromise, how do individuals come up with an adequate integration scheme, and how do they apply this scheme to the given context? In statistical approaches to judgment and decision-making it is considered optimal to apply the integration scheme that minimizes expected loss incurred by a wrong decision or inaccurate judgment over a series of decisions or judgments. In this introductory section I will address the question how we can define an optimal solution for a group of individuals with known abilities, and what information about their abilities intelligent agents need to know and share to integrate individually held information optimally in statistical sense from an information integration theoretical perspective.

Recently a new approach to the problem of inter-individual information integration has been suggested. Each individual decision-maker in a social context is conceived of as an observer making a decision under uncertainty (Bahrami et al., 2012a). This approach draws on an analogy between the computational problem of inter-individual information integration and that of multi-sensory integration, faced by an individual's nervous system when combining different sensory inputs into a single percept (Ernst and Bülthoff, 2004). This functional analogy makes it possible to adopt

computational models originally developed for multi-sensory integration to joint decision making and enables one to directly compare processes of intra- and inter-individual information integration (see section 1.1 of this manuscript).

Similarly, computational models originally developed for the problem of multi-sensory integration of quantitative estimates (van Beers et al., 1996, 1998, 1999) can be adapted to the problem of inter-individual formation of joint judgments. Due to the fact that judgments inherently comprise a choice, further predictions regarding joint judgment formation can be derived based on theories of inter-individual information integration that were originally applied to joint categorical decision-making (Bahrami et al., 2012a; a comprehensive discussion on how the processes of judgment and choice are related and the argument on why this cross-over of predictions is justified can be found in section 1.2 of this manuscript).

The experiments reported in this chapter tested several predictions derived from the theory of inter-individual information integration where joint judgment formation is viewed as a process of integration of individual judgments, and each individual judgment is modeled as an estimation made under uncertainty. The predictions addressed the role of complementarity and the role of information available in the environment in the process of joint judgment formation. Alternative predictions were derived from Bahrami's theory of joint decision-making (Bahrami et al., 2012a) that highlights the role of meta-cognition and communication.

3.1.2 Models for judgment integration

In this section I will review the most common models proposed as normative solutions to the problem of inter-individual integration of judgments focusing on two-person scenarios, along with empirical evidence obtained in experimental studies on group judgment and decision-making in support of each model. Because the current theoretical approach is based on the analogy between

inter-individual information integration and multi-sensory integration, relevant theoretical developments and empirical studies in the field of multi-sensory integration will also be addressed.

Before reviewing different models of integration it is necessary to briefly discuss why integrated group judgments are expected to result in a more precise judgment than the individual judgments the group judgement is derived from. One way to approach this issue is to consider why intra-personal integration of information from different sensory channels leads to more accurate perception. The statistical advantage of information integration in multi-sensory integration or perceptual information stems from the inherently stochastic nature of this process (Ernst, 2006; Knill & Richards, 1996; Maloney, 2002; Yuille & Kersten, 2006). Probabilistic models of multi-sensory perception assume that perceived environmental properties such as the size of an object are derived from a noisy estimate of the true property state that generated the input in each sensory channel. A common assumption is that this noise is Gaussian. Under this assumption, independent estimates are unbiased, but the amount of noise determines the precision with which the estimations are made: more noise leads to lower precision, that means, to higher variability of independent estimates of the same property value. Unless the noise in sensory channels is perfectly correlated, averaging the estimates across several channels is expected to filter out some of the noise and to thereby make the integrated percept more robust (Ernst & Bülthoff, 2004).

The same statistical principles underlie the "random error" model addressing the integration of different individuals' judgments (Einhorn, Hogarth, & Klemptner, 1977; Yaniv, 2004b). When individuals provide an uncertain judgment of a quantity (i.e. the number of antelopes in group A before deciding whether to start hunting group A or group B), one can view each individual judgment as a combination of the true quantity value and some random error added to it. If the error in individual judgment is truly idiosyncratic (i.e., the source of error is not shared across individuals), averaging individual judgments will reduce the random error in the group judgment.

The group judgment derived this way is also expected to be more accurate than individual judgments if the common bias such as a general tendency to systematically overestimate or underestimate the true quantity is not too large (Dawes, 1979; Hogarth, 1978; Stewart, 2001; Yaniv, 2004b).

In social psychology the empirical discovery of this statistical effect resulting from averaging independent individual judgments is attributed to Francis Galton. The discovery that Galton made in a natural set-up (local livestock in Plymouth) was that the average of individual judgments of the weight of a dressed ox that was put up for sale was more accurate than any individual weight judgment and even more accurate than judgments made by experts (Galton, 1907). Since Galton's paper was published in 1907 this phenomenon has been known as the "Wisdom of the Crowd" effect. This effect has been replicated numerous times with judgments in various domains (Ariely et al., 2000; Clemen, 1989; Gordon, 1924; Stroop, 1932).

However, in terms of collective benefit, the averaging procedure that Galton (1907) applied to the collected judgments is not necessarily the best solution to arrive at a group judgment in any circumstances, and it is not always expected to lead to a collective benefit where group judgments are more accurate than judgements of the most accurate individual. Other solutions to the problem of how individuals could integrate their judgments to achieve more accurate group judgements have been proposed in the literature. We now turn to three different computational models, which I will refer to as *integration schemes*. These models have been proposed as actual mechanisms through which individuals form a group judgment from discrepant individual judgments: *Simple Averaging*, *Take-the-Best*, and *Weighed Averaging*. The review will cover (1) the theoretical basis and formal definition of each integration scheme, (2) the advantages and disadvantages of each integration scheme regarding the expected group benefit, and (3) and empirical evidence for the application of each scheme in a situation of joint judgment formation by real groups.

Simple Averaging

One way to combine two or more discrepant judgments is to take their average. This integration scheme, hereafter referred to as *Simple Averaging*, is a particular case of a more general class of linear integration schemes. Let J_i be i^{th} individual judgment of the judged property, also known as the criterion variable, value Y. Given n individual judgments $J_1, J_2, \ldots, J_n$, linear integration schemes are schemes where the joint judgment J_{joint} is derived through an additive combination:

$$J_{\text{joint}} = \sum_{i=1}^{n} w_i J_i \ , \tag{3.1}$$

where $w_1, w_2, \ldots, w_n$ are the weights assigned to the n individual judgments, and $\Sigma w_i = 1$. For the *Simple Averaging* scheme $w_i = 1/n$ for all i. In a two-person scenario the weights are $w_1 = w_2 = 0.5$.

To understand how linear integration schemes can improve the quality of group judgments, it is useful to consider the "random error" model in more detail. Let us return to the example where two judges, A and B, were judging the height of the Empire State Building (see section 1.2). Their goal is to provide the value as close as possible to the true height of the tower. Arriving at a value above or below the true value is equally undesirable. According to the "random error" model, A's individual judgment J_A can be modeled as $J_A = Y_T + e(\mu_A, \sigma_A)$, where Y_T is the true criterion value which in our example is the true height of the Empire State Building, e is the random error modeled as a random sample from a normal distribution with parameters mean μ_A and standard deviation σ_A, and similarly for judge B. The model parameters μ_i and σ_i correspond to the bias and precision associated with judge *i's* individual judgment. The precision of each individual judgment is directly related to the amount of uncertainty in a judge's internal representation of the estimated quantity: higher uncertainty leads to less precise judgments.

Let's now consider how exactly Simple Averaging is expected to make the joint judgment more accurate than the individual judgments the joint judgment is derived from. The statistical effect of canceling out random error takes place when individual judgments bracket the true value (Soll & Larrick, 2009; Larrick, Mannes, & Soll, 2012), that means, they are located at different sides relative to the true value on the underlying continuum. If the true tower's height Y_T is 443.2 m, suppose that Judge A underestimates its height and provides a judgment $J_A = 440$ m. The absolute error (unsigned error) of his judgment $|e_A| = |Y_T - J_A|$ is 3.2 m. Now suppose that Judge B overestimates the tower's height and provides judgment J_B equal to 450 m. The absolute error of her judgment in this case is 6.8 m. The average of the two judgments J_A and J_B is equal to 445 m., and the absolute error of the joint judgment in this scenario will be 1.8, which is smaller than the absolute error of either A's or B's individual judgment. When individual judgments bracket the true value, the absolute error of the average of the two judgments will always be smaller than the average of absolute errors of the individual judgments.

Now suppose both J_A and J_B occur on the same side of the continuum relative to the true value, for example both judge A and judge B underestimate the tower's height and provide judgments of 440 and 435 m. In this scenario the absolute error of the joint judgment derived through Simple Averaging is equal to the average of the absolute errors of J_A and J_B: $|e_{Joint}| = |443.2 - (0.5\times440 + 0.5\times435)| = 5.7$ m. The absolute error of the joint judgment derived through Simple Average can never be larger than the average of absolute errors of individual judgments.

If judges A and B are providing many judgments, for example, estimates of the height of many different towers with the same amount of noise in their individual perceptions, and form joint judgments through Simple Averaging, the absolute error of any single joint judgment will be at least as large as the average of absolute errors of individual judgments, it will sometimes be smaller than the average absolute error, and it will sometimes be less than the smaller absolute error of the two

judgments due to the bracketing effect. Thus, the joint judgment $J_{joint} = 0.5J_A + 0.5J_B$ is expected to contain smaller absolute error than the average absolute error of the individual judgments J_A and J_B.

Simple Averaging is a heuristic scheme (Gigerenzer et al., 1999), which means that unlike statistically proper integration solutions, to apply this scheme, individuals do not need to assess and match accuracies or other relevant statistical properties of their respective judgments. Intuitively, this scheme seems rather improper when an expert and a novice need to arrive at a consensual judgment because it seems that higher weight should be given to the expert's judgment. At the same time, when individual expertise, that means the precision with which individuals make their judgment, cannot be reliably determined, Simple Averaging is a safe strategy because it protects groups from assigning large weights to the judgments of the less competent individual (Einhorn, Hogarth, & Klemptner, 1977).

Assigning equal weights to the judgments of two or more individuals is most advantageous in a situation where both the difference between individual judgment precisions and individual biases are relatively small. When individual precisions are equal and individuals are unbiased in their judgments, Simple Averaging is the statistically optimal solution for integrating individual judgments into the joint judgment, and the joint judgment is expected to be more accurate than the judgment of the more accurate individual, causing the "Wisdom of the Crowd" effect. At the same time, this heuristic is notably robust to mild violations of the precision equality condition and to small non-zero biases. Thus, equal-weighing will bring close to maximum possible collective benefit if the optimal weights (reflecting differences in the precision of individual judgments) are not too extreme. This property can also be read the other way: moderate deviations from equal weights are relatively negligible for the collective benefit (Dawes & Corrigan, 1974; Soll & Larrick, 2009). These robust properties of Simple Averaging can be explained on the flat-maximum account (von Winterfeldts & Edwards, 1986). On this account, averaging flattens the function which links selected weight values

to the loss incurred by inaccurate judgments by its general filtering property. Consequently, in most circumstances Simple Averaging performs well even relative to more complex and fine-tuned methods for judgment integration (Armstrong, 2001; Clemen, 1989; Lin & Cheng, 2009).

For these desirable statistical qualities and its "robust beauty" (Dawes, 1979) several authors have proposed Simple Averaging as a normative solution to the problem of post hoc statistical aggregation of individual judgments, as a method for improving accuracy in estimation and prediction tasks (Armstrong, 1985; Ashton, 1986; Hogarth, 1978; Zarnowitz, 1984), and as a solution to the problem of judgment integration during group discussion (Einhorn & Hogarth, 1975). Einhorn and Hoggarth (1975) conclude that "< …> so long as the variability in the judges' decisions is not too different and (b) they all have positive validity (i.e., even minimal expertise), a simple average of their decisions will provide an excellent prediction" (p. 188). They further advocate this scheme as having

> the appeal of a democratic procedure which would be particularly suited to pooling the judgments of experts. It clearly avoids the awkward problem which could arise if one expert were to discover that his opinion was weighted one-half that of a colleague! (Einhorn & Hoggarth, 1975, p. 188)

Other authors (Einhorn, Hogarth, & Klemptner, 1977; Hinsz, 1999) also listed Simple Averaging as a viable model for the actual psychological process through which a group reaches a compromise when interacting and bargaining.

Given the low cognitive effort that implementing Simple Averaging demands from group members and considering its powerful potential to improve the quality of the group judgment, this scheme appears to be a reasonable and straightforward solution to the joint judgment formation problem. Furthermore, because individuals have been demonstrated to actively and spontaneously employ the strategy of averaging quantitative information available from multiple sources to form a single judgment (Birnbaum, Wong, & Wong 1976; Shanteau & Nagy, 1982), one could reasonably expect them to generalize this strategy to the group level. There is little evidence, however, that

groups actually use Simple Averaging when combining individual judgments into a group judgment. Interpreting existing empirical studies on small group judgment and decision-making with regard to which actual integration scheme is applied by groups is further complicated by the fact that, traditionally, small group research exclusively focused on the quality of the combined judgment and not on the actual mechanisms that combine individual judgments into group judgments[10]. As a consequence, researchers have often reported data aggregated across items and did not track how each group judgment was formed (a similar argument was made by Soll & Larrick, 2009).

Nevertheless, even from aggregated data one can infer whether a particular integration scheme was applied by spelling out qualitative predictions. If groups apply Simple Averaging to form a judgment, one of the clearest empirical predictions is that group judgment should be more accurate than the individual judgment of the average group member. In fact, this was rarely found in experiments. In his extensive review of experimental studies on group decision-making with quantitative tasks, Hastie (1986) concluded that in only two out of 14 reviewed studies were groups more accurate than the average group member.

There are few exceptions though. Schonbar (1945) found that dyads were more accurate than individuals at estimating line lengths, and she noted that averaging helped individuals to cancel out individual over- and under-estimation errors. Gigone and Hastie (1993) asked participants assorted into ad hoc groups to make consensus group judgments of the grades that students received for a course. For each case (student) individual group members were privately provided with a set of quantitative predictors, such as the student's school GPA or self-rated workload. These predictors served as the informational input for individuals' judgment of that student's course grade,

[10] This is what N. Anderson referred to as a research focus on *outcome generality* rather than *process generality* (see Editor's Note 2 in Graesser, 1991). Research on outcome generality aims at "*generalizing experimental outcomes more or less directly*" and it is a concern that many psychological studies implicitly aim at outcome generality. General theory, according to Anderson, should seek process generality, not outcome generality.

and then individual judgments were combined into a group judgment. Each group provided judgments (one judgment per student) within a short deliberation time where the group judgment was discussed: less than two minutes per judgment on average. The results showed that Simple Averaging of individual judgments provided a very good fit to the group judgments. It is, therefore, possible that groups employ a Simple Averaging scheme under time pressure.

Other than that, the general picture is that groups do not rely on Simple Averaging much when forming group judgments. One potential reason for this is that people possess wrong conceptions about the effects of averaging. Accordingly, a study conducted by Larrick and Soll (2006) showed that people believe that averaging individual judgments will lead to an average level of accuracy rather than a group benefit. Similarly, lay people are skeptical of averaging information from sources susceptible to bias and noise (Soll, 1999). Altogether, one can conclude that people underestimate the power of averaging, and when forming joint judgments groups do not rely on this scheme as frequently as one would expect.

Simple Averaging is by no means a universal scheme for integrating individual judgments. Several factors can undermine its efficiency. One drawback of Simple Averaging is that its deviation from optimality increases rapidly when judges diverge in their precision. Another drawback is that the efficiency of integration through equal weighing decreases when there is some shared tendency to make similar errors, for example, a shared bias or high correlations between individual errors. When there are substantial correlations between errors across judgments of different individuals, equal-weighing of individual judgments is expected in only a narrow range of inter-individual precision ratios to lead to a level of accuracy that is as high as or higher than that of the most accurate individual (Soll & Larrick, 2009). Accordingly, if the joint goal is to maximize accuracy of the joint judgment, under certain conditions a better strategy for integrating information across individuals is to simply take the judgment from the more accurate member.

Take-the-best

If two individuals can reliably identify which individual is more accurate they can form a joint judgment by selecting the judgment of the more accurate individual. This scheme, which I'll refer to as *Take-the-best* after Soll and Larrick (2009) is the second integration scheme reviewed in this section. Soll and Larrick (2009) adapted this model from Gigerenzer and Goldstein (1996) to the situation of inter-individual integration of judgments as a viable psychological alternative to linear integration schemes. Originally applied to multiple cue integration in a situation of categorical choice, this integration scheme was proposed as an efficient decision strategy in environments where the most informative cue is more likely to point to the correct alternative even when all other cues point into a different direction. Such environments are often referred to as non-compensatory environments (Hogarth & Karelaia, 2006; Martignon & Hoffrage, 1999, 2002).

In a situation of continuous judgments this scheme can be favored over Simple Averaging when the difference in precision of individual judgments is very large; or when individuals share the same bias and this bias substantially contributes to the overall error. Under both conditions Simple Averaging is expected to result in a less accurate judgment than the judgment of the more accurate individual.

The study by Soll and Larrick (2009) suggests that when individuals have formed their own judgment and are exposed to somebody else's judgment afterwards, they prefer to select one of the two judgments (the one they believe is more accurate) rather than to average the two judgments. A similar tendency might apply to groups forming joint judgments. Empirical evidence that groups selected the judgment of the most competent member as the group judgment was obtained in experiments by Einhorn, Hogarth, and Klemptner (1977) and later by Uecker (1982). However, Einhorn et al. (1977) report that three out of 20 groups in their sample were more accurate than the

most accurate member in that group, which is not consistent with the Take-the-best strategy and suggests that an averaging process took place in these groups.

The conditions under which the integrated judgment derived through the Take-the-best scheme is expected to have higher accuracy than the integrated judgment derived through Simple Averaging, or vice versa, depends on the interplay of several factors which include a) the probability with which the more accurate judge can be identified; b) the level of precision of one judge relative to the precision of another judge; and c) a tendency to make similar errors. A detailed analysis with a specification of conditions where one or the other strategy should be favored over another was performed by Soll and Larrick (2009).

Weighted Averaging

The third scheme for integrating individual judgments, reviewed in this sub-section, is *Weighted Averaging*. Because, under this scheme, the group judgment can be described as an additive composite of individual judgments (Eq. 3.1 applies), this scheme also belongs to the general class of linear integration schemes. The main difference compared to the Simple Averaging scheme reviewed earlier, is that individual judgments do not equally contribute to the group judgment. Rather, individual judgments are weighed differently, depending on their quality.

The idea that a group decision or judgment can be described as a weighted composite of individual contributions where individual group members are not generally expected to contribute to the group decision or judgment equally can be found in the works of several authors (Cartwright & Zinder, 1968; Einhorn et al., 1977; Graesser, 1991; Steiner, 1972).

Graesser (1991) took a formal approach to the problem and provided a conceptualization of group decision-making and bargaining in terms of *social averaging*, which is summarized in her Social Averaging theorem (p. 2-3). The theorem asserts that group decisions can be represented as a

weighted average of individual members' most preferred decisions. This can be generalized to continuous judgments by allowing unequal w_i terms in Eq. (3.1). For the two-person scenario this implies that

$$J_{\text{joint}} = \frac{w_1 J_{\text{A}} + w_2 J_{\text{B}}}{(w_1 + w_2)}, \tag{3.2}$$

where $w_1 \neq w_2$ condition is allowed. Graesser (1991) tested the Social Averaging theorem by investigating how dyads would integrate their quantifiable preferences to reach a compromise group decision. She found that a weighted average of individual preferred decisions provided a perfect fit to compromise solutions reached by a dyad, while a version of Simple Averaging provided quite a poor fit. However, the situation of conflicting preferences modeled in Graesser's experiment (1991) is different from a situation of judgment under uncertainty considered in the current work. Because in a situation of conflicting preferences there is no true answer, there can be no prescriptive solution on how to minimize expected loss incurred by a wrong joint answer. In other words, it is impossible to specify *how* individuals *should* weigh their individual judgments to form the joint judgment.

In a situation where the group goal is to provide a judgment as close as possible to some true value, which is the situation of judgment under uncertainty, the specification of a proper weighing policy is possible and it follows from statistical properties of the "random error" model of an individual judgment. As a general statistical principle, to minimize random error in the combined estimate several sources with unequal precision should be weighed proportional to their reliabilities (inversely proportional to their variances):

$$w_{\text{i}} = \frac{1/\sigma_{\text{i}}^2}{\sum_{\text{j}=1}^{\text{n}} 1/\sigma_{\text{j}}^2}, \tag{3.3}$$

where $j \leq i \leq n$, and σ_{i}^2 corresponds to variance of errors inherent in estimates from the i^{th} source.

This principle of weighing by reliability has a prominent role in research on multi-sensory integration, as it provides a precise prediction on how the nervous system should combine noisy inputs from distinct sensory modalities into a combined percept (Ernst & Bülthoff, 2004; Oruç̧, Maloney, & Landy, 2003). In an influential study, Marc Ernst and Martin Banks (2002) addressed intra-personal information integration across visual and haptic sensory modalities. Participants were instructed to compare the stimulus size against a standard object, but the visual information projected on a mirror and haptic information delivered through a forced-feedback device was manipulated independently. Despite the fact that sensory information available through the haptic channel was much less reliable than the sensory information available in the visual channel, the haptic information was not ignored but weighted proportionally. Furthermore, when information from both modalities was available, estimates of the stimulus size were more precise than estimates based on any single modality. When extra noise was added to the visual channel, the relative weight of the haptic percept in the integrated estimate was increased, confirming that the nervous system is dynamically calibrating the relative contributions of the sensory channels according to their reliability. Following up on this study, optimal integration has been demonstrated between visual and auditory modalities (Alais & Burr, 2004), vision and proprioception (van Beers et al., 1999) and between stereo and texture signals within the single visual modality (Knill & Saunders, 2003).

The conceptual analogy between the problems of intra- and inter-individual information integration based on the common "random error" statistical model makes it possible to apply theoretical principles which govern information integration processes on the individual level to the group level. Along these lines, a similar principle of weighing by reliability has been suggested for optimal integration of judgments from two or more individuals with unequal expertise (Klein & Sprenger, 2015; Libby, Trotman, & Zimmer, 1987).

Suppose that Judge A and Judge B make a series of judgments, for example, of the height of n towers. For each distinct tower i the individual judgments $J_{A,i}$ and $J_{B,i}$ will contain some random error that is independent from errors in judgments made for other towers. Judgments from the less precise judge will tend to have larger absolute errors due to larger variance of errors σ^2. If for each tower the two judges integrate their individual judgments through Weighted Averaging so that individual judgments contribute to the joint judgments weighed by their precision (see Eq 3.2 and Eq 3.3), the joint judgments are expected to have smaller variability than the judgments of the more precise judge. This implies that if systematic errors (biases) are small, each single joint judgment is expected to be closer to the true value than the initial individual judgments. Furthermore, there is no other linear integration scheme which would allow reducing random error in the integrated judgments more than through this weighing policy. This implies that if the two judges are unbiased, this integration scheme is the statistically optimal way to integrate individual judgments into a joint judgment.

A version of a Weighted Averaging scheme was tested by Sniezek and Henry (1989). In their experiment participants were instructed to individually estimate the frequency of deaths from 15 causes and to then make group judgments. One of the results was that 30% of the groups in their sample were more accurate than the most accurate individual judgment from each group. Sniezek and Henry (1989) concluded that the quality of group judgments was "as if unequal partially valid weights were assigned to initial individual judgments" (p.21), thus favoring a Weighted Averaging process. The authors did not evaluate, though, whether individual judgments were weighed exactly by their precision.

3.1.3 The role of confidence and meta-cognition

One of the major theoretical problems pertaining to research on small-group judgment and decision-making is to explain the recurrent failure of groups to properly integrate individual information, in particular quantitative judgments, to achieve maximum collective benefit, and to understand the conditions which increase the chances of successful integration. While optimal models of information integration predict that integrated decisions or judgments should never be inferior to those of the more accurate group member, this is rarely observed in empirical studies, as can be seen from several reviews (Ferrell, 1985; Hastie, 1986; Lorge et al., 1958). In their review of the studies on group judgment integration Lorge et al. (1958) resume: "At best group judgment equals the best individual judgment but usually is somewhat inferior to the best judgment" (p. 348).

Bahrami and his colleagues (2010) proposed a new theory postulating that group disadvantages are due to communication limitations: individuals cannot communicate independently the perceptual content informing their decision and the reliability associated with their decision, as would be required for optimal interpersonal information integration. They hypothesized that, instead, individuals can only communicate confidence in their decisions, which they defined as meta-cognitive awareness of one's probability of being correct.

Bahrami et al. (2010) tested their theory using a perceptual signal-from-noise discrimination task. In their experiment participants were instructed to jointly judge which of two successively displayed arrays of Gabor patches contained a patch that differed in its lightness from all other patches in the array. After the stimulus was presented to each participant individually the two participants were encouraged to communicate before they agreed on a joint judgment. The variability in individual accuracies in the initial experiment was not too large, and when both individuals in a group were exposed to the display with the same amount of noise in the stimulus, group accuracy exceeded accuracy of the best individual in the group.

In a further experiment Bahrami and colleagues (2010) added additional noise to one of the dyad member's display to increase the inter-individual discrepancy in relative precision (note that this manipulation is very similar to the manipulation where Ernst and Banks (2002) added noise to one of two sensory channels in their study on multi-sensory integration of perceptual information within individuals). Following their prediction, dyads could not outperform their best member (who performed the task without added noise) in this unequal condition. According to Bahrami's theory (2010, 2012a) this happens because confidence is a sub-optimal measure of individual uncertainty. When two individuals have different amounts of uncertainty, relying on shared confidence over a series of decisions will sometimes lead a dyad to a decision different from the one it should make if individual observations were integrated using the weighing by reliability principle. Furthermore, the detrimental effect of relying on shared confidence is larger for more unbalanced dyads: the larger the difference in individual uncertainty about their observations, the further group performance will deviate from optimality. The authors conclude: "Individuals with very different sensitivities[11] are best advised to avoid collaboration and instead should rely entirely on the more sensitive individual" (Bahrami et al., 2010, p.1085).

3.1.4 The role of complementarity in individually held information

As was discussed in sections 1.2 and 1.3, conclusions derived from the analysis of decisions or judgments that are made with regard to one particular dimension such as size or weight do not necessarily generalize to a situation where multidimensional judgments are made such as when judging the location of an object on two spatial dimensions. In the latter case the true criterion value is characterized by a vector, for example, the two spatial coordinates of the location at which an object is located. More specifically, the conclusion reached by Bahrami et al. (2010) that individual

[11] One can substitute *sensitivity* with *precision* to apply this conclusion to integration of judgments.

collaboration will yield less collective benefit in a situation when individual sensitivities (or, equally, precision of internal representations) are different, appears to be in contradiction with the theorem of Maximum Collective Benefit that captures the benefit of having access to complimentary information in the multidimensional case (section 1.3). The Theorem predicts that two individuals should most benefit from collaboration when the precisions of their internal representations are different, namely, when there is a structural asymmetry so that the dimension of high precision for one individual is the dimension of low precision for another individual, and vice versa. So how can the two statements be reconciled? The conclusion reached by Bahrami et al. (2010) is not incompatible with the theorem. Instead, in the light of the theorem, Bahrami's conclusion holds only in situations either when the dimensions of high and low precision for the two individuals are collinearly aligned or when the judgments of at least one individual are equally precise on all dimensions (isotropic).

To illustrate the relation between complementarity of individually held information and the amount of collective benefit from interpersonal judgment integration, consider the following example. Judges A and B are making multidimensional judgments that can be captured by a vector, for example judgments about the position of a dart on a dartboard. The target is specified with two coordinates – x (corresponding to the left-right dimension of the board) and y (corresponding to the height of the board). As with one-dimensional judgments, the goal is to provide a judgment that is as close to the true value as possible. In two dimensions a proper way to measure closeness of the judgment location to the true value, which is served by absolute error on a one-dimensional continuum, is absolute distance from the true position to the judged position, also known as Euclidean distance. Suppose that our coordinate system is always centered at the true value, that means, the target coordinates are [0 0]'. If we denote the absolute distance with d, then $d^2 = j_x^2 + j_y^2$, where j_x and j_y are judgments of the x and y position of the dart respectively. Now, assume that

Judge A and Judge B are unbiased, but have some uncertainty about the true position of the target: this implies that their judgments will be symmetrically scattered around the true position, with some random x and y error. This uncertainty can be visualized with confidence ellipses (red and blue ellipses in Figure 13, p. 128): the axis of elongation of an ellipse corresponds to the dimension of the largest spread of expected judgment errors (the dimension of low precision), and the orthogonal axis corresponds to the dimension of the lesser spread of judgment errors. The axes of the confidence ellipse do not need to be aligned with the axes of our coordinate system: if judgment errors on different dimensions are correlated, the elongation axis of the confidence ellipse will be oriented in the direction of the largest variance of judgments. To put it formally, the axes of the ellipse are aligned with the eigenvectors of the variance-covariance matrix of judgment errors.

For the two judges A and B, the dimensions of high and low precision of their judgments may not be aligned, for example, because they are looking at the dartboard from different perspectives, leading to different confidence ellipses for each judge. Without a loss of generality, assume that for Judge A (blue ellipse in Figure 13) the dimension of high precision is always aligned[12] with the horizontal axis, and the dimension of low precision is aligned with the vertical axis of our coordinate system. To put it formally, the variance of Judge A's judgment will be larger on the vertical axis than on the horizontal axis, and there will be no correlation between the errors on both axes. In the simplest case the same is true for Judge B (her uncertainty is illustrated with the red ellipse in Figure 13). This implies full structural overlap between the two dimensions of varying precision across A's and B's judgments. When these dimensions are not aligned, the structural difference in precision can be captured as the angle ϕ between the major (or minor) axes of the two confidence ellipses characterizing uncertainty of the two judges in two dimensions. When the

[12] If it is not, we can always align our coordinate system with the axes of A's confidence ellipse so that it is oriented exactly as in Figure 13 with an affine transformation.

dimension of Judge A's highest precision is the dimension of Judge B's lowest precision and vice versa, that means, individually held information is complementary, the angle ϕ is equal to 90°.

As Figure 13 on p. 128 illustrates, even when individual judgments are equally precise overall, the amount of structural overlap has consequences for the accuracy of the joint judgments that Judges A and B can reach by optimally integrating their individual judgments (far left and right of the figure). The abscissa axis in Figure 13 plots the angle ϕ between the dimensions of high precision in individual judgments of Judge A and Judge B. The ordinate axis plots expected absolute distance d^2 between the true position and joint judgments of the position integrated through a Weighed Averaging[13] scheme with weights corresponding to the reliability of each judge on each spatial dimension. The total individual uncertainty about the true position can be characterized by the variability parameters $\sigma_{A,Y}$ and $\sigma_{A,X}$ for Judge A, and $\sigma_{B,X}$ and $\sigma_{B,Y}$ for judge B. In Figure 13 the first three parameters are fixed for all black curves. The last parameter $\sigma_{B,Y}$ is parameterized through coefficient c, so that $\sigma_{B,Y} = c\sigma_{A,Y}$. Values of c larger than 1 imply that Judge B has more overall uncertainty than Judge A, and her judgments are less precise. Different curves on the graph are plotted for different values of c.

The graph illustrates two things. First, less structural overlap in individual uncertainty leads to more collective benefit: in the range between 0° and 90° all curves decrease reaching a global minimum at 90°. This implies that maximum collective benefit is reached when two judges have access to complementary information on different dimensions. The second observation is that collective benefit depends less on the difference in overall precision of individual judgments between Judge A and Judge B as structural overlap becomes smaller. When the angle ϕ is 0°, that means, individual dimensions of high and low precision are collinear, Bahrami's conclusion holds:

---

[13] The formal definition of the multi-dimensional version of this scheme will be provided later, and is not important for the current illustration purpose.

the difference in overall precision between Judge A and Judge B greatly influences the benefit obtained from integrating judgments. This can be seen by comparing the five curves representing different c values on the far left and the far right of Figure 13. Lower c values imply a smaller difference between the individual uncertainties of the two judges and lead to larger collective benefit (distance between black lines and blue line for a given $\phi = 0°$ and $\phi = 180°$). At the same time, the differences between the different curves are very small in the range between 30° and 150°, and are minimal at $\phi = 90°$. This pattern implies that when dimensions of individual precision are not collinear, reliable information held by one judge can compensate for unreliable information held by another judge and that this compensation effect should be maximal when complementarity of information held by the two judges is maximal.

Two qualitative predictions follow from the theorem: first, accuracy of joint judgments should depend less on the difference in individual overall precision between the two judges when individually held information is complementary rather than redundant. Second, individuals should achieve more collective benefit from collaboration when individually held information is complementary rather than redundant.

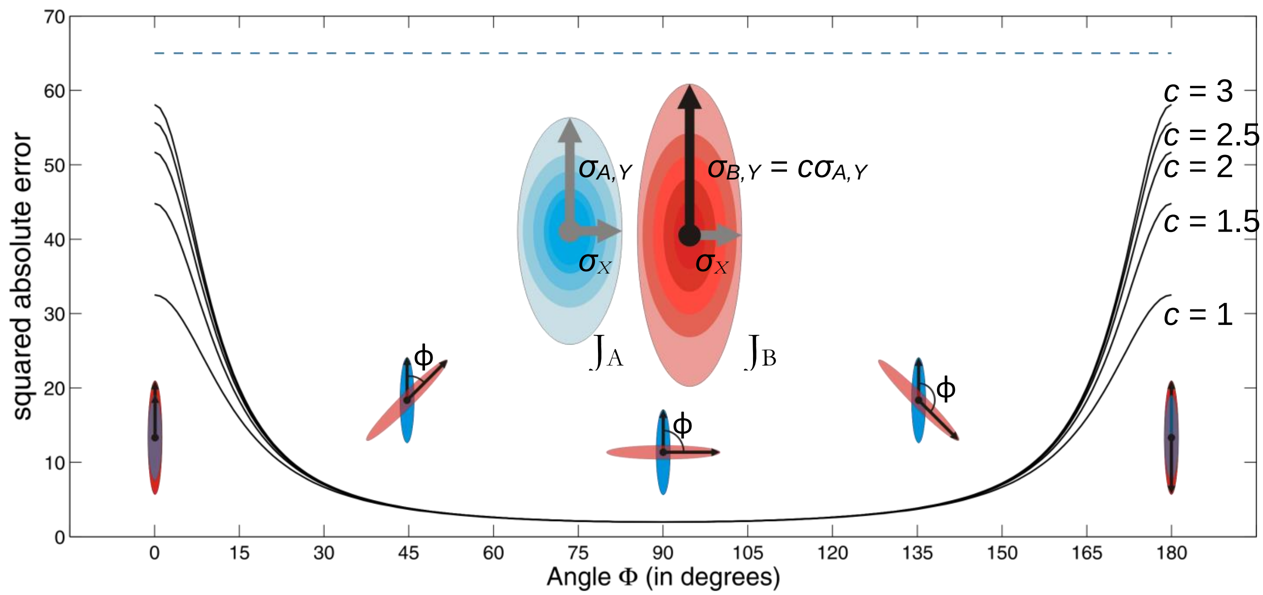


Figure 13. Illustration of the theorem of Maximum Collective Benefit. The bigger ellipses in the center of the figure represent variability of random error in Judge A's (J_A, in blue) and Judge B's (J_B, in red) 2D judgments on a horizontal (x) and vertical (y) dimension. The ordinate axis plots expected squared absolute error in optimally integrated joint judgments under the assumption that individual judgments are unbiased. The abscissa axis plots the angle ϕ between the major axes (black arrows on smaller ellipses) of confidence ellipses constructed for A's and B's judgments and it captures the structural overlap in the two judges' uncertainty about the true value of a position in two-dimensional space. ϕ = 0° implies redundant information distribution (small ellipses on the far left and far right), ϕ = 90° implies complementary information distribution (small ellipses in the center).

Distinct black curves correspond to different ratios of the relative precision of A's and B's individual judgments on the dimension of low precision. On the bigger ellipses, grey arrows denote fixed parameters and the black arrow indicates the argument that was varied to create different ratios. It was assumed that the two judges have equal standard deviation of error along the dimension where their judgments are precise, $\sigma_{A,X} = \sigma_{B,X} = \sigma_X = 1$. Standard deviation of error in A's judgments along the dimension of low precision was set to $\sigma_{A,Y} = 8$. Standard deviation of error in B's judgments along the low precision dimension was parameterized through a multiplication coefficient c: $\sigma_{B,Y} = c\sigma_{A,Y}$.

The blue dashed line on the top shows the expected squared absolute error in individual judgments of Judge A (the more precise of the two). The distance between the blue line and the black curves illustrates the amount of collective benefit resulting from optimal integration of individual judgments for different amounts of structural overlap in information held by the two judges (from ϕ = 0° to ϕ = 180°)

3.1.5 Communication vs. Environment-mediated interaction in joint judgment formation

As was noted in sections 3.1.1-3.1.3, when forming joint judgments individuals will need to coordinate their contributions to the joint judgment if they differ in the accuracy with which they can estimate the true quantity. Verbal communication is one such means for coordination which is available for interacting individuals. Previous research has shown that communicating confidence enables information integration across two people when they choose between discrete alternatives. Aligning different internal coordinate systems to have the same origin and the same units plays a major role in this integration (Ernst, 2010; Bahrami et al., 2012a).

If communicating confidence were the only way to integrate information interpersonally, inter-individual perceptual integration should fail when individuals are not able to communicate their confidence. Accordingly, Bahrami et al. (2010) tested whether depriving individuals from communication would hinder the information integration process. In line with their predictions, dyads could not outperform their best member and performed at the accuracy level of their best individual when they could not communicate. Furthermore, replacing verbal communication with confidence scores did not provide a functional substitute: dyads were still underperforming (Bahrami et al., 2012a). Together, the evidence supported the assumption that the inter-individual ability to align confidence and to come up with a joint reference system for expressing and interpreting confidence plays a crucial role in this process This also implies that benefit of integrating individual decisions and judgments into a group decision or group judgment will depend on individuals' meta-cognitive abilities and, in particular, on their ability to assess their subjective level of confidence and to communicate it to others (Fusaroli et al., 2012).

But is verbal communication of confidence the only means to efficiently coordinate individual contributions to a joint judgment? Answering this question with yes will bring us to the conclusion that without explicit communication of meta-cognitive information individuals with different level of competence will not achieve a collective benefit. However, a potential alternative route to achieve collective benefit is to rely on environment-mediated interactions when integrating individual percepts to form a group judgment. There are several ways in which interaction *in* a shared environment and *via* a shared environment can aid the integration process. In many instances the perceptual information to be integrated across individuals is continuous and available directly from the environment. At the same time, the environment itself provides a common spatial reference as well as a medium for information exchange.

This may remove the need to rely on meta-cognitive means of communication. When coordinating joint actions, the physical environment can be exploited for transmission of information necessary for interpersonal coordination via the haptic modality (Sebanz, Bekkering, & Knoblich, 2006; van der Wel, Knoblich, & Sebanz, 2011) or the visual modality (Millar, Oldham, & Randshaw, 2012). Force coupling in physical environments can make individual actions more efficient (Ganesh et al., 2014). Ganesh and colleagues (2014) provided participants with a continuous target tracking task where unbeknown to them, they and another participant were tracking the same target on independent displays. Coupling the two participants with a virtual elastic band was mutually beneficial and led to a reduction in participants' tracking error. A similar effect but with collective benefit was demonstrated by Wahn, Schmitz, König, and Knoblich (2016) who implemented an object manipulation task with joint control. Participants were manipulating the cursor with keyboard keys moving it to the target location. In the joint mode, control was distributed between the two participants in two spatial dimensions. The authors found a reaction time improvement and accuracy improvement in the joint mode compared to the individual

performance of the more skilled individual. This study demonstrated that sharing a common environment can be sufficient for achieving collective performance benefit even in the absence of verbal communication, at least in tasks which require precise motor control.

In the current study, we investigated to what extent and under what conditions environment-mediated interaction can replace verbal communication when providing joint perceptual judgments of object location in a shared environment. Our hypothesis was that efficient integration of individual perceptions to form a joint judgment does not necessarily require verbal communication and that coordination via a shared environment where individuals can gradually adjust their contribution to the final judgment can be sufficient to reach a compromise. By the Social Averaging theorem (Graesser, 1991), the compromise can be described as a weighted average of individual most preferred positions, that is, the compromise judgment should lie between the locations that individuals deem to be the correct ones. This should have a statistical effect of averaging on joint judgments. Assuming that perceived object location is subject to random deviations from the true location, the prediction then is that reaching a compromise should be sufficient to cancel out some random error inherent in individuals' internal estimations and improve judgments accuracy. If individuals contribute to the final judgments proportionally to the precision of their internal estimates of the location, they should also achieve optimal or close to optimal integration of individual perceptions.

3.1.6 Method overview and hypotheses

To investigate how environment-mediated interactions can support the process of joint judgment formation in a situation of redundant and complementary distribution of information in a group, an empirical study was conducted which involved a series of four experiments. This study had several aims. The first aim was to test whether two individuals can achieve optimal integration

of individual perceptions, resulting in collective benefit from collaboration, in a shared interactive environment under conditions of redundant and complementary access to multidimensional spatial information. The second aim was to test whether the opportunity for verbal communication is a necessary factor for successful integration of perceptual judgments or whether environment-mediated interactions can substitute verbal communication. Finally, the third aim was to test whether availability of feedback is a necessary factor for successful integration of perceptual judgments and whether verbal communication can substitute perceptual feedback in the process of joint judgment formation.

To these aims, the same experimental task as in Study I was used. Participants localized a target at the bottom of a virtual 3D container displayed as a 2D projection on a monitor. Participants were providing their judgments by moving a pointer on the top plane of the container which was parallel to the plane containing the target. The correct coordinates were the coordinates exactly above the target: if the pointer was located at these coordinates a perpendicular would connect its apex with the center of the target (see Figure 2A, p. 58). By introducing perspective distortions to the 2D projection we induced anisotropy in participants' perceptual uncertainty about the true target location. Participants had high uncertainty on their depth dimension (along their virtual line of sight): on this dimension participants' judgments were less precise and had higher variability of errors. On the dimension at the right angle to the depth dimension participants had low uncertainty: their judgments were more precise and had lower variability of errors.

In the Joint condition participants were locating the target together and had to provide a joint judgment for each trial. By manipulating difference in viewpoints on the virtual layout generated for two participants we created two conditions. In the redundant visual input condition two participants were provided with the same view on the container. In the complementary visual input condition orthogonal views on the container were provided to the two participants. In this

configuration, the dimension of high uncertainty of one participant was the dimension of low uncertainty for the other participant, and vice versa.

In the single-perspective condition (IS) participants provided location judgments from one viewpoint in isolation. This condition tested individual performance from one single perspective and served as a baseline for performance comparisons.

Four experiments were conducted in total. Experiments 1 and 4 involved complementary visual input, and Experiments 2 and 3 involved redundant visual input. Feedback on judgments accuracy was provided in Experiments 1 and 2, and was not provided in Experiments 3 and 4. Availability of verbal communication in the Joint condition was manipulated within-subjects in all four experiments. This design allowed us to investigate the interplay between communication and feedback in a situation of redundant and complementary information distribution separately.

Hypotheses

The first hypothesis is related to the participants' ability to track the geometric structure of joint uncertainty: if participants employ optimal or near-optimal strategies of information integration the accuracy of joint judgments should be higher than that of the more accurate individual in a given dyad. According to the theorem of Maximum Collective Benefit, this effect (if anything) should be stronger under conditions of complementary visual input (Experiments 1 and 4) than under conditions of redundant visual input (Experiments 2 and 3). An auxiliary hypothesis following from the theorem is that the amount of collective benefit should depend more on the difference in localization precision between two partners under a condition of redundant visual input than under a condition of complementary visual input.

The second hypothesis is related to the role of the environment in integrating individually held uncertain information. If a shared environment provides a sufficient medium for mapping individual

uncertainties, efficient information integration should be possible in the absence of verbal communication if minimal means to gradually adjust individual contributions to the joint judgment and negotiate the joint judgment are available. Alternatively, if verbal communication is necessary for successful inter-individual information integration, dyads should exhibit better performance when the opportunity for communication is provided than when it is not provided.

Finally, the third hypothesis addressed the role of feedback. According to the confidence sharing model (Bahrami et al., 2010), feedback is neither necessary nor sufficient for establishing a collective benefit from information integration, because exchange of individual uncertainties is mediated through meta-cognition. In contrast, feedback should play a crucial role if joint judgment formation is mediated by the shared environment. Depriving individuals of immediate feedback about the accuracy of their judgments should hurt the process of information integration and diminish the accuracy of joint judgments.

3.1.7 Deriving models of joint two-dimensional vector judgment formation

In addition to qualitative predictions about how well dyads can integrate their individual perceptions under different conditions, a formal approach to the problem of joint judgment formation allows one to make quantitative predictions on certain statistical properties of joint judgments based on statistical properties of contributing individuals' judgments. These properties can be evaluated in terms of the integration scheme by which individual judgments are integrated into the joint judgment. In this section I provide formal definitions of three models which are two-dimensional extensions of the schemes presented in section 3.1.2, and one additional model derived specifically for the multi-dimensional case. In an attempt to describe *how* groups form judgments, the relative fits of the four integration scheme models to group judgments were then compared. In analogy to the work obtained in the domain of multi-sensory integration (Ernst & Bülthoff, 2004),

we expect joint judgments to become more precise due to integration of information from different people. Our specific predictions derived from the different integration models will focus on predicting the properties of the variability of joint judgments from properties of the variability of judgments belonging to the contributing individuals.

For all the models the following notation is used: individual location judgments are denoted with vectors $\mathbf{j}_i = [x_\mathrm{i}, y_\mathrm{i}]'$ where $i \in \{1, 2\}$ is an individual member of a dyad, and X and Y are the two dimensions of a Cartesian coordinate system. Each member's judgment is modeled as a true location $\boldsymbol{\theta} = [x_0, y_0]'$ with added random error sampled from a bivariate distribution parameterized with mean $\boldsymbol{\mu}_\mathrm{i} = [\bar{x}_\mathrm{i}, \bar{y}_\mathrm{i}]'$ and the following variance-covariance matrix:

$$\boldsymbol{\Sigma}_\mathrm{i} = \begin{bmatrix} \sigma^2_\mathrm{i,x} & \rho_\mathrm{i,xy}\sigma_\mathrm{i,x}\sigma_\mathrm{i,y} \\ \rho_\mathrm{i,xy}\sigma_\mathrm{i,x}\sigma_\mathrm{i,y} & \sigma^2_\mathrm{i,y} \end{bmatrix}, \tag{3.4}$$

where $\sigma_\mathrm{i,d}$ is the standard deviation of random error on dimension $d \in \{x, y\}$, and $\rho_\mathrm{i,xy}$ is the correlation between X and Y errors for individual i. Thus, the mean of the distribution $\boldsymbol{\mu}_\mathrm{i}$ reflects individual perceptual bias for individual i, and Σ_i characterizes variability in individual i's judgments and reflects his or her perceptual uncertainty about the correct target location. Individual judgments are then combined to form the joint judgment with one of the following schemes.

Simple averaging. Just like in the uni-dimensional version of this rule reviewed in section 3.1.2, this rule implies that dyad members are giving equal weights to each other's judgment on both dimensions. The joint judgment is derived by averaging individual judgments of the two dyad members: $\mathbf{j}_\mathrm{joint} = 0.5\mathbf{j}_1 + 0.5\mathbf{j}_2$. The distribution of joint judgments is expected to have the following variance-covariance properties[14]:

$$\boldsymbol{\Sigma}_\mathrm{joint} = 0.25\boldsymbol{\Sigma}_1 + 0.25\boldsymbol{\Sigma}_2 \,, \tag{3.5}$$

[14] Note that coefficients in Eq. (3.5) are equal to 0.25 due to the general statistical property that if a random variable is multiplied by a scalar, its variance is multiplied by that scalar raised to the power 2.

where Σ_1 and Σ_2 is the variance-covariance matrix (Eq. 3.4) of random error characterizing variability in judgments of the first and second dyad member respectively.

Take-the-best (TTB). This scheme prescribes that the joint judgment on each trial is taken from the dyad member with the highest overall precision: $\mathbf{j}_{\text{joint}} = \mathbf{j}_1$. This simply predicts that the distribution of the joint judgments will have the same variability properties as the distribution of the more accurate member's individual judgments: $\mathbf{\Sigma}_{\text{joint}} = \mathbf{\Sigma}_1$, where $(\sigma_{1,x}^2 + \sigma_{1,y}^2) < (\sigma_{2,x}^2 + \sigma_{2,y}^2)$.

Dimension Distribution (DD). This integration scheme has no one-dimensional equivalent (it reduces to the Take-the-best scheme, in fact), and is applicable to multi-dimensional judgments. This scheme assumes that each individual is contributing the more reliable component of his percept to the joint judgments and discards the less reliable component of his percept. In the current task, because participants experience largest uncertainty about the location on their depth dimension, this strategy resembles the intersection method used in land navigation: drawing two rays from two observers' positions to the target and locating the target at the intersection of the two rays (Mooers & Robert, 1972). Under the assumption that individual errors are uncorrelated and that judgments have higher variability on the depth dimension than on the orthogonal dimension[15] predictions on the variability of joint judgments can be approximated as follows.

Denote the larger and smaller eigenvalues of variance-covariance matrices characterizing individual i's uncertainty with $\lambda_{i,1}$ and $\lambda_{i,2}$ respectively, $i \in (1,2)$. Without a loss of generality, we align our coordinate system with eigenvectors of the variance-covariance matrix of random error in the judgments of the first dyad member. In the current task, this essentially implies aligning the Y-axis of our coordinate system with the first dyad member's virtual line of sight. In this case, the variance-covariance matrix of his judgments can be rewritten in the following form:

[15] We tested and confirmed the validity of these assumptions in Study I reported in Chapter II of this manuscript.

$$\boldsymbol{\Sigma}_1 = \begin{bmatrix} \lambda_{1,2} & 0 \\ 0 & \lambda_{1,1} \end{bmatrix}. \tag{3.6}$$

When the two viewpoints are orthogonal the precision dimensions of two dyad members are complementary. Consequently, the Dimension Distribution scheme predicts the following composition of the variance-covariance matrix characterizing joint judgments:

$$\boldsymbol{\Sigma}_{\text{joint}} = \begin{bmatrix} \lambda_{1,2} & 0 \\ 0 & \lambda_{2,2} \end{bmatrix}. \tag{3.7a}$$

A geometric formulation of the same prediction is that joint judgments should be described with a confidence ellipse where the two axes are equal to the minor axes of the confidence ellipses characterizing random error in judgments of the individual members of the dyad.

When the viewpoints are the same as in the redundant condition, the major axis of the confidence ellipse characterizing joint judgments should be equal to the smaller major axis of confidence ellipses characterizing variability in dyad members' individual judgments, and the same for the minor axis. In the formal description, the following composition of the variance-covariance matrix characterizing joint judgments is predicted:

$$\boldsymbol{\Sigma}_{\text{joint}} = \begin{bmatrix} \min(\lambda_{1,2}; \lambda_{2,2}) & 0 \\ 0 & \min(\lambda_{1,1}; \lambda_{2,1}) \end{bmatrix}. \tag{3.7b}$$

If participants apply this scheme in the redundant information distribution condition, their joint judgments are expected to be more precise than the judgments of the more precise dyad member, but generally less precise than joint judgments derived from averaging individual judgments. If participants apply this scheme in the complementary information distribution condition, their joint judgments are expected to be nearly as precise as if individual judgments were optimally integrated (see below).

Multidimensional Optimal Weighing (MOW). This model is the multi-dimensional extension of the Weighted Averaging scheme, where individual judgments are weighed by their reliability. The

derivation of the joint judgment can be written out as the linear combination of individual judgments:

$$\mathbf{j}_{\text{joint}} = (\mathbf{W}_1 + \mathbf{W}_2)^{-1}\,(\mathbf{W}_1\mathbf{j}_1 + \mathbf{W}_2\mathbf{j}_2). \tag{3.8a}$$

This model assumes that participants in a dyad have full access to the structural properties of their uncertainties about the location and combine them in a statistically optimal fashion. This implies that the weight matrix $\mathbf{W}_i$ assigned to a judgment from individual *i* is the inverse of the variance-covariance matrix $\boldsymbol{\Sigma}_i$ that characterizes variability in his individual judgments (van Beers et al., 1999): $\mathbf{W}_i = \boldsymbol{\Sigma}_i^{-1}$. In this way, weight matrix $\mathbf{W}_i$ accommodates all information about the extent of an individual's uncertainty and its geometric properties. If the weights are selected in a statistically optimal way and the Multidimensional Optimal Weighing integration scheme is applied, the precision of joint judgments is expected to be always at least as high as that of the more precise dyad member or higher. The predicted composition of the variance-covariance matrix of joint judgments in this case is

$$\boldsymbol{\Sigma}_{\text{joint}} = (\boldsymbol{\Sigma}_1^{-1} + \boldsymbol{\Sigma}_2^{-1})^{-1}. \tag{3.8b}$$

The model also provides the upper baseline for the level of precision a particular dyad can reach in their joint judgment based on statistically optimal integration of information given the individual uncertainties associated with the two dimensions of the judgment.

3.2 Experiment 1

In the first experiment we investigated how dyads would jointly localize a target on a plain under the condition of simultaneous access to complementary visual information. Dyads were doing the task either with an opportunity for verbal communication or without verbal communication by means of shared control over the response object (the object they were to position in the correct location in a shared virtual environment). We also introduced an Individual Double-perspective (ID)

condition where individuals had sequential access to two viewpoints providing complementary visual information. This allowed us to make a comparison between processes of inter-individual and intra-individual integration of the same information. Feedback on judgment accuracy was available for groups and individuals.

Our main question was whether verbal communication can add anything above and beyond interactions via the shared environment in the process of inter-individual integration of complementary visual information.

3.2.1 Method

Participants

Thirty-two English-speaking students (18 females) aged between 19 and 27 years (M = 22.9) were tested in pairs. The two participants in a pair were always of the same gender. One dyad was replaced due to very poor performance (more than 3 *SD*s in Absolute Error in one of the conditions). Participants in all dyads reported that they had not interacted with each other prior to participating in the experiment.

Material and apparatus

The shared virtual environment was presented using two Apple iMac computers (2.5 GHz Intel Core i5 iMac computers with 21.4" Display and AMD Radeon HD 6750M 512 MB graphics) and an additional external monitor (BenQ RL2240H 21.5, connected to one of the Macs). The screen resolution was 1600 X 900 pixels on both displays and they were calibrated to have matched color output. Thrustmaster T16000M ambidextrous joysticks were used as input devices with which participants provided their location judgments. The virtual environment for the location task was generated using the perspective mode of the MatLab (The Mathworks, Inc.) software. The main element of this environment was a two-dimensional projection of a three-dimensional square-based

rectangular cuboid with a length × width × height ratio of 2×2×1. The cuboid was centered on the screen area (263×204 mm) and simulated a real-world cuboid (256×256×128 mm). The target appeared in the bottom plane of the cuboid with the constraint that it could not be closer to the edges of the plane than 1/5 of the length of each edge of the plane.

The visible length of the sides of the cuboid on the screen varied as a function of the virtual camera orientation relative to the cuboid following the laws of perspective projection. Different viewpoints on the cuboid were generated by varying properties of the camera orientation as defined by azimuth (polar angle in the x-y plane) and elevation (angle above and below the x-y plane). Throughout the experiment, elevation was kept constant at 14°. The azimuth values were varied to generate different viewpoints on the layout. We used two sets of viewpoints in the experiment, each of which included two azimuth values that were used to create pairs of viewpoints with 90° angular difference for the Joint condition and the Individual Double condition. The two viewpoint sets with their corresponding azimuth values are listed in Table 4. A schematic depiction of the virtual environment with two dyad members observing the shared layout from two orthogonal viewpoints (selected from Set 1) is provided in Figure 14. The camera view angle was set to a constant 6.3° to minimize angular distortions in the scene. The distance between cuboid center and virtual camera was held constant.

The proportion of color components for the fillings of the upper and lower planes of the cuboid in RGB code were [.1 .8 .8] and were semi-transparent. The side planes of the cuboid were fully transparent. The edges defining one of the side planes of the cuboid were colored green (line width 1.5 points), all other edges were colored black (line width 1 point). This served to provide cues with regard to the orientation of the cuboid to support participants in tracking the relation between different views on the cuboid.

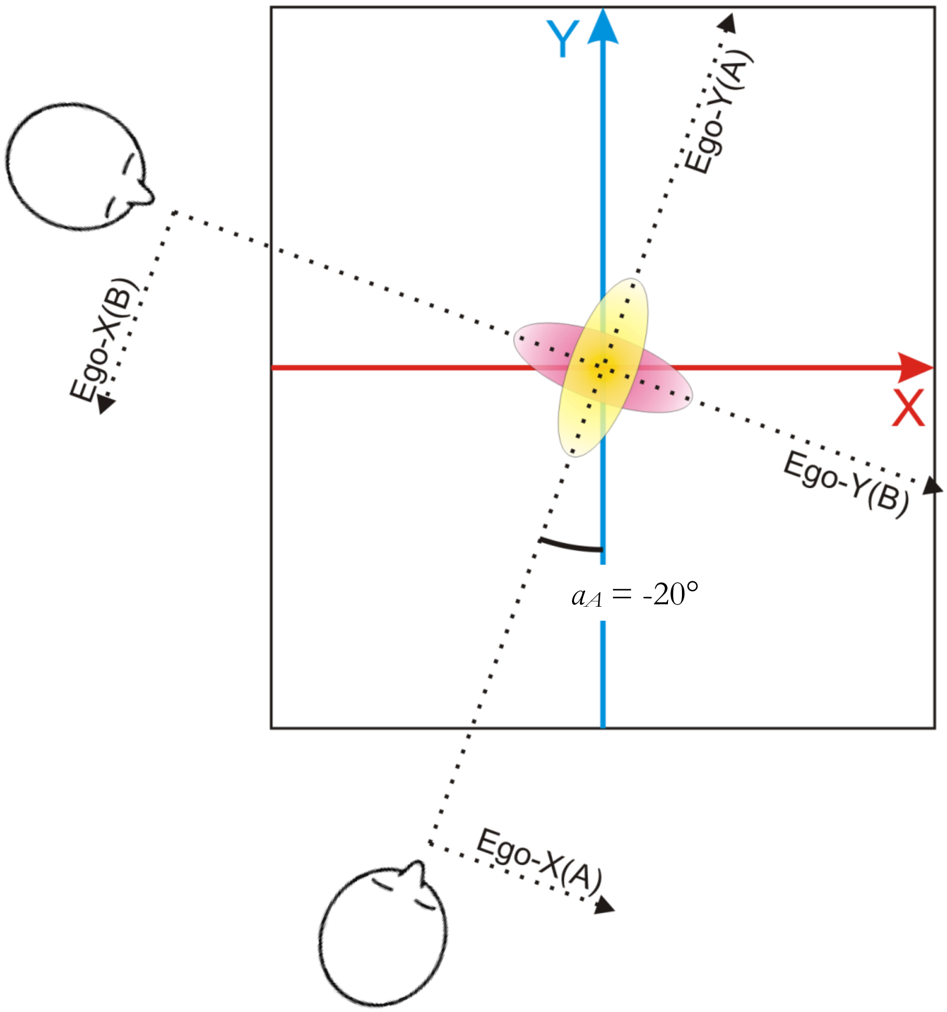


Figure 14. A schematic representation of the top-down view on the virtual layout with two orthogonal viewpoints. The colored solid lines represent the intrinsic coordinates of the cuboid; the dashed lines represent participants' respective egocentric coordinate systems. The two ellipses visualize idealized uncertainty distributions for the two participants with respect to the correct location of the pointer. Ego-Y (depth) dimension coincides with participant's virtual line of sight, and is the one where they experience largest uncertainty about the pointer's correct location. In this illustration the azimuth value of the viewpoint provided to dyad member A is -20°, and for dyad member B it is -110°.

Table 4. Viewpoint sets and their corresponding azimuth values used in experiments 1 and 4 (Study II).

	Set 1	Set 2
Viewpoint 1	-110	-70
Viewpoint 2	-20	20

Design

The experiment involved two within-subject factors. The first experimental factor was the mode in which participants were performing the object location task. In the Joint (J) condition each of the two individuals in a dyad had simultaneous access to one of two orthogonal viewpoints, and the two individuals provided a jointly agreed location judgment. In addition to the Joint condition we introduced an Individual Double-perspective (ID) condition where individual participants had sequential access to both orthogonal viewpoints by switching between them. This condition tested how well participants could integrate information from the two viewpoints intra-individually.

The second experimental factor was the availability of verbal communication during joint task performance (J). In the Communicative condition participants were allowed to talk to each other via a headset, in the Non-communicative condition there was no possibility for verbal communication.

Procedure

The two participants in a pair were seated in different rooms with separate monitors and joysticks throughout the whole experiment. At the beginning they were individually provided with a video instruction about the task, and then the experiment started. The main experiment was run in two sessions. Each session involved a block of 20 J trials and a block of 20 ID trials. In one of the sessions communication was allowed in the J condition and in the other part communication was not allowed in the J condition.

Before and after the main experiment each individual in a pair was asked to provide a location judgment from one particular viewpoint (individual single, IS, 20 trials before and 20 trials after). The individual single judgments collected at the end of the experiment (IS) were used as a baseline for the ID and J conditions and reflect an estimate of participants' individual performance

limit after being familiar with the location task. The whole experiment took about 2 hours for one dyad.

Each of the two pre-defined pairs of viewpoints (see Table 4) was exclusively assigned to one part of the experiment. Within a given part, two individuals from a given dyad were provided with orthogonal viewpoints in the Joint condition. Each individual performed the IS trials of the respective part from the same viewpoint as the one occupied in the J condition. In the ID condition participants switched between the two viewpoints from a set.

The order of the Communicative and Non-communicative part, the order of the task modes (J and ID), and the assignment of viewpoint pairs were fully counter-balanced across participants.

At the beginning of each block (J, ID in the communicative and non-communicative part) participants received detailed instructions about their task in the ensuing block. Once the instructions were understood participants performed 20 trails of the corresponding experimental condition. Each trial started with displaying the pointer in one the four randomly selected corners of the cuboid's upper plane together with a new randomly positioned target on the cuboid's bottom. Participants had 120 s to locate the target by moving the pointer with the joystick. Starting 10 s before the end of the trial warning beeps sounded each second if no judgment had been provided yet. Pilot experiments had shown that joint as well as individual judgments could be comfortably made within this time. After the judgment was delivered or after the time limit was reached, participants received immediate feedback in the form of a colored line connecting the apex of the pointer with the bottom plane of the cuboid. This line was visible for 5 s.

In the J condition participants read the instructions independently from their own screen (presentation of the instructions was synchronized, and was delayed if one of the partners was not yet finished with her preceding condition). Once both participants had indicated that they were ready they visited each other's rooms to look at their partner's setup. This served as a reminder

about the two different viewpoints they had on the same virtual environment. When they returned to their own room the first joint trial started. The two partners observed the same layout each from his own visual perspective, and jointly controlled the pointer. The inputs from the two joysticks were summed up: that means if both partners tilted their joysticks by the same angle in the same direction, the pointer would move double the distance compared to when one partner tilted the joystick by the same angle. If a participant judged that the pointer was at the correct location, he pressed the joystick button. At this moment a "DONE" caption appeared in a small box in the left upper corner of both partners' displays ("Player1: DONE"). The pointer turned grey and was disabled for control for 5 s. During those five seconds the partner could accept the proposal by pressing the joystick button and the judgment was taken for that trial with immediate feedback. If the proposal was not accepted within 5 s, the pointer was activated again, and moving it cancelled the DONE status, so that a new proposal had to be made. A proposal could be accepted also after the disability period if the pointer was not moved. If the two partners did not agree on the pointer's location within the time limit, the pointer's last position before the end of the trial was taken as the joint judgment on that trial. Depending on its status a judgment was coded as either "consensual", "proposed" (if at the end of the trial only one partner was in the DONE status), or "divided" (no partner was in the DONE status at the end of the trial). In the communicative J block the partners talked to each other via headsets with no constraints on the content of their communication. In the non-communicative J block headsets were disabled, and partners needed to come to an agreement without verbal communication.

In the two blocks of the ID condition individual participants could switch between the two viewpoints and thus make a judgment based on sequentially integrated information. The initial perspective was always the one participants worked from in the IS and J conditions of the respective experimental part. Participants triggered the camera rotation around the cuboid to another

viewpoint by pressing “z” and “x” keys on the keyboard. After the viewpoint rotation the joystick control was remapped accordingly. There was no limit on the number of switches per trial. We introduced an opaque confirmation window, which appeared when participants pressed the joystick button and obscured most part of the layout, to avoid accidental judgment submissions caused by subjects’ confusion of the control buttons. When the confirmation window was active the pointer control was disabled, and participants had to press the “y” key to confirm their judgment or “n” to continue locating the target. There was no procedural difference for the ID mode in the Communicative and Non-communicative parts of the experiment.

In the IS condition participants provided their judgment from one viewpoint. A single press on the joystick button was required for the location judgment.

Dependent measures

The dependent measures fall into different categories. A first set of measures was directly related to performance, characteristics of the location judgments and their derivatives. In order to determine *how well* dyads form their judgments, the accuracy of joint judgments was compared to that of individual judgments. To determine which integration scheme provides a more accurate description of the process of joint judgment formation, the variability of joint judgments was compared to the variability of joint judgments predicted by the candidate models based on dyad members’ individual performance. The latter analysis addresses the question of *how* individuals integrate individually perceived information into joint judgments.

Further measures were not directly related to participants’ performance in the task. These included the average time taken to make a judgment and the number of individual location proposals per target needed to reach an agreement in the Joint condition. We collected and analyzed these data

to gain insight into how participants were modifying their information integration strategies under different conditions, which might not be directly reflected in their performance.

3.2.2 Data preparation

The raw data obtained were the spatial coordinates of participants' location judgments coded as two-dimensional coordinates in the cuboid's x-y plain. From these coordinates we computed absolute error (absolute error), the Euclidean distance between the judgment coordinates and the true coordinates (see Figure 2A on p. 58). The mean of absolute error provides a measure of the accuracy of participants' judgments.

The directional error on a specified axis was simply coded as the distance between the target coordinate and the judgment coordinate (negative for undershoot and positive for overshoot). Directional error confounds directional biases due to sensorimotor calibration and variability in the judged coordinates.

To obtain a variability measure of participants' location coordinates that reflects participants' uncertainty (Gigone & Hastie, 1997) we removed the bias component from the directional error. In the present experiment participants had a general bias to overshoot on the Ego-Y axis resulting in a positive bias on their depth dimension. This corresponds to previous findings on individual performance in tasks that required participants to judge depth in the visual modality (e.g. van Beers et al., 1996, 1998, 1999), and is thought to occur when participants observe a visual layout under a slant angle.

We also removed variability that resulted from participants' gradual learning from feedback to compensate for this bias to overshoot, which results in a drift of the geometrical center of their judgments within a particular condition. In order achieve this we subtracted variance resulting from such drifts from the total variance of judgments (e.g. van Beers et al., 1996, 1998, 1999), which also

removes general bias. We first calculated the best linear fits for the egocentric *x* and *y* coordinates for all trials in a particular condition as a function of the serial number of a trial. The fits were determined using the method of least squares. This procedure was performed for each participant in the IS and ID conditions, and for each dyad in the J condition. The drifting center coordinates were then subtracted from the raw egocentric coordinates to derive the unbiased coordinates. We applied Chauvenet's criterion (Taylor, 1997, p. 166-168) to the unbiased coordinates to identify errors that resulted from accidentally pressing the joystick button. According to this criterion, in a sample consisting of *n* data points, a data point may be discarded as a measurement error if the probability of obtaining the particular deviation from the mean is less than $1/(2n)$. The criterion was applied separately for each participant (or dyad in the J condition) in each condition. After removing the data points that were identified as outliers, unbiased coordinates were re-calculated.

Because the location judgments in our task are described with bivariate distributions, we adopted an established measure for assessment of variability for the two-dimensional spatial data: the area under the standard deviational (SD) ellipse (Ebdon, 1988; Kent & Leintner, 2007). This is an ellipse in which the semi-axes are equal to the square roots of the eigenvalues of the sample variance-covariance matrix. We derived the SD ellipses from the unbiased coordinates and then computed their area as a scalar measure of judgment variability that captures participants' uncertainty about the target location in two dimensions. Because an area is a quadratic measure it results in skewed data distributions on a linear scale. Therefore, we used the square root of the SD ellipse area as the dependent variable in our analyses of judgment variability.

The individual performance parameters of participants in the IS and ID condition belonging to one dyad were averaged when comparing them to the performance of the corresponding dyads in the J condition.

3.2.3 Results

Participants came to a consensual joint judgment in 99.35% of trials in the condition with communication, and in 99.68% of trials without communication. Data where dyads did not reach a consensual judgment and trials where a judgment had not been made before the time limit were excluded before further data preparation and analysis. After excluding these trials and outlying data points 97.1% of original data were retained for the ensuing analyses.

First, we analyzed absolute error of the location judgments (see Figure 15). Performance in the ID and J conditions was compared by means of a two-way repeated-measurement analysis of variance (ANOVA) with the factors Condition (ID vs J) × Communication (Communication vs No-Communication). This test showed no significant main effects or interactions between the two factors (ps >.179). Therefore, we collapsed the data from the communicative and non-communicative block for the ensuing analyses. A comparison of absolute error in the J condition with absolute error in the IS condition revealed that absolute error in the J condition (M = 7.14 mm, SD = 4.60 mm) was significantly lower than in the IS condition (M = 21.1 mm, SD = 6.18 mm), $t(15) = -11.0$, $p < .001$. Absolute error in the ID condition (M = 8.92 mm, SD = 5.24 mm) was also significantly lower than absolute error in the IS condition, $t(31) = -8.21$, $p < .001$.

Furthermore, we compared the absolute error of dyads in the J condition with the absolute error of the better individual in the dyad (i.e., the individual with lower absolute error, see Figure 16 on p. 150). Performance in the Communication and No-communication condition was compared separately. Paired-sample t-tests showed that dyads achieved lower absolute error than the more accurate individuals in the IS condition (M = 16.0 mm, SD = 5.08 mm) both with communication (M = 7.26 mm, SD = 5.68 mm), $t(15) = -5.81$, $p < .001$, and without communication (M = 7.03 mm, SD = 4.28 mm), $t(15) = -6.82$, $p < .001$.

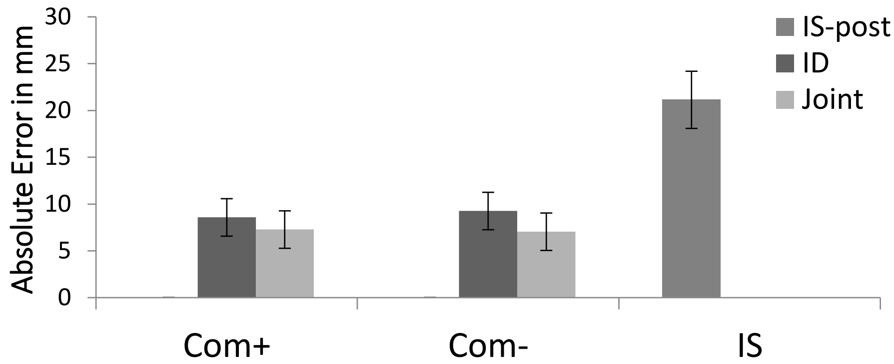


Figure 15. Results of Experiment 1 (Study II): absolute error. “Com+” shows the results for the Communication condition and “Com-” shows the results for the No-communication condition. Error bars in Com+ and Com- stand for within-subjects 95% confidence intervals of the respective 2 × 2 ANOVA. Error bar in the IS condition is the 95% confidence interval.

Next, the variability of location judgment errors in the ID and J conditions was compared by means of a two-way repeated-measurement analysis of variance (ANOVA) with the factors Condition (ID vs J) × Communication (Communication vs No-Communication). This test revealed no significant main effects or interactions between the two factors (*p*s > .381). Therefore, we collapsed the variability data from the Communicative and the Non-communicative block. Variability of judgment errors in the ID condition (M = 8.05 mm, SD = 3.73 mm) was significantly lower than in the IS condition (M = 12.7 mm, SD = 3.74 mm), $t(31) = -6.08$, $p < .001$. Variability of judgment errors in the J condition (M = 7.24 mm, SD = 5.57 mm) was also significantly lower than in the IS condition, $t(15) = -6.15$, $p < .001$.

In the Communicative condition subjects required fewer proposals of target locations to reach an agreement (M = 1.38, SD = 0.71) than in the Non-communicative condition (M = 1.84, SD = 0.61), $t(31) = -3.67$, $p = .001$.

The average time required to complete a trial was analyzed by means of a repeated-measurement ANOVA with factors Condition (ID vs J) and Communication (Communication vs No-Communication). This test revealed neither significant main effects nor an interaction between the two factors (*p*s > .224).

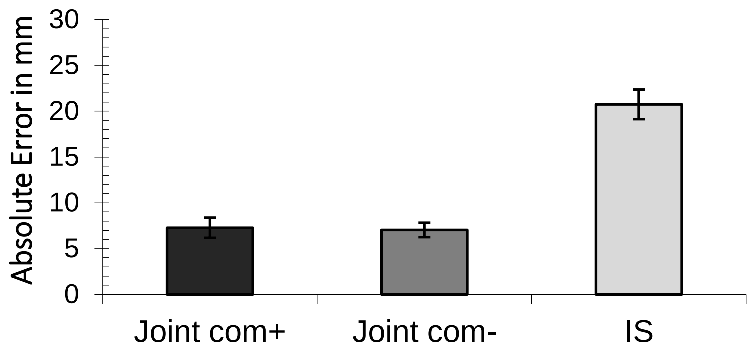


Figure 16. Results of Experiment 1 (Study II): comparison of performance between the better individual from a dyad and the joint performance from that dyad. “Com+” – Communication condition, “Com-” – No-communication condition. Error bars are SE of the means.

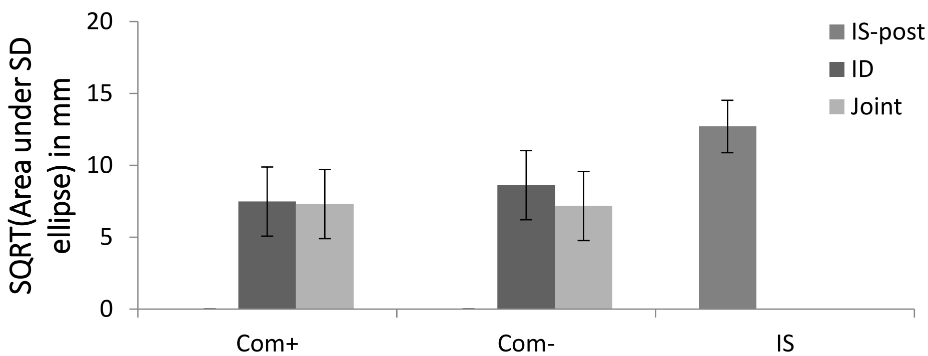


Figure 17. Results of Experiment 1 (Study II): variability of judgment errors. “Com+” shows the results for the Communication condition and “Com-” shows the results for the No-communication condition. Error bars in Com+ and Com- stand for within-subjects 95% confidence intervals of the respective 2 × 2 ANOVA. Error bar in the IS condition is the 95% confidence interval.

Model Comparisons

We used individual judgments provided from one viewpoint at the end of the experiment (IS) to generate model predictions for the variability of errors in joint judgments resulting from different integration schemes of individual percepts into a joint judgment.

Prior to testing how well our models fit empirical data we analyzed whether participants’ judgments can be accurately modeled with bivariate normal distributions, which is a necessary condition for the validity of our models’ predictions. To do so we tested with Henze-Zirkler's

Multivariate Normality Test (Henze & Zirkler, 1990; Trujillo-Ortiz, Hernandez-Walls, Barba-Rojo & Cupul-Magana, 2007) whether each individual's distributions of unbiased errors of judgments obtained in the IS condition violated the assumption of a bivariate normal distribution. Only for four of our 32 participants were violations of bivariate normality significant with *p* values falling in a range between .049 and .013 that would not survive an adjustment for multiple comparisons. We therefore conclude that, generally, our assumption of multivariate normality holds for individual error data, and that our integration models are applicable to the data obtained.

An example of the comparison between variability of error in empirical joint judgments and that predicted by the candidate models is provided in Figure 19 on p. 153. Figure 19A displays the unbiased errors of individual judgments for two typical dyad members in a coordinate system aligned with one individual's virtual line of sight. Closed curves are 95% confidence ellipses characterizing variability of errors in the two individuals' judgments. Figure 19B displays predictions from different integration models for the variability of joint judgments and the empirically observed joint judgments of the dyad in the J condition. Black closed curves are 95% confidence ellipses characterizing variability of errors in individual judgments, the same as on the top figure. The different colored 95% confidence ellipses correspond to the variability in joint judgments predicted by different models. For the depicted dyad most of the empirically observed judgment errors in the J condition (each colored diamond corresponds to one judgment) fall within the inner white ellipse that corresponds to the predictions of the optimal information integration scheme.

Because we found no difference in performance measures between the Communication and the No-communication condition, square roots of SD ellipses were averaged across the two conditions. Pair-wise *t*-tests showed that variability of errors in joint judgments was smaller than predicted by the Simple Averaging model, $t(15) = -8.85$, $p < .001$; and smaller than predicted by the Take-the-Best model, $t(15) = -3.21$, $p = .006$. At the same time, it was larger than predicted by the

Multidimensional Optimal -Weighing model, $t(15) = 2.69$, $p = .01$, and larger than predicted by the Dimension Distribution model, $t(15) = 2.53$, $p = .023$.

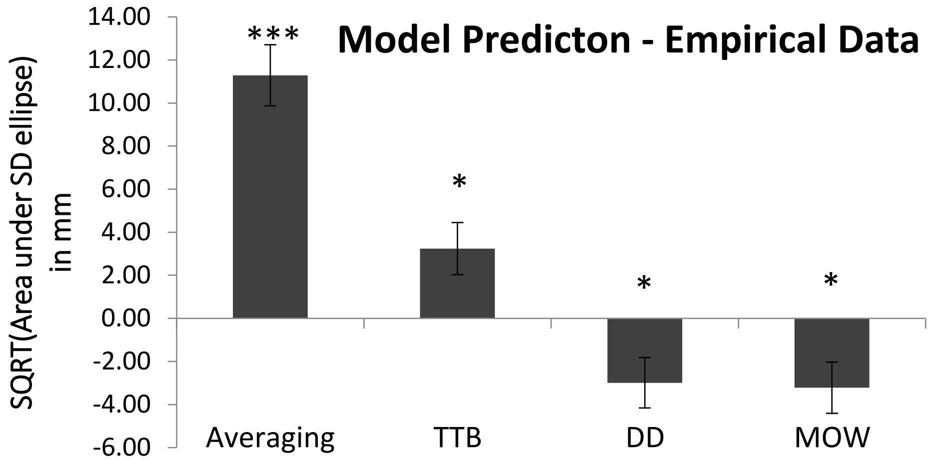


Figure 18. Results of Experiment 1 (Study II): comparisons with model predictions. Average differences between the predictions for the area of the SD ellipse from different integration schemes and empirical data obtained from dyads in the joint judgment condition J of Experiment 1. Asterisks indicate whether these differences were significantly different from 0 (paired samples *t*-test). Error bars stand for the standard error of the mean difference. Averaging – *Simple Averaging*, TTB – *Take-the-Best*, DD – *Dimension Distribution*, MOW – *Multidimensional Optimal Weighing*.

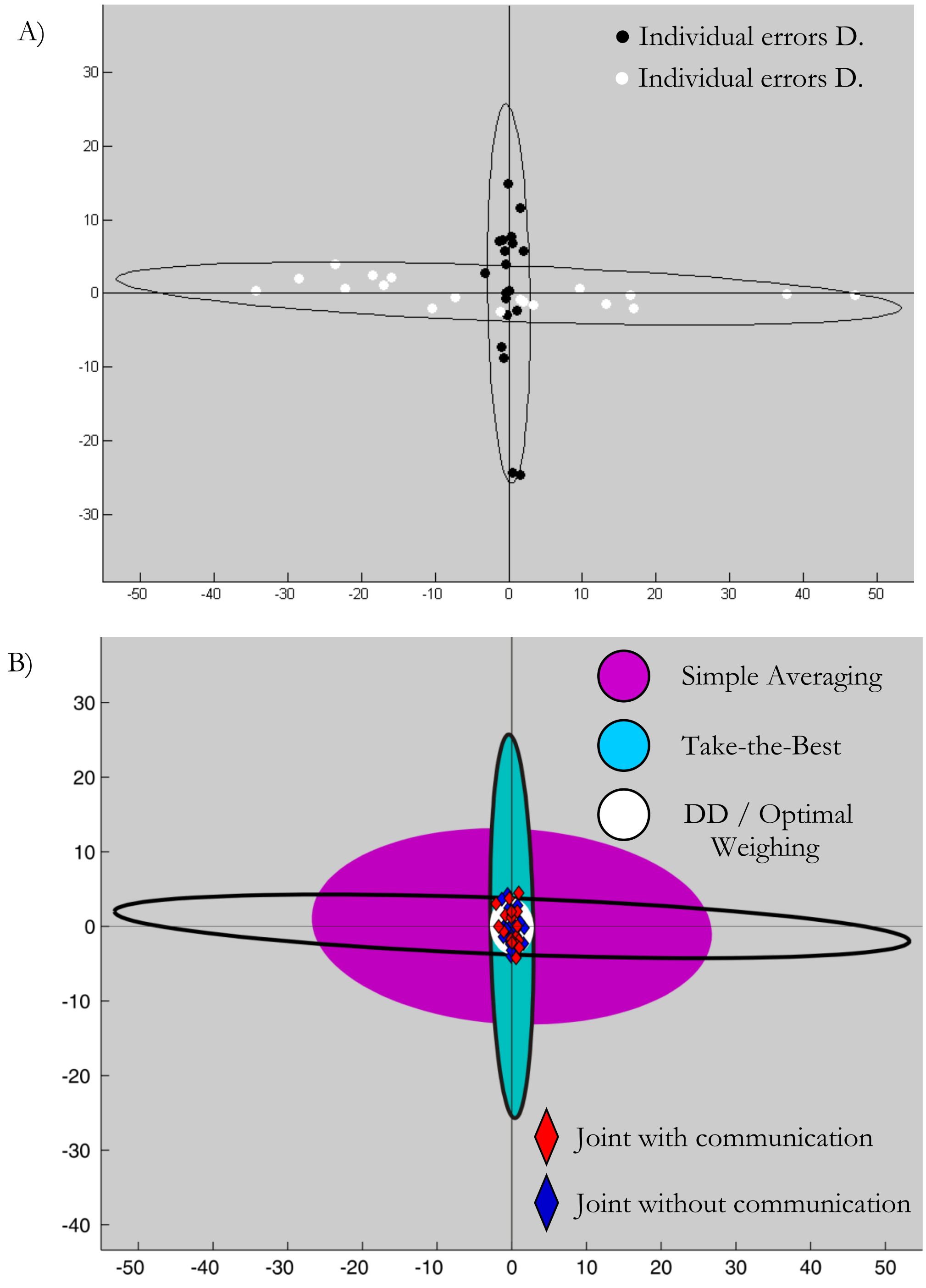


Figure 19. Model fits to empirically observed variability in joint judgment errors for a representative dyad. A) Errors of judgments (in mm) collected in the Individual condition from two members of one dyad in the first member's egocentric coordinate system (the ordinate axis is aligned with the virtual line of sight). The two ellipses are the 95% confidence ellipses for the two individuals. B) The 95% confidence ellipses predicted by different integration schemes. The ellipses with black contours illustrate the expected bivariate distributions of judgment errors if either of two subjects was solely responsible for joint judgments. The confidence ellipses predicted by the Dimension Distribution and Optimal Weighing models almost completely overlap and therefore are not plotted

separately. Colored diamonds represent the empirically observed errors of the dyad judgments in the Joint condition with and without communication.

3.2.4 Discussion

The results clearly show that pairs of individuals took the advantage of being able to observe a shared environment from two different viewpoints when jointly locating objects. They integrated information available to them to form joint judgments which were remarkably more accurate than individual judgments provided from one viewpoint. Moreover, the results provided evidence that dyads achieved higher accuracy than the better individual in the dyad by means of environment-mediated interactions when verbal communication was not possible.

Dyads' performance was as accurate as performance of individuals who had access to both viewpoints, which means that interpersonal integration of information was as efficient as with-individual integration of the same information. The time needed to provide a judgment was the same in the joint condition and the condition where individuals could switch between different perspectives. The opportunity to verbally communicate did not affect processing time either.

Availability of verbal communication did not affect dyads' performance: dyads were not less accurate when they were silent than when they were talking. There was a difference, though, in the number of proposed locations during a trial. In the absence of verbal communication, individuals made more use of rejecting another's proposed judgments before agreeing on a joint judgment. This suggests that minimal means for negotiating the collective judgment can be sufficient to substitute verbal communication, at least in a situation where individuals need to contribute different judgment components. In this situation strong resistance of one individual to compromise his most preferred position on a particular dimension would convey information to the partner that the first individual is much more reliable and should dominate on this dimension. This process of information exchange through differential propensity to compromise one's own most preferred position is

compatible with Graesser's (1991) Proportional Compromise model for bargaining and decision-making based on her theorem of Social Averaging. According to this model, firm rejections of compromise positions by one group member would convey important information to other members on the strength of the first member's preference. In application to the situation of collective judgments under uncertainty as in our task, this would mean that firm rejections of location offers by a dyad member would convey some information on that member's expertise.

The extent to which dyads reduced variability of their judgments compared to the individual estimations of their members enabled us to make quantitative predictions about whether different integration schemes can explain the observed results or whether some of them can be ruled out. We can with confidence reject the Simple Averaging model as well as the Take-the-Best model because the empirically observed variability in joint judgments was considerably smaller than the variability predicted by these models. Thus we have strong evidence that both individuals contributed to the joint judgment. However, the reduction in variability was smaller than predicted by the statistically optimal or near optimal Dimension Distribution model.

Several possible explanations can be provided for the observed sub-optimality of joint judgments. In addition to general failures to integrate information properly (Wallsten, Budescu, Erev, & Diedrich, 1997) and potential social loafing (Latané, Williams, &Harkins, 1979) in some of the participants, there are also some factors that may specifically impair performance in the present task. One is a potential increase of response noise resulting from joint control over the pointer and the task to coordinate two joysticks, which could potentially impede providing fine-grained judgments (Wahn et al., 2016). The second specific factor that may make us underestimate individuals' real ability to integrate individual percepts to form a joint judgment is the fact that the individual data the model predictions were based on was collected in the last phase of the experiment when individuals had already had a chance to improve their task-related abilities. As a

result, the models may overestimate the precision of individual internal estimations that were integrated into joint judgments during the earlier joint phases of the experiment. We will find some evidence supporting this interpretation in further experiments reported here.

3.3 Experiment 2

In the second experiment we tested whether dyads would benefit from integrating individual information when the individuals in the dyad receive the same perceptual input. Under these conditions, if the abilities of two paired individuals are equal or near-equal, simply averaging the two individual location estimates would be sufficient to reduce the random error of the joint judgment, making it more accurate.

We hypothesized that interacting via the environment should be sufficient for dyads to arrive at a compromise. By the Social Averaging theorem (Graesser, 1991) this should have a statistical effect of averaging on the joint judgments. Thus, we predicted that joint judgments should be more accurate than judgments of an average individual member. If individuals are able to coordinate their individual contributions to the joint judgment proportionally to their precision, they should achieve collective benefit from the integration.

This experiment also allowed us to test a hypothesis based on the theory by Bahrami et al. (2010) that depriving people of verbal communication will hinder the process of inter-individual information integration because it will preclude individuals from communicating their confidence about the perceptual input. If this hypothesis is true, dyads would not obtain collective benefit when deprived from the opportunity for verbal communication, and would show deteriorated performance compared to the Joint condition with Communication.

Finally, by comparing the results of Experiment 2 with the results from Experiment 1, we tested a hypothesis derived from the theorem of Maximum Collective Benefit. The hypothesis stated

that collective benefit should depend more on the difference in localization precision between two partners under conditions of redundant visual input than under conditions of complementary visual input. Thus, the correlation between inter-individual precision ratio and collective benefit obtained by a dyad in the joint Condition should be stronger in Experiment 2 than in Experiment 1.

3.3.1 Method

Participants

Thirty-two English-speaking students (14 males) aged between 20 and 26 years (M = 22.1) were tested in pairs. The two participants in a pair were always of the same gender. Two additional dyads were replaced due to poor performance (accuracy lower than 3 *SD* from sample average).

Materials and apparatus

These were the same as in Experiment 1 except that only two viewpoint azimuth values (110° and 20°) were used, one in the communication condition and one in the no-communication condition. The order of the two viewpoints was counter-balanced across participants.

Procedure

The procedure was the same as in Experiment 2 except that individuals belonging to one dyad were assigned to the same viewpoint (which depended on the experimental part) in each condition, and thus shared identical perceptual input in the Joint condition.

In the Individual Double condition participants were provided with only one viewpoint which made this condition practically equal to the Individual Single condition in terms of the perceptual information that was available for one individual. However, the same judgment-

confirmation dialog appeared upon pressing the joystick button as in Experiment 1 to make the ID conditions in the two experiments procedurally similar.

3.3.2 Results

Participants came to a consensual joint judgment in 98.44% of trials in the No-communication condition and in 100% of trial in Communication condition. After excluding non-consensual trials, trials where the response was not made within the time limit, and outliers, 97.56% of original data was retained for the ensuing analyses.

We first analyzed absolute error of the location judgments. Performance in ID and J conditions was compared by means of a two-way repeated-measurement analysis of variance (ANOVA) with the factors Condition (ID vs J) × Communication (Communication vs No-Communication). The test revealed a significant main effect of Condition, $F(15) = 11.6$, $p = .004$, partial $\eta^2 = .437$. Absolute error in the J condition ($M = 15.8$ mm, $SD = 7.35$ mm) was smaller than in the ID condition ($M = 18.5$ mm, $SD = 6.65$ mm). The effect of Communication was not significant ($p = .221$) and there was no interaction between the two factors ($p = .376$), consequently the data was collapsed across the communication and non-communication block for the ensuing analysis. The difference in absolute error between the IS and the J condition was not significant ($p = .452$). Absolute error in the IS condition ($M = 16.5$, $SD = 6.55$ mm) was significantly lower than in the ID condition, $t(15) = -2.04$, $p = .049$.

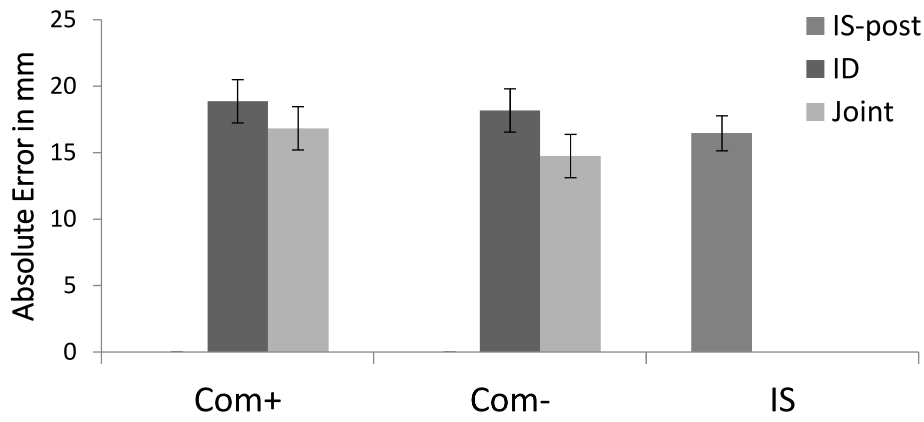


Figure 20. Results of Experiment 2 (Study II): absolute error. "Com+" – Communication condition, "Com+" – No-communication condition. Error bars for Com+ and Com- stand for within-subjects 95% confidence intervals of the respective 2 × 2 ANOVA. Error bar in the IS condition displays 95% confidence interval.

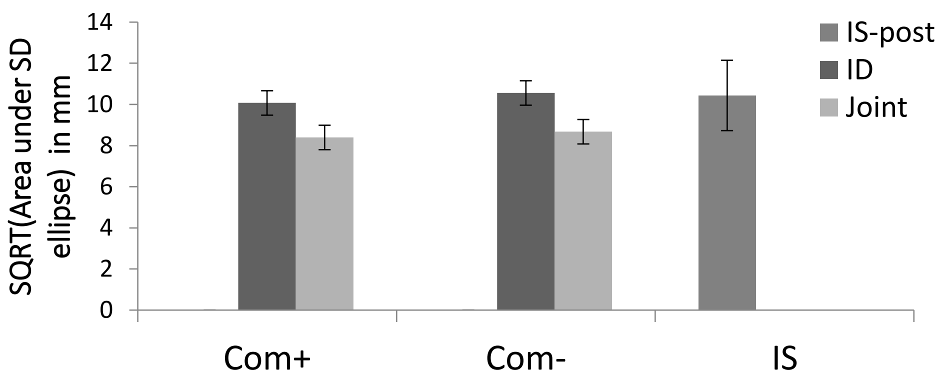


Figure 21. Results of Experiment 2 (Study II): variability of judgment errors. "Com+" shows the results for the Communication condition and "Com-" shows the results for the No-communication condition. Error bars in Com+ and Com- stand for within-subjects 95% confidence intervals of the respective 2 × 2 ANOVA. Error bar in the IS condition is the 95% confidence interval.

Next, the variability of errors in location judgments in the ID and the J condition was compared by means of a two-way repeated-measurement analysis of variance (ANOVA) with the factors Condition (ID vs J) × Communication (Communication vs No-Communication). The main effect of Condition was significant, $F(15) = 13.9$, $p = .002$, partial $\eta^2 = .481$. Variability of judgments was smaller in the J condition ($M = 8.53$ mm, $SD = 2.59$ mm) than in the ID condition ($M = 10.3$, $SD = 3.03$ mm). The main effect of Communication was not significant ($p = .289$) and

there was no interaction between the two factors (p = .724), consequently we collapsed the data across the communication and no-communication block. Variability of judgment errors in the J condition was significantly lower than in the IS condition (M = 10.2 mm, SD = 3.37 mm), t(15) = 3.46, p = .003. There was no difference in variability of judgment errors between the ID and IS condition (p = .886).

Although variability of judgment errors in the J condition was on average larger than variability of judgment errors of the more precise dyad members in the IS condition (M = 7.91 mm, SD = 3.0 mm), it should be mentioned that 7/16 dyads reached higher level of precision (smaller area under the SD ellipse in the J condition) than that of the more precise dyad member.

There was no difference in the number of judgment proposals between the Communication and the No-Communication condition (p = .141). The mean number of proposals per trial[16] was 0.93, SD = 0.47.

The average time required to complete a trial was analyzed by means of a two-way repeated-measurement analysis of variance (ANOVA) with the factors Condition (ID vs J) × Communication (Communication vs No-Communication). The analysis revealed a main effect of condition, F(15) = 16.8, p = .001, partial $\eta 2$ = .529, a main effect of Communication, F(15) = 6.79, p = .020, partial $\eta 2$ = .312, and a significant interaction between the two factors, F(15) = 10.9, p = .005, partial $\eta 2$ = .421. RTs were slower in the J condition (M = 41.3 s, SD = 19.9 s) than in the ID condition (M = 25.6 s, SD = 6.91 s). A paired-sampled t-test showed that RTs were slower in the Communicative J condition (M = 46.6 s, SD = 22.2 s) than in the Non-communicative J condition (M = 36.0 s, SD = 17.4 s), t(15) = 3.24, p = .005.

[16] Note, that it is not necessary that both dyad members make a proposal on every trial. Minimally, it would be sufficient that one member is making the proposal, and the second member accepts it. In this case, no proposals will be counted for the second participant on this trial.

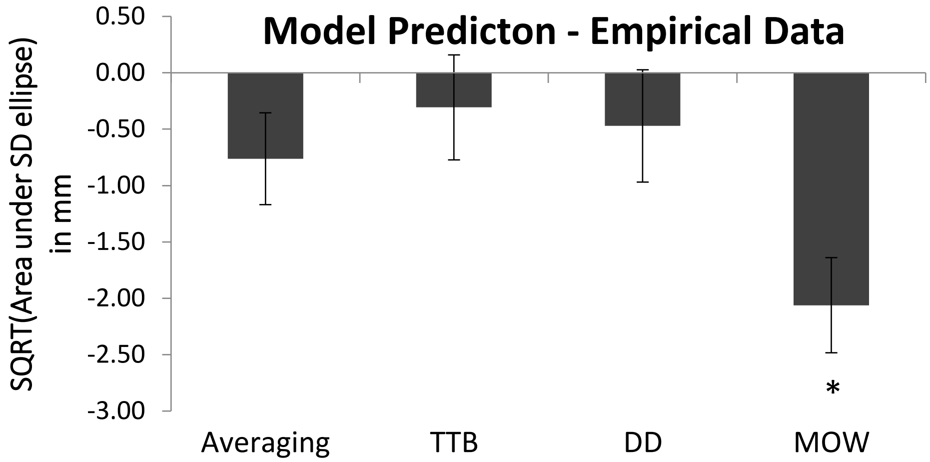


Figure 22. Results of Experiment 2 (Study II): comparisons with model predictions. Average differences between the predictions for the area of the SD ellipse from different integration schemes and empirical data obtained from dyads in the joint judgment condition of Experiment 2. Asterisks indicate whether these differences were significantly different from 0 (paired samples *t*-test). Error bars stand for the standard error of the mean difference. Averaging – *Simple Averaging*, TTB – *Take-the-Best*, DD – *Dimension Distribution*, MOW – *Multidimensional Optimal Weighing*.

Model Comparisons

Since we had clear evidence of learning taking place throughout the experiment (lower error in the IS condition than in the ID condition), the parameters of participants' performance derived from the IS condition at the end of the experiment provide a conservative estimate of participants' individual location precision in the preceding experimental parts. Because in this experiment the ID condition essentially probed participants' location accuracy provided with perceptual input from a single viewpoint, we took variability properties of judgments collected in this condition to estimate participants' individual localization precision in the J condition of the respective experimental part. These estimates were used as inputs to generate predictions for the variability of errors in joint judgments that would be expected from different integration schemes.

Because there was no difference in performance measures between the Communication and the No-communication J conditions, the empirical data and model predictions were collapsed across the two conditions. Variability of errors in joint judgments was larger than predicted by the

Multidimensional Optimal Weighing model, $t(15) = 4.89$, $p < .001$. All tests comparing empirical data with predictions derived from other models revealed non-significant differences (ps > .08).

Comparison with Experiment 1

We further tested one of the predictions following from the theorem of the Maximum Collective Benefit: the quality of joint judgments should depend less on inter-individual absolute differences in abilities under the condition of complementary visual input. To address this question, we analyzed how well difference in individual accuracy between two dyad members can predict the amount of collective benefit. We computed the collective benefit (Bahrami et al., 2010; Wahn et al., 2016) in judgment accuracy as the ratio of the more accurate partner's absolute error in the IS condition to the average absolute error of the dyad in the Joint condition. We then computed an index of inequality in accuracy between the two individuals in each dyad as the ratio of absolute error of the more accurate partner to absolute error of the less accurate partner. This analysis revealed that there was a significant strong correlation between collective benefit and inter-individual inequality in accuracy in Experiment 2, Pearson's $r(16) = .995$, $p < .001$; but not in Experiment 1, $r(16) = .163$, $p = .547$. The difference between the two correlation coefficients was significant, $z = 7.22$, $p < .001$.

3.3.3 Discussion

As in the case of collaboration performed from orthogonal viewpoints, participants took advantage of the opportunity to integrate individual perceptions when they shared the same viewpoint. This is evident because the accuracy of joint judgments was higher than the average individual accuracy in the ID condition. This effect demonstrates that even under conditions of

access to redundant information dyads could reduce some random error by integrating their individual perceptions.

In contrast to previous findings which pointed to a crucial role of verbal communication in the process of inter-individual integration of perceptual information (Bahrami et al., 2010), in our experiment the opportunity for communication did not have any effect on accuracy. This is not due to participants not talking to each other. The fact that participants generally actively used the verbal communication channel is evident in longer response times in the Communication condition.

Unexpectedly, in contrast to Experiment 1 accuracy in the J condition was not higher than in the IS condition. This can be explained by learning taking place through the whole course of the experiment. Further support for this explanation comes from lower errors in the IS condition compared to the ID condition which were identical in terms of the available information. This suggests that individuals' localization precision in the joint phases of Experiment 1 was overestimated, and participants might have been closer to optimal performance in the Joint condition than it appeared from the data pattern. In other words, participants might have been doing their best given their abilities in the Joint conditions of Experiment 1, while our models provided benchmarks that were biased towards higher precision.

From the comparison of the present data with the model predictions of different integration schemes we can conclude that joint judgments were less precise than predicted by the statistically optimal model. Because we did not track how joint judgments were formed from individually perceived estimates of the correct location, a limitation of the present analysis is that we cannot tell for certain whether both participants averaged their private estimations on each trial or whether the more accurate member dominated in the final judgment on most trials (Hastie & Kameda, 2005; Laughlin & Ellis, 1986): both strategies could result in the empirically observed pattern of accuracy and variability of judgment errors.

It is worth noting that in contrast to previous findings on perceptual decision-making (Bahrami et al., 2010), dyads were not more accurate than the better member from the dyad, regardless of whether they communicated or not. One aspect to keep in mind is that the efficiency of the Simple Averaging strategy is reduced if the individual contributors to a joint judgment are systematically biased in the same direction. With larger biases, the Take-the-Best scheme is more likely to outperform the Simple Averaging scheme (Soll & Larrick, 2009). Given that participants exhibited a strong positive bias in the direction of their line of sight in the present task, it is likely that at least for some dyads minimizing variance did not lead to an equal minimization of absolute error.

Finally, when individual perceptions are redundant the collective benefit crucially depends on the differences between individual abilities, as predicted by Bahrami et al.'s (2010) theory. When two individuals in a dyad shared the same visual information, the accuracy ratio reflecting the discrepancy in two dyad members' localization abilities predicted the collective benefit accrued by the dyad. In line with our predictions we found that in Experiment 1, where the two percepts were complementary to each other, there was a much weaker dependency between the quality of joint judgments and inter-individual differences in absolute task-related abilities. The evidence for this comes from the non-significant correlation between the accuracy ratio reflecting the discrepancy in two dyad members' localization abilities and the collective benefit accrued by the dyad.

3.4 Experiment 3

Experiment 2 showed that communication had no effect on efficiency of information integration when partners within a dyad shared an identical perceptual input and when direct feedback was available. Previous research on perceptual decision-making showed that in the absence of feedback communicating individual confidence is sufficient to achieve collective benefit (Bahrami

et al., 2010, 2012b). The third experiment tested whether verbal communication can lead to a collective benefit in the absence of feedback when continuous judgments rather than categorical decision-making is required. As in Experiment 2 we provided partners with the same perceptual input, but this time they did not receive feedback at the end of the trial during the joint and individual phases of the experiment.

3.4.1 Method

Participants

Thirty-two English-speaking students (16 males) aged between 19 and 25 years (M = 22.6) were tested in pairs. The two participants in a pair were always of the same gender.

Materials and apparatus

The same stimuli and set-up as in Experiment 2 were used.

Procedure

The same procedure as in Experiment 2 was applied. There was no feedback at the end of the trials.

3.4.2. Results

Participants reached a consensual joint judgment in 96.88% of trials in the Non-communication condition and in 100% of trials in the Communication condition. After exclusion of joint trials where a consensual judgment was not reached, trials where the response was not made within the time limit, and outlying data points, 97.8% of the trials were retained for ensuing analyses.

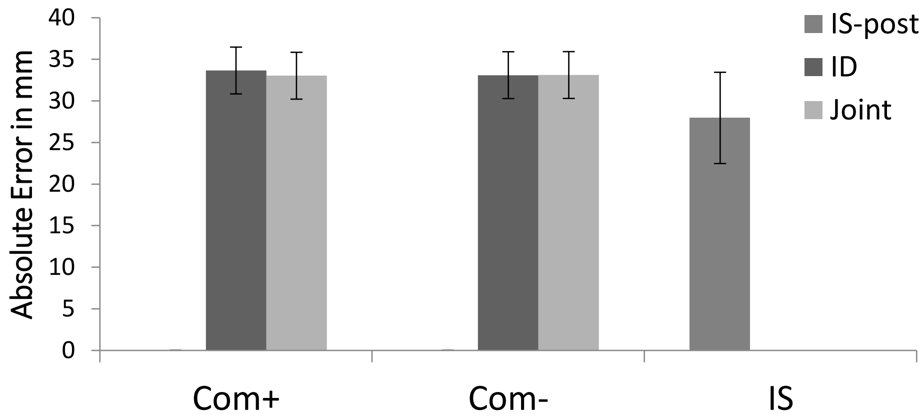


Figure 23. Results of Experiment 3 (Study II): absolute error. "Com+" shows the results for the Communication condition and "Com-" shows the results for the No-communication condition. Error bars for Com+ and Com- stand for within-subjects 95% confidence intervals of the respective 2 × 2 ANOVA. Error bar in the IS condition displays 95% confidence interval.

First, absolute error of the location judgments was analyzed. Performance in the ID and J conditions was compared by means of a two-way repeated-measurement analysis of variance (ANOVA) with the factors Condition (ID vs J) × Communication (Communication vs No-Communication). The test showed no significant main effects and no interaction between the two factors (*p*s > .7). A paired-sample *t*-test showed no significant difference between ID and J condition. Absolute error of judgments in the IS condition (M = 28.0 mm, SD = 16.9 mm) was lower than in the J condition (M = 33.2 mm, SD = 12.0 mm), $t(15) = -3.19$, $p = .006$; and lower than in the ID condition (M = 33.4 mm, SD = 15.9 mm), $t(31) = -2.89$, $p = .007$.

Variability of judgment errors was analyzed by means of a two-way repeated-measurement analysis of variance (ANOVA) with the factors Condition (ID vs J) × Communication (Communication vs No-Communication). This analysis showed no significant main effects and no interaction between the two factors (*p*s > .062). There was no significant difference in variability of judgment errors between the ID and the IS condition, and between the J and the IS condition.

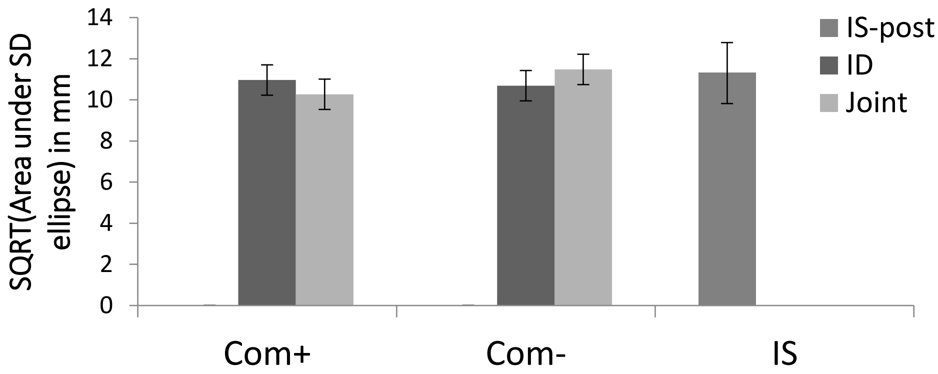


Figure 24. Results of Experiment 3 (Study II): variability of judgment errors. "Com+" shows the results for the Communication condition and "Com-" shows the results for the No-communication condition. Error bars in Com+ and Com- stand for within-subjects 95% confidence intervals of the respective 2 × 2 ANOVA. Error bar in the IS condition is the 95% confidence interval.

Next, we looked at measures not related directly to performance. Participants required more proposals of target location in the Non-Communication condition (M = 0.86, SD = 0.41) than in the Communication condition (M = 0.68, SD = 0.25), t(31) = 2.73, p = .01.

The average time required to complete a judgment was analyzed by means of a two-way repeated-measurement analysis of variance (ANOVA) with the factors Condition (ID vs J) × Communication (Communication vs No-Communication). The analysis revealed a significant interaction between the two factors, F(15) = 16.4, p = .001, partial η^2 = .522. , and a significant main effect of Condition, F(15) = 32.7, p < .001, partial η^2 = .686. Dyads took more time to make judgments (M = 37.0 s, SD = 12.6 s) than individuals did in the ID condition (M = 24.7 s, SD = 8.0 s). The main effect of Communication was not significant. RTs in the Communicative J condition (M = 40.6 s, SD = 14.2 s) were longer than in the Non-communicative J condition, t(15) = 2.62, p = .019.

Model comparisons

Variability properties of individual judgments collected in the ID condition were used as inputs for the models to generate predictions on variability of judgment errors in the J condition. Pair-wise comparisons showed that variability of judgment errors in the J condition was significantly larger than predicted by all of the tested models (*p*s ≤ .001).

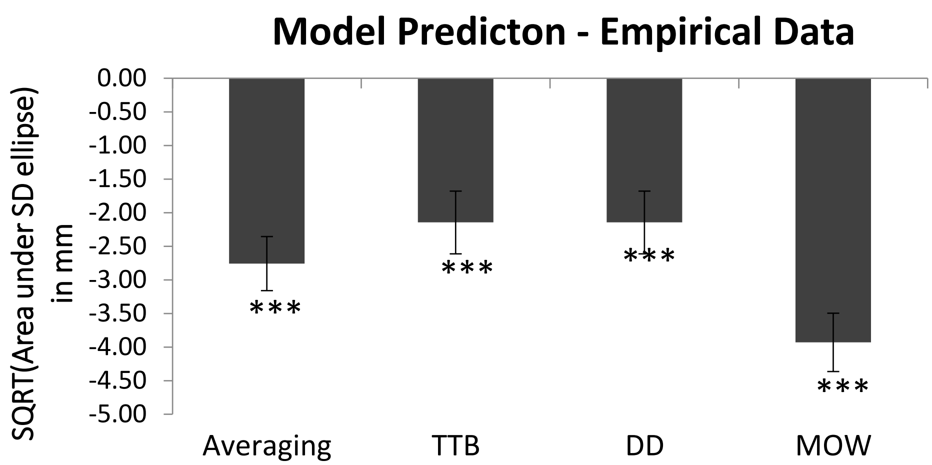


Figure 25. Results of Experiment 3 (Study II): comparisons with model predictions. Differences between the predictions on the area of the SD ellipse from different integration schemes and empirical data obtained from dyads in the Joint mode of the task in Experiment 3. Asterisks indicate significance of the corresponding paired-sample *t*-test. Error bars stand for the SE of mean difference. Averaging – *Simple Averaging*, TTB – *Take-the-Best*, DD – *Dimension Distribution*, MOW – *Multidimensional Optimal Weighing*.

3.4.3 Discussion

The results demonstrate that participants failed to integrate information available to them when deprived from immediate feedback even when they could communicate with each other. Verbal communication did not help, although participants tried to use communication to integrate their perceptions and to agree on joint judgments. Attempts to communicate are reflected in their longer response times when communication was allowed and the larger number of proposals made when communication was disallowed. Contrary to the findings of previous research on joint

decision making (Bahrami et al., 2010), communication could not compensate for the lack of feedback when forming continuous joint judgments. This suggests that in contrast to a situation of dichotomous decision-making, availability of feedback on accuracy is necessary for inter-personal integration of information, at least in the context of perceptual judgments.

The lack of a collective benefit is particularly surprising because even simple averaging of two internal estimates has a potential to at least reduce the variability of joint judgments. However, in the present experiment participants could not achieve such a reduction without being given feedback on accuracy of individual and joint judgments.

The only improvement observed in Experiment 3 was that participants were more accurate when providing individual judgments at the end of the experiment. The fact that the observed decrease in absolute error was not accompanied by a reduction in variability of errors suggests that the improvement occurred because participants gradually calibrated their sensory system to looking from a slant angle which, in turn, led to a decrease in their overshooting bias.

3.5 Experiment 4

Experiment 3 demonstrated that availability of direct feedback is a necessary condition for individuals to successfully integrate individual estimates into a joint judgment when receiving the same perceptual information. In the fourth experiment we tested whether feedback on accuracy of individual and joint judgments is also a necessary condition for successful integration of complementary visual information. It is possible that dyad members would completely rely on the partner in locating the target on their depth dimension because they have very high uncertainty on this dimension. This would amount to applying the Dimension Distribution scheme, which is very efficient under the condition of complementarity between the two partner's precision dimensions. In this case, each dyad member's imprecision on the depth dimension would be mutually compensated

by the reliable component of the partner's percept, leading to close to optimal performance. It is yet an open question whether this coordination would require verbal communication, or whether distribution of dimensions could emerge from simple interactions in the environment.

To address these questions, we provided participants with orthogonal viewpoints on the virtual task layout and deprived them from feedback.

3.5.1 Method

Participants

Thirty-two English-speaking students (16 females) aged between 20 and 30 years (M = 22.6) were tested in pairs. The two participants in a pair were always of the same gender.

Materials and apparatus

The same stimuli and set-up as in Experiment 1 were used.

Procedure

The same procedure as in Experiment 1 was applied with the exception that no feedback about the precision of participants' judgments was provided.

3.5.2 Results

Participants came to a consensual joint judgment in 98.75% of trials in the condition with communication, and in 100% of trials without communication. After exclusion of joint trials where a consensual judgment was not reached, trials where the response was not made within the time limit, and outlying data points, 97.75% of the trials were retained.

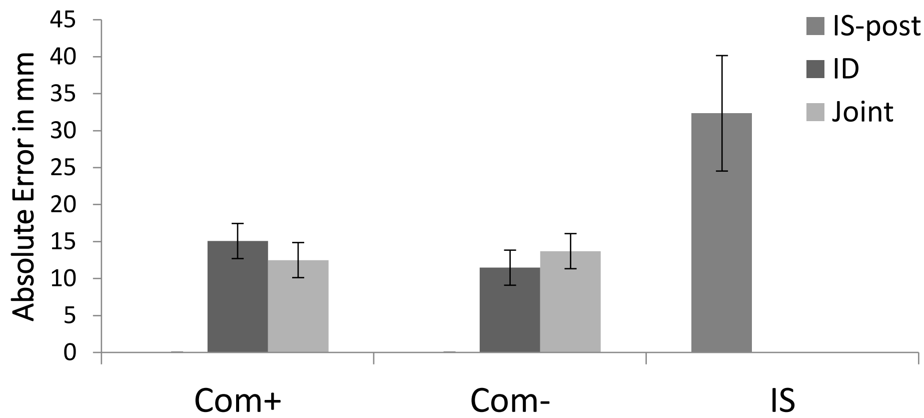


Figure 26. Results of Experiment 4 (Study II): absolute error. "Com+" shows the results for the Communication condition and "Com-" shows the results for the No-communication condition. Error bars for Com+ and Com- stand for within-subjects 95% confidence intervals of the respective 2 × 2 ANOVA. Error bar in the IS condition displays 95% confidence interval.

First, absolute error of location judgments was analyzed. A two-way repeated-measurement analysis of variance (ANOVA) with the factors Condition (ID vs J) × Communication (Communication vs No-Communication) revealed no significant main effects (*p*s > .245), but a significant interaction between the two factors: $F(15) = 4.73$, $p = .046$, partial $\eta^2 = .24$. When communication was allowed absolute error of judgments in the J condition was lower than in the ID condition, while there was an opposite relation between the conditions when communication was disallowed (see Figure 26). Absolute error of judgments in the IS condition (M = 33.3 mm, SD = 19.9 mm) was significantly larger than in the J condition (M = 13.1 mm, SD = 10.9 mm), $t(15) = 5.19$, $p < .001$; and larger than in the ID condition (M = 13.3 mm, SD = 14.1 mm), $t(31) = 5.4$, $p < .001$.

The absolute error of dyad judgments in the J condition was further compared with absolute error of judgments of the more accurate dyad member. Dyads achieved lower absolute error of judgments in the J condition than the more accurate dyad members in the IS condition (M = 23.0 mm, SD = 13.1 mm) both with communication (M = 12.5 mm, SD = 12.0 mm), $t(15) = -3.48$, $p = .003$, and without communication (M = 13.7 mm, SD = 10.6 mm), $t(15) = -4.28$, $p = .001$.

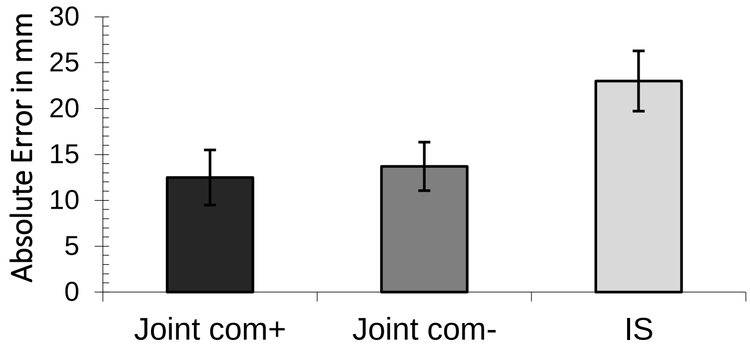


Figure 27. Results of Experiment 4 (Study II): comparison of performance between the better individual from a dyad and the joint performance from that dyad. Error bars are SE of the means.

Next, the variability of judgment errors was analyzed by means of a two-way repeated-measurement analysis of variance (ANOVA) with the factors Condition (ID vs J) × Communication (Communication vs No-Communication). This test revealed no significant main effects but a significant interaction between the two factors, $F(15) = 6.38$, $p = .023$, partial $\eta^2 = .298$. When communication was allowed, judgments provided by dyads in the J condition ($M = 8.77$ mm, $SD = 6.68$ mm) were significantly less variable than judgments provided by individuals in the ID condition ($M = 12.7$ mm, $SD = 9.65$ mm), $t(15) = -2.76$, $p = .015$, but not when communication was disallowed. Judgments were less variable in the Communicative J condition than in the Non-Communicative J condition ($M = 10.4$ mm, $SD = 7.19$ mm), $t(15) = -2.38$, $p = .031$. Variability of judgment errors in the IS condition ($M = 14.8$ mm, $SD = 7.71$ mm) was significantly larger than in the J condition ($M = 9.59$ mm, $SD = 6.81$ mm), $t(15) = 7.18$, $p < .001$; and significantly larger than in the ID condition ($M = 11.8$ mm, $SD = 11.7$ mm), $t(31) = 2.08$, $p = .046$.

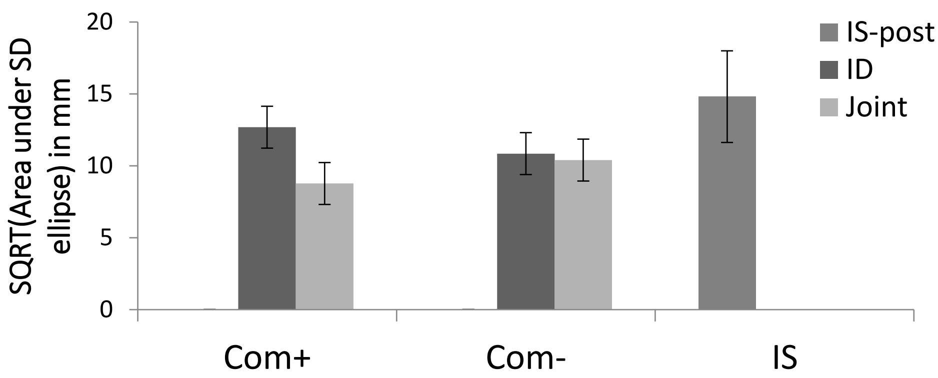


Figure 28. Results of Experiment 4 (Study II): variability of judgment errors. "Com+" shows the results for the Communication condition and "Com-" shows the results for the No-communication condition. Error bars in Com+ and Com- stand for within-subjects 95% confidence intervals of the respective 2 × 2 ANOVA. Error bar in the IS condition is the 95% confidence interval.

Next, we looked at dependent measures not directly related to performance. We found that there was no difference between Communicative and No-Communicative J conditions in the number of proposals made per trial (p = .389).

A repeated-measurement ANOVA with the factors Condition (ID vs J) and Communication (Communication vs No-Communication) on RTs revealed a significant main effect of condition, $F(15) = 4.56$, $p = .05$, partial $\eta^2 = .233$. RTs were significantly shorter in the J condition (M = 30.9 s, SD = 12.4 s) than in the ID condition (M = 36.7 s, SD = 12.2 s). An interaction between the two factors fell short of significance by a margin, $F(15) = 4.18$, $p = .059$, partial $\eta^2 = .218$, and the main effect of Communication was not significant. RTs in the J condition with communication (M = 27.7 s, SD = 11.2 s) were significantly shorter than in the J condition without communication (M = 34.1, SD = 15.6), $t(15) = -2.29$, $p = .037$.

Model Comparisons

We used variability parameters of individual judgments collected in the IS condition to generate predictions on variability of errors in dyad judgments.

Because we found a significant difference in variability of judgment errors between Communicative and Non-Communicative J conditions, we compared empirical data from the two conditions with model predictions separately. Pair-wise *t*-tests showed that variability of errors in joint judgments in the Communication condition was significantly smaller than predicted by the Simple Averaging model, $t(15) = -6.38, p < .001$; and smaller than predicted by the Take-the-Best model, $t(15) = -3.56, p = .022$. At the same time, it was larger than predicted by the Multidimensional Optimal Weighing model, $t(15) = 2.84, p = .012$, and than predicted by the Dimension Distribution model, $t(15) = 2.51, p = .024$. Variability of errors in joint judgments in the Non-Communication condition was also significantly larger than predicted by the Multidimensional Optimal Weighing model, $t(15) = 3.99, p = .001$, and than predicted by the Dimension Distribution model, $t(15) = 3.74, p = .002$; and it was lower than predicted by the Simple Averaging model, $t(15) = -6.58, p < .001$. However, it was not significantly different from predictions from the Take-the-Best model, $p = .319$.

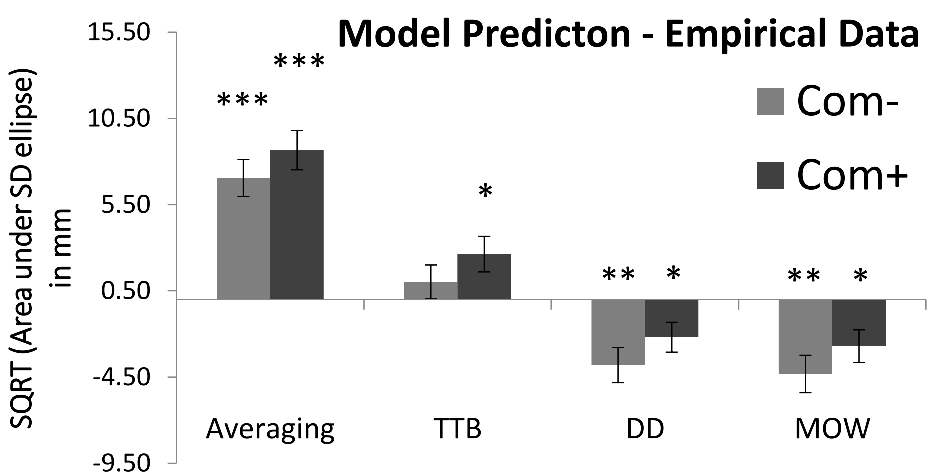


Figure 29. Results of Experiment 4 (Study II): comparisons with model predictions. Differences between the predictions on the area of the SD ellipse from different integration schemes and empirical data obtained from dyads in the Joint mode of the task in Experiment 4. "Com+" stands for Communication, and "Com-" stands for No-communication Joint condition. Asterisks indicate significance of the corresponding paired-sample *t*-test. The area data were transformed prior to the test. Error bars stand for the SE of mean difference. Averaging – *Simple Averaging*, TTB – *Take-the-Best*, DD – *Dimension Distribution*, MOW – *Multidimensional optimal weighing*.

3.5.3 Discussion

The results of the fourth experiment demonstrate that individuals can successfully integrate complementary information available to them into a joint judgment and reach a collective benefit even when feedback on accuracy is absent and when they cannot communicate. We conclude that neither feedback nor verbal, or more generally, explicit communication is absolutely necessary to obtain a collective benefit from inter-personal information integration. Availability of complementary information is sufficient to obtain such benefit.

Although verbal communication was not a necessary precondition to reach collective benefit, it had a facilitating role for inter-individual information integration when feedback was lacking. The comparison of the empirically observed variability of errors in joint judgments with the predictions derived from different integration schemes demonstrated that verbal communication allowed dyads to reach higher precision in their judgments than would be expected if the judgment of the generally more precise dyad member was consistently selected. When dyads could not communicate, the precision of their judgments was not different from precision of the more precise dyad member. This also implies that if reduction of random error in individual estimates was the only effect of integrating individual percepts, dyads would not have less absolute error in their judgments, than the more accurate individual from the dyad. However, this was not the case: even without communication dyads produced less absolute error than the more accurate dyad member. This suggests that not only random error but also systematic error (bias) was reduced as a result of information integration.

In contrast to the previous Experiments, individuals took less time to form joint judgments when verbal communication was permitted which provides an indication that communication may have sped up integration in the present experiment where complimentary information was provided

to two individuals and feedback about joint and individual accuracy was not available. This possibly reflects the fact that verbal communication helped individuals to construct structural models of their uncertainty about the location that would otherwise be constructed on the basis of feedback on individual accuracy.

3.6 General Discussion of Study II

In a series of four experiments we investigated how pairs of individuals would integrate two-dimensional spatial location judgments via verbal and environment-mediated interactions by sharing control over the joint response. Across experiments we manipulated the amount of structural overlap between individual perceptions by providing either orthogonal views or the same view on the virtual layout to two individuals forming a dyad. We also manipulated the availability of feedback about the accuracy of judgments. Within each experiment we varied the availability of verbal communication. This enabled us to test a variety of hypotheses about interpersonal integration of multidimensional perceptual information into joint judgments.

The first hypothesis regarded the relationship between the amount of collective benefit that can be achieved through information integration and the extent of complementarity in the information available to each individual. As predicted by the theorem of Maximum Collective Benefit, individuals in pairs that received complementary information from different viewpoints (Experiment 1 and 4) were more successful and benefitted from the integration more than individuals in pairs that received redundant information through the same viewpoint (Experiment 2 and 3). In line with earlier findings by Bahrami et al. (2010) when individuals shared the same perceptual input (Experiment 2), there was a significant relationship between the amount of collective benefit and inter-individual differences in the ability to precisely locate a target in space: dyads that were composed of individuals who were similarly accurate did better than dyads that were

composed of individuals who greatly differed in their accuracy. However, the significance of inter-individual differences in precision completely vanished when two individuals had complementary information (Experiment 1) so that the dimension of high uncertainty for one individual was the dimension of low uncertainty for another individual, and vice versa. In this situation, the availability of complementary information dominated any differences in individual precision.

We further established that access to complementary spatial information from different perspectives is sufficient to enable the "two heads better than one" effect, as dyads systematically outperformed their better members when complementary information was available to them (Experiment 1 and 4), even in the absence of verbal communication and feedback about the quality of individual and joint judgments. This finding indicates that individuals in a dyad could form a structurally accurate representation of their multi-dimensional uncertainty without external feedback, and that acting in a shared environment was sufficient to coordinate individual contributions to the joint judgment according to the respective information available to each individual.

The second and third hypotheses regarded the role of communication and feedback in the process of inter-individual information integration. In contrast to previous findings on collective perceptual decision-making, verbal communication did not play a decisive role in this process in the present experiments. Depriving participants of the opportunity to verbally communicate had little impact on the quality of their integrated judgments. In the two experiments that involved perception of complementary information (Experiment 1 and 4), dyads systematically outperformed the better individuals with and without verbal communication. In Experiment 2 where dyads looked from the same viewpoint and thus perceived the same information, the lack of opportunities to verbally communicate did not reduce the benefit from inter-individual integration when immediate trial-by-trial feedback was available. In Experiment 3 where immediate feedback about the accuracy of judgments was not available verbal communication could not compensate for the lack of feedback.

It is quite remarkable that in the absence of feedback dyads failed to integrate their individual perceptions resulting in an accuracy of joint judgments that was no better than that of the average dyad member. This interplay between communication and feedback is markedly different from what has been suggested by previous studies of joint decision-making based on perceptual information (Bahrami et al., 2010, 2012b). Bahrami et al. (2010, 2012b) found that only when individuals in a dyad could verbally communicate dyads could outperform their better members. Without verbal communication, the level of joint performance was at the level of the more competent dyad member (Bahrami et al., 2010, 2012b). Bahrami et al. (2010, 2012b) also report that dyads integrated information optimally even in the absence of trial-by-trial feedback, when allowed to verbally communicate, suggesting that feedback is neither necessary nor sufficient for achieving collective benefit from interpersonal information integration, but that communication is necessary and sufficient to achieve such a benefit. Our findings suggest just the opposite: communication was neither necessary nor sufficient to achieve collective benefit from inter-individual integration of perceptual information, while the availability of feedback was a necessary condition to improve the quality of joint judgments compared to an average individual when individual percepts were redundant (Experiment 2).

At the same time, verbal communication did make a small but significant contribution to the process of inter-individual information integration when participants integrated complementary perceptions without receiving feedback about the accuracy of judgments. In this situation the opportunity to verbally communicate enabled two individuals to integrate information from two viewpoints better than the average individual could do when having consecutive access to both viewpoints. Furthermore, we found that only when allowed to verbally communicate did participants' joint judgments become more reliable than judgments of the better individual in a dyad.

This suggests that verbal communication can replace to some extent information that would be otherwise extracted from feedback on accuracy of individual and joint judgments.

Finally, our results suggest that participants may have relied on different integration schemes depending on whether they perceived redundant (same) or complementary information when providing joint judgments. When perceiving the same information (Experiments 2 and 3) the level of precision of joint judgments was at the level predicted by simple heuristics that operate without taking into account geometrical properties of individual uncertainty and do not require fine-tuned coordination of individual contributions, such as Simple Averaging or selecting the judgment from the more accurate individual in a dyad. When perceiving complementary information (Experiments 1 and 4), participants applied integration schemes that allowed them to come closer to statistically optimal joint judgments based on their individual precisions.

One way to explain why different integration schemes were used in the same task is provided by the flat-maximum account (von Winterfeldts & Edwards, 1986). On this account, people will prefer to follow simple heuristics when this is not expected to be very detrimental compared to more fine-tuned but more demanding strategies. We found this behavior in Experiment 2, where variability of error in joint judgments was higher than predicted by the optimal integration scheme, but consistent with the application of simpler integration schemes. In this experiment, the expected difference in precision between applying simpler and more sophisticated integration schemes was small. In Experiments 1 and 4, the increase in precision that could be achieved by applying optimal or close to optimal integration schemes was much larger. This is further illustrated in Figure 30, which displays the expected absolute error in joint judgments (y-axis) resulting from different integration schemes as a function of structural overlap between individual uncertainties. Thus, participants seem to have chosen more fine-tuned integration schemes that demanded evaluation of

structural properties of individual uncertainty only if applying them promised substantial benefit in joint precision.

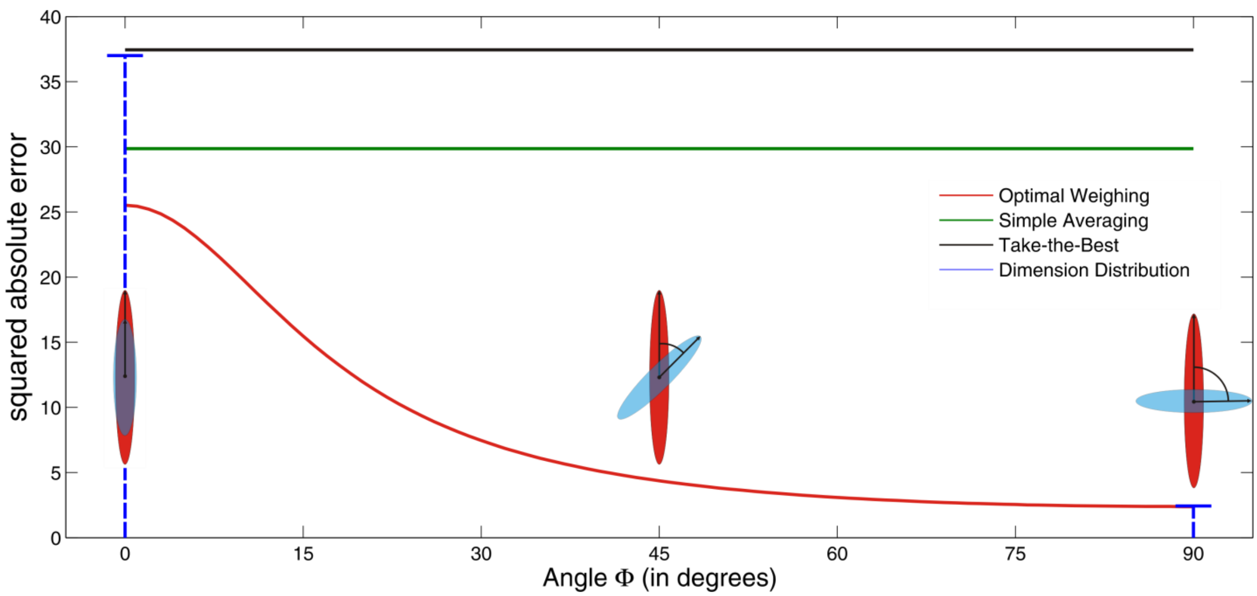


Figure 30. Predictions on judgment accuracy derived from different integration schemes (an illustrative example). The ordinate axis plots expected absolute error of joint judgments derived through different integration schemes under the assumption that individual judgments are unbiased. The abscissa axis plots the angle ϕ between the major axes (black arrows on ellipses) of confidence ellipses constructed for A's and B's judgments and captures the structural overlap in the two judges' uncertainty about the true value of a position in two-dimensional space. ϕ = 0° implies redundant information distribution (ellipses on the far left), ϕ =90° implies complementary information distribution (ellipses on the far right). Distinct curves show predictions from different integration schemes. The Dimension Distribution model is plotted for two angles (0° and 90°) where its predictions can be derived analytically. The exact vertical order of the lines representing predictions from Simple Averaging and Take-The-Best model depends on variances of two judges' errors on both dimensions. Individual precision of the two judges is parameterized with standard deviation of random error in their judgments: $\sigma_{red,x}$ = 1 (minor axis of the red ellipse); $\sigma_{red,y}$ = 9 (major axis of the red ellipse); $\sigma_{blue,x}$ = 1.2 (minor axis of the blue ellipse); $\sigma_{red,y}$ = 6 (major axis of the blue ellipse). These ratios are representative of empirical data obtained with the experimental task we used in the current study.

Failure to reach optimal level of performance

There could be several causes why participants in all experiments systematically underperformed relative to a statistically optimal benchmark. Whereas the current experiments demonstrated a "two heads are better than one" effect under the condition of access to

complementary information, we could not replicate an important aspect of Bahrami et al.'s (2010) finding in the domain of perceptual decision-making: in the present experiments dyads were no better than their more competent member when they were provided with redundant visual information. As has already been mentioned, one potential cause for this discrepancy is that participants tend to use sub-optimal but less demanding strategies that are expected to perform sufficiently well (von Winterfeldts & Edwards, 1986). This could explain why dyads did not resort to a Multidimensional Optimal Weighing scheme in Experiment 2, as optimal weighing would require a fine-tuned calibration of individual contributions while forming joint judgments.

In Experiment 1 and 4, where complementary information was accessible to the two individuals in a dyad, a near-optimal but cognitively less demanding strategy of Dimension Distribution predicted a level of precisions that is almost indistinguishable from optimal weighing. Most participants failed to systematically follow this strategy, even though the data clearly demonstrate (particularly in Experiment 1) that they coordinated individual contributions to the joint judgment to mutually compensate for uncertainty on the depth dimensions. One possible explanation for why participants assigned non-optimal weights to the individual contributions when forming joint judgments is that individuals in a dyad gave too much weight to inaccurate components of their partner's percept. This could be due to a social obligation to let the partner have a say on her preferred position even on her unreliable depth dimension, or the "urge to contribute" (Bahrami et al., 2012a).

In addition to social influences and a reluctance to completely ignore even obviously erratic opinions (Asch, 1952), the failure to achieve optimal integration could be due to communicating sub-optimal measures of uncertainty, such as confidence (Bahrami et al, 2010). For instance, Bahrami et al. (2012a) found that using a non-verbal mode of communication helped people to ignore detrimental influences. However, we did not find this effect in the present study: when dyads

coordinated their joint judgment through shared control they performed no better than when verbally communicating with one another.

It is also possible that the precision with which dyad members could form their individual judgments during the joint phases of the experiments was overestimated because it was estimated based on their individual performance at the very end of the experiment, at least in Experiments 1 and 4. As Experiment 2 and Experiment 3 showed, there is considerable learning and perceptual calibration taking place throughout the course of the experiment; the quality of participants' judgments at the end of the experiment might not accurately represent their quality in the preceding phases. In other words, the dyads in Experiment 1 and 4 may have been much closer to optimal integration than our results suggest because we overestimated their individual accuracy in this phase of the experiment, assuming it to be the same as individual accuracy in later phases of the experiment.

Finally, there are some methodological differences between the approach taken by Bahrami et al. (2010, 2012a, 2012b) and the approach taken in the present study, suggesting that it could be problematic to directly compare the results of the two studies. In the experiments by Bahrami et al. (2012b) collective benefit was demonstrated on data aggregated over 80 trials involving joint decisions with communication (without communication it took dyads even longer to reach a collective benefit). In our experiments the overall number of trials was 40 if trials from the two communication conditions are summed up. It is possible that after a larger number of trials more dyads could have gained a collective benefit in Experiment 2 where they had access to the same visual information and received feedback about the accuracy of judgments. Our optimism towards this possibility is supported by the fact that almost half of the dyads in Experiment 2 reached higher level of precision (but not lower absolute error!) than that of the most precise dyad member at the solitary phase at the end of the experiment.

There are further important methodological differences that I will discuss in more detail in the following sub-section.

Response mode

The task we used in this study and the task used by Bahrami et al. (2010, 2012a, 2012b) differ in how participants provided their response. Participants in our task were responding by providing point estimates on a continuous scale, while the oddball detection task (Bahrami et al., 2010, 2012a, 2012b) requires making a choice between two categorical alternatives (the oddball could be present on either of two screens). While normative theories assume no distinction between the processes of judgment and the choice based on a judgment, there are reasons to suspect that the two response modes invoke different information integration processes.

Hinsz (1999) emphasized that the two processes have different outcome goals. When groups select from discrete alternatives, the computational task can be described as choice-making with the ultimate goal of achieving agreement that can be described as consensus. In the course of achieving agreement group members exert influence on each other when the initial individual choices are in conflict. Some group member(s) will need to change their choice, while other(s) will not change their choice. In contrast, when group members make a selection from a set of alternatives having an underlying continuum, the task can be described as judgment-making, and the ultimate goal of this process is agreement that can be described as compromise. When seeking compromise, all group members are expected to adjust their initial positions to arrive at a group decision. These considerations led the author to conclude that “there are qualitative differences in group decision processes associated with discrete and continuous response measures” (Hinsz et al., 1997, p. 51).

Einhorn and Hogarth (1981) made a similar argument. When continuous measures are used, a conflict between different individual judgments gets usually resolved through compensatory

integration schemes, such as linear combination of evidence, for example, Simple Averaging or Weighted Averaging. While the judgment made on the basis of the individually available information is supposed to aid with (and ideally determine) the process of choice, it is neither necessary nor sufficient for choice. In extreme cases, informed judgments can be ignored at the stage of choice. According to Einhorn and Hogarth (1981) a situation of choice imposes additional sources of conflict on the decision-maker, and the process of this conflict resolution follows more non-compensatory strategies such as dominance of only one information source. For example, making a choice invokes a stronger sense of commitment than a situation of judgment (Beach & Mitchell 1978; Janis & Mann, 1977). Billings and Scherer (1988) provided experimental evidence that the requirement to give responses in the format of choice fostered solitary decision-makers to use more non-compensatory strategies for information integration compared to a situation of quantitative judgment. The same tendencies could generalize to the group level in a situation of collective choice and judgment.

In the experiments conducted by Bahrami et al. (2010, 2012a, 2012b) the response consisted in a dichotomous choice and shutting down the verbal communication channel could have caused dyads to switch to non-compensatory strategies, for example, dominance by the more competent dyad member. In the present task responses were made on a multidimensional continuous scale, and participants could continue to integrate individual judgments through compensatory (linear combination) schemes even when the verbal communication channel was shut down. Although the method we applied does not allow for a direct inference on whether joint judgments were formed through compensatory or non-compensatory information integration schemes, we also did not find any evidence that these processes were different between purely environment-mediated interactions and those assisted with verbal communication. In Experiments 1 - 3 we found no difference in performance between the communicative and non-communicative modes for interaction, and

therefore have no reason to assume that joint judgments were formed through qualitatively different processes under the two modes. In Experiment 4 we found that joint judgments would have the same variability if judgments of the generally more precise dyad member was selected on each trial. However, this would not allow dyads to be more accurate than the generally more accurate dyad member, which was clearly the case. Thus, Experiment 4 also provides no evidence to assume qualitatively different integration processes under the two modes of interaction. While process-tracing techniques would be absolutely necessary to make a definite conclusion, the results of our four experiments suggest that when the task requires a response on a continuous scale, environment-mediated interactions can support the same, probably compensatory, processes of information integration as interactions coordinated with verbal communication.

A further difference to the method employed by Bahrami et al. (2010) is that we provided our participants with a possibility for arriving at a joint judgment when no verbal communication was allowed. Unexpectedly, we found evidence that participants were more intensively using the coordination strategy of bargaining through proposals and rejections as a substitute for verbal communication only in Experiment 1, where both complementary visual information and feedback were provided, and in Experiment 3 where there was no feedback and no perceptual complementarity. In both experiments we did not find evidence that verbal communication can add anything over and above non-verbal interactions. Whether and how well a mere bargaining could substitute verbal communication in the task employed by Bahrami et al. (2010, 2012a, 2012b) or a similar one is an interesting question for understanding the difference between processes taking place in a situation of group choice and group judgment.

Given that our task is likely to favor compensatory strategies for information integration, it was a surprising finding that dyads failed to demonstrate a robust collective benefit when their members had access to the same perceptual information (Experiments 2 and 3). As our theoretical

predictions suggest (see Figure 29 on p. 174) if dyads had been following the strategy of Simple Averaging, for most dyads (for 12 out of 16 dyads in both Experiment 2 and Experiment 3) the joint judgments would have had less variability of error than judgments of their more capable member.

In Experiment 3, where individuals had no opportunity to objectively assess and compare individual accuracies within a dyad because feedback was missing, Simple Averaging would have provided a powerful and simple heuristic to improve joint judgments to the level equal to or higher than the average precision of the two individuals in a dyad. Yet the variability of errors in joint judgments suggests that in Experiment 3, as well as in Experiment 2, dyads were generally following more non-compensatory strategies with possibly only one individual dominating the group judgment on a given trial.

We suspect that one potential cause of this collective failure is that participants were not prompted to provide individual judgments before forming a joint judgment. The lack of an opportunity to externalize internal estimates of the correct location may have caused people to fail to develop and voice their disagreeing opinions, an effect known as groupthink (Bahrami et al., 2012a; Turner & Pratkins, 1998). While we on purpose did not include probes of private location judgments into joint trials to avoid the primacy effect (Anderson, 1991) and individuals' strong commitment to their initial judgment (Sorkin, Hays, & West, 2001) this decision may have undermined the efficiency of collaboration in dyads. The drawback of this decision could be exactly the lack of participants' commitment to their perceived correct locations; and because compensatory integration strategies are supposedly more cognitively demanding (Einhorn, Kleinmuntz, & Kleinmuntz, 1979), it led to groupthink as a consequence. It is then possible that in contrast to a situation of categorical choice, in a situation of collective judgment prompting an individual judgment prior to a negotiation phase can be helpful, as it would facilitate usage of more compensatory and, hence, more beneficial strategies for forming the group judgment.

Further, the role of communicating confidence might be larger in decision-making tasks than in judgment tasks. In contrast to the studies by Bahrami et al. (2010, 2012a, 2012b), Sniezek and Henry (1989) did not find that exchange of confidence in individual judgments affected the quality of group judgment nor did they find indications that it had an important role in the process of joint judgment formation. In particular, Sniezek and Henry (1989) reported that groups did not use confidence information to determine the weights assigned to individual judgments to form joint judgments. Moreover, they found that there was no systematic relationship between individuals' confidence in their judgments and the actual accuracy of their judgments. They conclude that "confidence is not a valid cue to judgment accuracy and therefore should not be used as a basis for variable weighing" (p. 21). This might in part explain why verbal communication in our task was not as important as in the experiments by Bahrami et al. (2010, 2012a, 2012b) and provides a further indication that there are qualitative differences in the processes of information integration between joint judgments on a continuous scale and joint choices between nominal alternatives.

It would therefore seem worthwhile to further investigate the effect of the response mode on inter-individual information integration using the present task and method. This line of research could increase our understanding of the qualitative differences between the cognitive processes operating in the context of discrete choice and continuous judgment. It could also help us to understand why certain procedural factors, such as verbal communication or immediate feedback, become critical and can lead to collective failure in one context but not the other. Ultimately, understanding how such procedural factors affect the within- and inter-individual information flow compromising group information integration processes, is informative about how people select, evaluate, integrate, and convert into action sensory and social information (e.g., in Bahrami et al., 2012a; Pescetelli, Rees, & Bahrami, 2016), and what cognitive mechanisms support these processes. One possibility is that in a situation of collective decision or judgment human cognition relies on

mechanisms specialized not in solving information integration problems, but in solving a different type of problems (e.g. social dominance or social cohesion), leading to a collective failure. Whether this applies to a situation of collective judgment, collective choice, or both, requires further investigation and would comprise interesting questions for future research.

The role of meta-cognition

A further important question is what kind of information was exchanged between individuals in a dyad and how it was exchanged in the present study. The results make it very unlikely that participants were verbally communicating their confidence, as seems to have been the case in the perceptual decision task investigated by Bahrami et al. (2010, 2012a, 2012b). Confidence, if defined in terms of information on probability of one's internal estimate being correct, lacks information on the structure of individual uncertainty and it would not be particularly helpful where individuals needed to distribute their contributions to a joint judgment in a two-dimensional space (Experiment 1 and 4). At the same time, verbal communication could not substitute feedback on joint and individual accuracy (Experiment 3), as suggested in Bahrami et al. (2012b). Instead, we hypothesize that through verbal communication participants could exchange their theories (Beppu & Griffith, 2009) about where the target could be located which in this situation involved structured relations between the target location and individual perceptions.

Verbal communication is a powerful social tool that can facilitate alignment of joint attention, perspective-taking, and the construction of higher-order situational models of the joint task (Tylén et al., 2010). As a means of inter-individual coordination, it is particularly efficient in situations requiring coordination of complementary abilities, where individuals need to take different roles and do different things to achieve a joint goal (Sebanz, Bekkering, & Knoblich, 2006; Tylén et al., 2010). Verbal communication can also facilitate or explicitly induce perspective-taking (Beveridge

& Pickering, 2013; Dietz, Roepstorff, & Wallentin, 2010), and thus provide a mental model of the partner's current perception and the uncertainty associated with it. It is plausible that in Experiment 4, where the individuals within a dyad had complementary perceptions but were not provided with feedback, participants had a coarse representation of anisotropy of their own and their partner's uncertainty even without receiving feedback about the accuracy of judgments. This representation was sufficient to achieve a collective benefit, even when verbal communication was not possible. However, when verbal communication was available it assisted individuals in building cognitive models with a more accurate structure and thereby allowed dyads to construct a joint situational model comprising complementary spatial dimensions in a shared coordinate system.

It is further possible that individuals' implicit confidence, defined as automatic and non-conscious assessment of individual uncertainty (Bach & Dolan, 2012), may have been conveyed via the actions controlling the cursor movement that reflected each participant's attempt to locate the target. There is evidence that human observers can rely on low-level kinematic characteristics of a third person's actions to infer the latter's confidence (Patel, Fleming, & Kilner, 2012). Action kinematics also been used as an implicit measure of individual's confidence in their own decision (Dale, Roche, Snyder, & McCall, 2008; Freeman, Dale, & Farmer, 2011; McKinstry, Dale, & Spivey, 2008; van der Wel, Sebanz, & Knoblich, 2014). Although participants did not achieve collective benefit under the condition of redundant information distribution (Experiments 2 and 3), coordinating their response through shared control was not less efficient than when mediated with assistance of verbal communication. This suggests that dyads could possibly coordinate the joint judgment by gradually adjusting individual contributions and paying attention to the degree of each other's behavioral ductility in this process.

Because we did not replicate the original collective benefit result from Bahrami et al. (2010), it is then still an open question whether coordination without explicit communication, such as via a

shared environment as in the reported experiments, can effectively replace verbal communication in a situation where communicating explicit confidence allows groups of individuals attending to the same perceptual input to outperform their more capable member, as in Bahrami et al. (2010). This is an intriguing possibility, especially given that implicit representation of one's own accuracy and precision can be more accurate in low-level perceptual judgments, and possibly decisions, than meta-cognitive representations (Maloney, Trommershäuser, & Landy, 2007). It is then possible that individual contributions to decisions reached through environment-mediated interactions, or through joint action, would be more proportionate to implicit and, hence, more accurate measures of individual uncertainty. One prediction that follows from this hypothesis is that simple joint decisions formed through coordination without explicit communication would be more accurate than those reached through verbal communication.

Summary and concluding remarks

In a series of four experiments we investigated how dyads would integrate individual perceptions of a target location in two-dimensional space to form joint judgments. We found that when dyads had complementary perceptions from orthogonal viewpoints so that the dimension of one partner's high uncertainty was the dimension of the other partner's low uncertainty, and vice versa, dyads systematically outperformed their better members even in the absence of verbal communication, of feedback, or both. Contrary to previous findings which suggested a crucial role of meta-cognition supporting the process through verbal communication of confidence (Bahrami et al., 2010, 2012a, 2012b) and only a minor role of feedback about individual accuracy, in the present experiments environmental feedback was both necessary and sufficient for dyads to outperform the average individual in a situation where they had access to the same perceptual information. Verbal communication was neither necessary nor sufficient to achieve a collective benefit, although it aided

integration of complementary perceptions in the absence of feedback. We suggest that these differences between the present results and those reported in Bahrami et al. (2010, 2012a, 2012b) are due to differences in response mode and due to the opportunity for environment-mediated interaction in a shared virtual environment that existed in our experiments.

Because we neither replicated the original collective benefit result reported by Bahrami et al. (2010), nor found a difference between environment-mediated interactions and verbal communication, our findings raise the possibility that in a task where collective benefit can be robustly achieved, such as in Bahrami et al. (2010, 2012a, 2012b), environment-mediated interactions can effectively substitute verbal communication in the process of interpersonal information integration. An implication of this hypothesis would be that a shared environment can provide a sufficient and effective means of establishing the common reference system necessary for optimal information integration across minds (Ernst, 2010). Because environment-mediated interaction is not bound to verbal symbolic systems, this mode of interaction would have the potential of removing linguistic and cultural barriers in realizing the "wisdom of the crowd" (Surowiecki, 2004).

Finally, answering the question when two heads are better than one, we emphasized that a robust and pronounced collective benefit is achieved when individuals have different views that provide them with complementary information. If this is the case, group judgments seem to be almost completely tolerant to individual weaknesses and make the best of individual abilities. This can serve as a reminder that the true advantage of group interactions requires revealing hidden diversity (Page, 2008) and that one way to reach it is to combine multiple and different perspectives on the problem.

Conclusions

Two empirical studies were conducted to investigate how environment-mediated interactions can support the process of inter-individual information integration under different conditions of structural overlap in individually available information.

With the three experiments in the first study we investigated how indirect interactions can support the process of judgment revision. In these experiments individuals provided independent location judgments in a shared environment, and could revise their judgments after observing another's judgments. Our main finding was that whether people can properly integrate observable information in the environment produced by another individual crucially depends on their uncertainty about their own judgment. When this uncertainty is low, they discard information of poor quality but integrate high-quality information, resulting in an improved quality of their judgments. When people's uncertainty about their own judgment is high, they do not discriminate between information of high and low quality in another's judgment. Under high uncertainty, people rely on observable social information as much as on their own perception even in situations where the information provided by others should be weighed much higher than their own judgment.

Surprisingly, we also found that individuals cannot integrate their own judgments containing complementary information in a sequential manner any better than judgments made by others. This points to the possibility that the known inability of people to rely on external advice more than on their own opinion, known as an egocentric discounting effect (Yaniv & Kleinberger, 2000) may not be purely social in nature. Rather, it might reflect a general dominance of immediately available evidence over imaginary evidence (or evidence which needs to be mentally simulated), in the process of information integration.

In the four experiments conducted in the second study, we investigated how environment-mediated interactions can support the process of joint judgment formation. In these experiments

individuals in ad hoc dyads were providing joint judgments in a shared environment. The main questions we asked concerned the role of feedback and communication in the process of inter-individual information integration, and their interplay under conditions of redundant and complementary information distribution. We found that this interplay crucially depended on how visual information was distributed across dyad members.

When two dyad members were provided with the same visual input, whether individuals could successfully integrate their percepts depended on whether they were provided with feedback on their accuracy. Successful information integration resulting in higher accuracy than that of the average dyad member was demonstrated only when feedback was available. At the same time, there was no evidence that verbal communication added anything to the process above and beyond environment-mediated interaction. Removing the verbal communication channel did not hurt the process of information integration when feedback was available, and it could not substitute feedback when feedback was not available.

When two dyad members were provided with complementary visual input, dyads could successfully integrate complementary information and systematically provided more accurate judgments than the better individual from a dyad. This collective benefit effect was demonstrated in the absence of communication, in the absence of feedback, and in the absence of both. Whereas neither availability of feedback about the individual and joint judgment accuracy nor verbal communication was necessary for successful integration of complementary visual information, the opportunity to verbally communicate facilitated the process in the absence of feedback, allowing dyads to make even more accurate joint judgments. Moreover, in the absence of feedback, being able to verbally communicate allowed two individuals to integrate visual information from two orthogonal viewpoints more efficiently than when they were individually switching between the two orthogonal viewpoints. Our interpretation of this finding is that verbal communication allowed

dyads to construct more accurate shared mental models of the task which allowed them to more properly distribute individual contributions to the final judgment.

The interplay between verbal communication and immediate feedback observed in our study is different to what has been reported in previous research on perceptual decision-making (Bahrami et al., 2010, 2012), where a significant impact of verbal communication on group performance and on the amount of collective benefit was found, even when dyad members had access to redundant visual information. One potential explanation for this discrepancy may be the difference in response mode implemented in our study and experiments by Bahrami et al. (2010, 2012a, 2012b): the experimental method developed by Bahrami et al. (2010, 2012a, 2012b) required a choice between two dichotomous alternatives, whereas our method required judgments on a continuous scale. We speculate that the requirement of providing a joint response on a continuous scale favors more compensatory processes of inter-individual information integration, that can be efficiently realized under non-verbal modes of interaction. Together, results of our experiments and findings from studies reported in Bahrami et al. (2010, 2012a, 2012b) point to the possibility that there might be a qualitative difference between information integration processes taking place in a situation of continuous judgment and that of categorical choice.

We could not replicate the collective benefit effect reported in previous studies on collective decision-making (Bahrami et al., 2010, 2012a, 2012b), where dyads systematically outperformed their better members under conditions of redundant information distribution. Our best explanation for this discrepancy is that participants in our experiment were simply not provided with sufficient time to establish a stable information integration procedure, and that individual learning effects canceled out collective benefit effects during the joint phases of the experiment. In experiments by Bahrami et al. (2012b) a reliable collective benefit was demonstrated based on data aggregated over 80 trials

(the smallest reported bin) of joint decisions in one condition. This is almost twice as many joint trials as participants completed in our experiment in total.

A skeptical interpretation of our findings would be that only environment-mediated interactions are not sufficient to yield collective benefit when two observers have access to redundant information. However, since we also did not find an effect of verbal communication on group performance, there is an intriguing possibility that environment-mediated interaction can effectively substitute verbal communication in collective judgment and decision-making under the right conditions. In the light of this hypothesis it would be interesting to further investigate how environment-mediated interactions compare to verbal communication in the context of a task where a robust collective benefit can be achieved, as in Bahrami et al. (2010, 2012a, 2012b).

Our analysis of the psychological mechanisms governing the process of inter-individual information integration is limited, however, in part due to methodological shortcomings of the current study. Since we did not probe individual judgments in joint trials in Study II, our evidence on how dyads integrated individual percepts into a joint judgment is indirect and not conclusive. In particular, we do not know whether both dyad members contributed on both spatial dimensions to the final judgment, and whether information integration changed over time. To make a conclusive inference, an outcome-tracing technique, where the derivation of joint from individual judgments is traced at each trial would have been crucial. The experimental task we developed easily accommodates for this type of analysis. Furthermore, in principle, the continuous response mode would allow one to track in future experiments exactly how joint location judgments were reached in the shared environment. This way, recordings of response adjustments could provide insights on the dynamical aspects of how the formation of joint judgments evolves in time. Pursuing this line of research may further enhance our understanding of the psychology of collective judgment and decision-making. Complementing this behavioral analysis of collective judgment formation in time

with elaborated models of information integration within individual minds, for example, with Bayesian models of belief updating (Peterson & Phillips, 1966; Slovic & Lichtenstein, 1971), can address the perhaps theoretically most pressing question on how the phenomena we observed on the group level are caused by information processing on the individual level.

Comparing the two studies reported in the current thesis allows us to draw some conclusions with regards to the role of the constraint of having to achieve an agreement on a joint decision or judgment. When pairs of participants had access to redundant visual information, they integrated individual information fairly well, regardless of whether they revised their own judgments after observing another's judgments (Study I) or whether they needed to agree on a common judgment (Study II). However, when pairs of participants had access to complementary visual information, only when *constrained* to provide a common judgment could they take the full advantage of the complementary information distribution between them and integrate individually held information in a near-optimal way.

One reason for this may be that without the constraint to achieve a judgment that reflected shared agreement, individuals were reluctant to give enough weight to the reliable component of their partner's judgment. When the constraint to agree on a joint judgment was present, they were pushed to do so by the partner. This would also explain the results of an earlier experiment by Sniezek and Henry (1990) where they tested the two-stage model of group judgment. They found that individuals generally do not converge at the stage of judgment revision, that their adjustments to initial judgments are relatively small, and that improvement mostly takes place during judgment formation while having to agree on a joint judgment.

A further relevant finding in the context of the present discussion is that even if individuals agree on a group judgment, they do not necessarily unanimously believe this to be the correct judgment. This was demonstrated in Arthur Jenness' experiment (1932) where he asked participants

to estimate the number of beans in a bottle. In one of his experimental groups[17] Jenness found that discussion made group (consensus) judgments more accurate relative to pre-discussion individual judgments, but post-discussion individual judgments did not improve relative to pre-discussion judgments. This is an astonishing demonstration of the fact that benefit from interpersonal information integration seems to be more related to the process of compromising individual positions than to the process of revision of individual positions in the light of exchanged evidence, and that it is the group that becomes wise, not the group members.

The fact that people generally do not (or do insufficiently) spontaneously converge towards one position after exchanging their evidence suggests that human social cognition may not be tailored to function in the "swarm" mode as in many other animal species (Couzin, 2009). Instead, human social cognition may be geared towards seeking a shared solution to intellectual problems. Collective decision-making in humans appears to be supported by a different kind of social mechanisms, that enables humans to pool and integrate individually accessible information and that ensure that two heads are better than one when agreements are sought. Psychological mechanisms that allow individuals to virtually constrain their action with a shared consensus requirement might comprise part of a larger mental toolkit which enables human minds to function in the "we mode" (Galotti & Frith, 2013) – a special regime of information processing where an organism conceives itself as part of a larger social entity. A cultural practice of efficient collective judgment and decision-making might be one aspect of the "we mode" enabled specifically in the human species or its immediate ancestors.

The current research has two broader implications that are related to the present focus on collective judgment and decision-making in environment-mediated modes of interaction and the

[17] The effect was found only in ad hoc groups selected for maximum agreement. Why this effect was not found in groups selected for maximum disagreement is difficult to interpret (see Lorge et al., 1958).

degree of structural overlap in individually held information. Questions regarding environment-mediated interactions are particularly relevant in the new era of remote electronic communication, where individuals inhabit a common global virtual environment. New modes of interactions are currently being studied to foster information exchange across large numbers of intelligent agents including human users of the World Wide Web. In this context, indirect interactions between virtual partners have the advantage that they are not bound to symbolic means of communication, that they can overcome cultural and linguistic barriers, and that they can pool information from larger and more diverse crowds. The present results suggest that such information pooling could lead to close to optimal integration of information from individuals when appropriate non-verbal ways of achieving agreement are built into the virtual environments.

The second aspect concerns the problem of integrating individually held information with variable amount of structural overlap. In particular, we studied integration of two-dimensional point estimates (judgments) under conditions of redundant and complementary information distribution. So far, the problem of complementarity has been mostly neglected in research on collective judgment and decision-making. One exception is the study by Budescu et al. (2003). This study focused on the relationship between overlap in forecasters' sources of information and confidence in an integrated forecast. The goal, however, of integrating judgments (including forecast judgments) across individuals is to make them more accurate, not just more confident. It therefore seems worth further investigating how people integrate information with different amounts of structural overlap in individually held information with a focus on quality of group decision or judgment in more complex multi-dimensional problems.

Several authors (Hastie, 1986; Hill, 1982) have suggested that inter-individual collaboration in intellectual tasks should be most beneficial when individuals' knowledge complements each other. The current work puts this assertion on strong theoretical and empirical grounds. The theorem of

Maximum Collective Benefit specifies that two people confronted with a situation of two-dimensional judgment (remember the Christmas tree example?) should accrue maximum collective benefit precisely when dimensions of individual high and low uncertainty complement each other. One may wonder whether the theorem can be extended to a larger number of dimensions. However, in the multi-dimensional case things get more complicated, and the necessary and sufficient conditions cannot be specified exactly. Generally, we can state that a group of individuals is expected to gain more benefit from collaboration, when lapses in individual expertise are compensated by (an)other member(s)'(s) expertise, and thus complement each other.

Empirically, our results demonstrate that two people can effectively integrate complementary quantitative information and provide a more reliable and accurate judgment than the most capable individual. We can expect that two experts with complementary expertise would benefit more from collaboration than two experts with much overlap in areas of expertise. However, our results also indicate that two experts would have to work in a tight coordination, and would need to come up with a shared solution to realize the collective benefit. The mere opportunity to interact and learn from each other is not sufficient, and the lack of agreement constraint may seriously undermine the benefit from collaboration.

To summarize, our studies demonstrate that a robust and pronounced collective benefit can be achieved when individuals have access to different information sources that complement each other. In such situations a collective benefit can be achieved even without symbolic communication. When individuals have complementary expertise, collaborating to form joint judgments makes the best of individual knowledge and discards individual weaknesses. This serves as a reminder that the true advantage of group interactions rests on latent diversity and that one way to reveal the power of this diversity is to combine multiple and different perspectives in the face of problems.

List of References

Akino, T., & Yamaoka, R. (2005). Trail discrimination signal of *Lasius japonicus* (Hymenoptera: Formicidae). *Chemoecology*, *15*(1), 21-30.

Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current Biology*, *14*(3), 257-262.

Anderson, N. H. (1981). *Foundations of information integration theory*. New York: Academic Press.

Anderson, N.H. (1991). A cognitive theory of judgment and decision. In N.H. Anderson (Ed.), *Contributions to information integration theory: Volume I: Cognition* (pp. 105-142). Hillsdale, NJ: Lawrence Erlbaum Associates.

Anderson, N. H., & Graesser, C. C. (1976). An information integration analysis of attitude change in group discussion. *Journal of Personality and Social Psychology*, *34*(2), 210-222.

Ariely, D., Tung Au, W., Bender, R. H., Budescu, D. V., Dietz, C. B., Gu, H., & Zauberman, G. (2000). The effects of averaging subjective probability estimates between and within judges. *Journal of Experimental Psychology: Applied*, *6*(2), 130-146.

Armstrong, J. S. (1985). *Long-range forecasting: From crystal ball to computer*. New York ETC.: John Wiley & Sons.

Armstrong, J. S. (2001). Combining forecasts. In J. S. Armstrong (Ed.), *Principles of forecasting: A handbook for researchers and practitioners* (pp. 417–439). New York: Kluwer.

Asch, S. (1952). *Social psychology*. Englewood Cliffs, NJ: Prentice Hall.

Asch, S. E. (1956). Studies of independence and conformity: I. A minority of one against a unanimous majority. *Psychological monographs: General and Applied*, *70*(9), 1-70.

Ashton, R. H. (1986). Combining the judgments of experts: How many and which ones?. *Organizational Behavior and Human Decision Processes*, *38*(3), 405-414.

Avraamides, M. N., Adamou, C., Galati, A., & Kelly, J. W. (2012). Integration of spatial relations across perceptual experiences. In C. Stachniss, K. Schill, & D. Uttal (Eds.), *Spatial cognition VIII: International Conference, Spatial Cognition 2012, Kloster Seeon, Germany, August 31-September 3, 2012, Proceedings*. Berlin: Springer. (pp.416–430). Springer-Verlag Berlin Heidelberg.

Bach, D. R., & Dolan, R. J. (2012). Knowing how much you don't know: a neural organization of uncertainty estimates. *Nature Reviews Neuroscience*, *13*(8), 572-586.

Bahrami, B., Olsen, K., Bang, D., Roepstorff, A., Rees, G., & Frith, C. (2012a). What failure in collective decision-making tells us about metacognition. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1594), 1350-1365.

Bahrami, B., Olsen, K., Bang, D., Roepstorff, A., Rees, G., & Frith, C. (2012b). Together, slowly but surely: The role of social interaction and feedback on the build-up of benefit in collective decision-making. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(1), 3-8.

Bahrami, B., Olsen, K., Latham, P. E., Roepstorff, A., Rees, G., & Frith, C. D. (2010). Optimally interacting minds. *Science*, *329*(5995), 1081-1085.

Batchelder, W. H., & Romney, A. K. (1986). The statistical analysis of a general Condorcet model for dichotomous choice situations. In B. Grofman & G. Owen (Eds.), *Decision research* (Vol. 2, pp. 103-112). Greenwich, CT: JAI Press.

Beach, L. R., & Mitchell, T. R. (1978). A contingency model for the selection of decision strategies. *Academy of Management Review*, *3*(3), 439-449.

Bechterev, W., & Lange, A. (1924) Die Ergebnisse des Experiments auf dem Gebiete der kollektiven Reflexologie. *Zeitschrift fur Experimentelle und Angewandte Psychologie.*, 24, 305-344.

Beckers, R., Deneubourg, J. L., & Goss, S. (1993). Modulation of trail laying in the ant Lasius niger (Hymenoptera: Formicidae) and its role in the collective selection of a food source. *Journal of Insect Behavior*, *6*(6), 751-759.

Beckers, R., Deneubourg, J. L., Goss, S., & Pasteels, J. M. (1990). Collective decision making through food recruitment. *Insectes Sociaux*, *37*(3), 258-267.

Beppu, A., & Griffiths, T. L. (2009). Iterated learning and the cultural ratchet. In *Proceedings of the 31st annual conference of the cognitive science society* (pp. 2089-2094). Austin, TX: Cognitive Science Society.

Beveridge, M. E. L., & Pickering, M. J. (2013). Perspective taking in language: integrating the spatial and action domains. *Frontiers in Human Neuroscience, 7,* 113-123.

Billings, R. S., & Scherer, L. L. (1988). The effects of response mode and importance on decision-making strategies: Judgment versus choice. *Organizational Behavior and Human Decision Processes*, *41*(1), 1-19.

Birnbaum, M. H., Wong, R., & Wong, L. K. (1976). Combining information from sources that vary in credibility. *Memory & Cognition*, *4*(3), 330-336.

Bonabeau, E., Dorigo, M., & Theraulaz, G. (1999). *Swarm intelligence: From natural to artificial systems*. New York: Oxford University Press.

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.

Brauner, E., & Scholl, W. (2000). Editorial: The information processing approach as a perspective for groups research. *Group Processes & Intergroup Relations*, *3*(2), 115-122.

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.

Brown, R. E. (1979). Mammalian social odors: a critical review. *Advances in the Study of Behavior*, *10*, 103-162.

Bruce, R. S. (1935). Group judgments in the fields of lifted weights and visual discrimination. *The Journal of Psychology*, *1*(1), 117-121.

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.

Calenbuhr, V., & Deneubourg, J. L. (1990). A model for trail following in ants: Individual and collective behaviour. In W. Alt & G. Hoffmann (Eds.), *Biological motion: Proceedings of a workshop held in Königswinter, Germany, March 16–19, 1989* (pp. 453–470). Berlin, Heidelberg: Springer Berlin Heidelberg.

Calvert, G., Spence, C., & Stein, B. E. (2004). *The handbook of multisensory processes*. MIT press.

Cartwright, D., & Zander, A. (1968). *Group dynamics.* New York: Harper & Row.

Castelfranchi, C. (2003). *When doing is saying – the theory of behavioral implicit communication*. Unpublished manuscript.

Castelfranchi, C. (2009, September). Tacitly communicating with our intelligent environment via our practical behavior and its traces. In *Proceedings of the 2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology-Volume 03* (pp. 323-326). IEEE Computer Society.

Clemen, R. T. (1989). Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting*, *5*(4), 559-583.

Conradt, L., & Roper, T. J. (2005). Consensus decision making in animals. *Trends in Ecology & Evolution*, *20*(8), 449-456.

Couzin, I. D. (2009). Collective cognition in animal groups. *Trends in Cognitive Sciences*, *13*(1), 36-43.

Couzin, I. D., & Krause, J. (2003). Self-organization and collective behavior in vertebrates. *Advances in the Study of Behavior*, *32*, 1-75.

Dale, R., Roche, J., Snyder, K., and McCall, R. (2008). Exploring action dynamics as an index of paired-associate learning. *PLoS ONE* 3, e1728. Retrieved from http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0001728

Dashiell, J. F. (1930). An experimental analysis of some group effects. *The Journal of Abnormal and Social Psychology*, *25*(2), 190-199.

Davis, J. H. (1996). Group decision making and quantitative judgments: A consensus model. In E. Witte & J. H. Davis (Eds.), *Understanding group behavior: Consensual action by small groups* (Vol. 1, pp. 35–59). Hillsdale, NJ: Erlbaum.

Davis, J. H., Hulbert, L., Au, W. T., Chen, X. P., & Zarnoth, P. (1997). Effects of group size and procedural influence on consensual judgments of quantity: The examples of damage awards and mock civil juries. *Journal of Personality and Social Psychology*, *73*(4), 703-718.

Dawes, R. M. (1979). The robust beauty of improper linear models in decision making. *American Psychologist*, *34*(7), 571-582.

Dawes, R. M., & Corrigan, B. (1974). Linear models in decision making. *Psychological Bulletin*, 81, 95–105.

De Biseau, J. C., Deneubourg, J. L., & Pasteels, J. M. (1991). Collective flexibility during mass recruitment in the ant *Myrmica sabuleti* (Hymenoptera: Formicidae). *Psyche*, *98*(4), 323-336.

Deneubourg, J. L., & Goss, S. (1989). Collective patterns and decision-making. *Ethology, Ecology & Evolution*, *1*(4), 295-311.

Deneubourg, J. L., Grégoire, J. C., & Le Fort, E. (1990). Kinetics of larval gregarious behavior in the bark beetle *Dendroctonus micans* (Coleoptera: Scolytidae). *Journal of Insect Behavior*, *3*(2), 169-182.

Detrain, C., & Deneubourg, J. L.(2009). Social cues and adaptive foraging strategies in ants. In S. Jarau, & M. Hrncir (Eds.), *Food exploitation by social insects: Ecological, behavioral, and theoretical approaches* (pp. 29–51). Boca Raton: CRC Press.

Deutsch, M., & Gerard, H. B. (1955). A study of normative and informational social influences upon individual judgment. *The Journal of Abnormal and Social Psychology*, *51*(3), 629-636.

Dietz, M., Roepstorff, A., & Wallentin, M. (2010). See it from my side! A fMRI study of perspective-taking using language. Poster presented at 16th Annual Meeting of the Organization for Human Brain Mapping, Barcelona, Spain.

Dipple, A., Raymond, K., & Docherty, M. (2014). General theory of stigmergy: Modelling stigma semantics. *Cognitive Systems Research*, *31*, 61-92.

Doms, M., & Van Avermaet, E. (1980). Majority influence, minority influence and conversion behavior: A replication. *Journal of Experimental Social Psychology*, *16*(3), 283-292.

Durrant-Whyte, H. F. (1988). Sensor models and multisensor integration. *The International Journal of Robotics Research*, *7*(6), 97-113.

Ebdon, D. *(*1988*). Statistics in geography (2nd ed. with corrections).* Oxford, UK: Blackwell.

Einhorn, H. J., & Hogarth, R. M. (1975). Unit weighting schemes for decision making. *Organizational Behavior and Human Performance*, *13*(2), 171-192.

Einhorn, H. J., & Hogarth, R. M. (1981). Behavioral decision theory: Processes of judgment and choice. *Journal of Accounting Research*, 1-31.

Einhorn, H. J., Hogarth, R. M., & Klempner, E. (1977). Quality of group judgment. *Psychological Bulletin*, *84*(1), 158-172.

Einhorn, H. J., Kleinmuntz, D. N., & Kleinmuntz, B. (1979). Linear regression and process-tracing models of judgment. *Psychological Review*, *86*(5), 465-485.

Elvers, G. C., & Sorkin, R. D. (1989, October). Detection and recognition of multiple visual signals in noise. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 33, No. 20, pp. 1383-1387). SAGE Publications.

Emporis standards. (2016*). Data Standards: Skyscraper (ESN 24419)*. Retrieved from https://www.emporis.com/building/standard/75/skyscraper

Ernst, M. O. (2006). A Bayesian view on multimodal cue integration. In G. Knoblich, I. M. Thornton, M. Grosjean, & M. Shiffrar (Eds.), *Human body perception from the inside out* (pp. 105-131). New York: Oxford University Press.

Ernst, M. O. (2010). Decisions made better. *Science*, *329*(5995), 1022-1023.

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, *415*(6870), 429-433.

Ernst, M. O., & Bülthoff, H. H. (2004). Merging the senses into a robust percept. *Trends in Cognitive Sciences*, *8*(4), 162-169.

Ferrell, W. R. (1985). Combining individual judgments. In G. Wright (Ed.), *Behavioral decision making* (pp. 111-145). Plenum, New York.

Fitzgerald, T. D., & Peterson, S. C. (1988). Cooperative foraging and communication in caterpillars. *BioScience*, *38*(1), 20-25.

Franklin, F. (1996). *Coordination without communication*. Unpublished manuscript. Retrieved from http://www.dbgroup.unimo.it/IIA/references/paper19.html

Freeman, J., Dale, R., & Farmer, T. (2011). Hand in motion reveals mind in motion. *Frontiers in Psychology*, *2*. Retrieved from http://journal.frontiersin.org/article/10.3389/fpsyg.2011.00059/full

Focardi, S., Deneubourg, J. L., & Chelazzi, G. (1985). How shore morphology and orientation mechanisms can affect the spatial organization of intertidal molluscs. *Journal of Theoretical Biology*, *112*(4), 771-782.

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.

Fusaroli, R., Rączaszek-Leonardi, J., & Tylén, K. (2014). Dialog as interpersonal synergy. *New Ideas in Psychology*, *32*, 147-157.

Gallotti, M., & Frith, C. D. (2013). Social cognition in the we-mode. *Trends in Cognitive Sciences*, *17*(4), 160-165.

Galton, F. (1907). Vox populi (the wisdom of crowds). *Nature*, *75*(7), 450-451.

Ganesh, G., Takagi, A., Osu, R., Yoshioka, T., Kawato, M., & Burdet, E. (2014). Two is better than one: Physical interactions improve motor performance in humans. *Scientific Reports*, *4*, Retrieved from http://www.nature.com/articles/srep03824

Gigerenzer, G., Todd, P., & the ABC Research Group. (1999). *Simple heuristics that make us smart*. New York: Oxford University Press.

Gigone, D., & Hastie, R. (1993). The common knowledge effect: Information sharing and group judgment. *Journal of Personality and Social Psychology*, *65*(5), 959-974.

Gigone, D., & Hastie, R. (1997). Proper analysis of the accuracy of group judgments. *Psychological Bulletin*, *121*(1), 149-167.

Glynn, S. J., & Henning, R. A. (2000, July). Can teams outperform individuals in a simulated dynamic control task?. In *Proceedings of the human factors and ergonomics society annual meeting* (Vol. 44, No. 33, pp. 6-141). Sage CA: Los Angeles, CA: SAGE Publications.

Gordon, K. (1924). Group judgments in the field of lifted weights. *Journal of Experimental Psychology*, *7*(5), 398-400.

Gorman, M. L., & Trowbridge, B. J. (1989). The role of odor in the social lives of carnivores. In J. L. Gittleman, (Ed.), *Carnivore behavior, ecology, and evolution* (pp. 57-88). Ithaca: Comstock Pub. Associates.

Graesser, C.C. (1991). A social averaging theorem for group decision making. In N.H. Anderson (Ed.), *Contributions to information integration theory: Volume II: Social* (pp. 1-40). Hillsdale, NJ: Lawrence Erlbaum Associates.

Graham, D. (1962). Experimental studies of social influence in simple judgment situations. *The Journal of Social Psychology*, *56*(2), 245-269.

Grassé, P.-P. (1959). La reconstruction du nid et les coordinations inter-individuelles chez *Bellicositermes natalensis* et *Cubitermes* sp. La th´eorie de la stigmergie: Essai d'interpr´etation du comportement des termites constructeurs. *Insectes Sociaux*, 6, 41–81.

Grassé, P.-P. (1963). Les phénomènes sociaux chez les Animaux. *Cahiers de l'Institut de Science économique appliquée*. Suppl. 139, V, pp. 7–23.

Halpin, Z. T. (1986). Individual odors among mammals: origins and functions. *Advances in the Study of Behavior*, *16*, 39-70.

Harries, C., Yaniv, I., & Harvey, N. (2004). Combining advice: The weight of a dissenting opinion in the consensus. *Journal of Behavioral Decision Making*, *17*(5), 333-348.

Harrison, S. J., & Richardson, M. J. (2009). Horsing around: spontaneous four-legged coordination. *Journal of Motor Behavior*, *41*(6), 519-524.

Harvey, N., & Fischer, I. (1997). Taking advice: Accepting help, improving judgment, and sharing responsibility. *Organizational Behavior and Human Decision Processes*, *70*(2), 117-133.

Harvey, N., Harries, C., & Fischer, I. (2000). Using advice and assessing its quality. *Organizational Behavior and Human Decision Processes*, *81*(2), 252-273.

Hastie, R. (1986). Experimental evidence on group accuracy. In B. Grofman & G. Owen (Eds.), *Decision research* (Vol. 2, pp. 129-157). Greenwich, CT: JAI Press.

Hastie, R., & Kameda, T. (2005). The robust beauty of majority rules in group decisions. *Psychological Review*, *112*(2), 494-508.

Henry, R. A. (1993). Group judgment accuracy: Reliability and validity of postdiscussion confidence judgments. *Organizational Behavior and Human Decision Processes*, *56*(1), 11-27.

Henry, R. A., Strickland, O. J., Yorges, S. L., & Ladd, D. (1996). Helping groups determine their most accurate member: The role of outcome feedback. *Journal of Applied Social Psychology*, *26*(13), 1153-1170.

Henze, N., Zirkler, B. (1990). A class of invariant consistent tests for multivariate normality. *Communications in Statistics-Theory and Methods*, *19*(10), 3595-3617.

Hill, G. W. (1982). Group versus individual performance: Are N+ 1 heads better than one?. *Psychological Bulletin*, *91*(3), 517-539.

Hinsz, V. B., Tindale, R. S., & Vollrath, D. A. (1997). The emerging conceptualization of groups as information processors. *Psychological Bulletin*, *121*(1), 43-64.

Hinsz, V. B. (1999). Group decision making with responses of a quantitative nature: The theory of social decision schemes for quantities. *Organizational Behavior and Human Decision Processes*, *80*(1), 28-49.

Hirokawa, R. Y, & Poole, M. S. (1986). *Communication and group decision making*. Newbury Park, CA: Sage.

Hogarth, R. M. (1978). A note on aggregating opinions. *Organizational Behavior and Human Performance*, *21*(1), 40-46.

Hogarth, R. M., & Karelaia, N. (2006). "Take-the-best" and other simple strategies: Why and when they work "well" with binary cues. *Theory and Decision*, *61*(3), 205-249.

Holland, O. E. (1996, March). Multiagent systems: Lessons from social insects and collective robotics. In *Adaptation, Coevolution and Learning in Multiagent Systems: Papers from the 1996 AAAI Spring Symposium* (pp. 57-62). Retrieved from https://vvvvw.aaai.org/Papers/Symposia/Spring/1996/SS-96-01/SS96-01-011.pdf

Hölldobler, B., & Wilson, E. O. (1990). *The ants*. Harvard University Press.

Huang, H., Ren, C., & Jin, S. (2008, June). “Signs” is the sign: Towards a unified view of stigmergy. In *ICIA 2008. International Conference on Information and Automation* (pp. 535-540). IEEE. Retrieved from https://doi.org/10.1109/ICINFA.2008.4608058

Ickes, W., & Gonzalez, R. (1994). "Social" cognition and social sognition from the subjective to the intersubjective. *Small Group Research*, *25*(2), 294-315.

Innami, I. (1994). The quality of group decisions, group verbal behavior, and intervention. *Organizational Behavior and Human Decision Processes*, *60*(3), 409-430.

Jackson, D. E., & Châline, N. (2007). Modulation of pheromone trail strength with food quality in Pharaoh's ant, *Monomorium pharaonis*. *Animal behaviour*, *74*(3), 463-470.

Jacobs, R. A. (2002). What determines visual cue reliability?. *Trends in cognitive sciences*, *6*(8), 345-350.

Janis, I. L., & Mann, L. (1977). *Decision making: A psychological analysis of conflict, choice, and commitment*. Free Press.

Jenness, A. (1932). The role of discussion in changing opinion regarding a matter of fact. *The Journal of Abnormal and Social Psychology*, *27*(3), 279-296.

Kameda, T., & Tindale, R. S. (2006). Groups as adaptive devices: Human docility and group aggregation mechanisms in evolutionary context. *Evolution and Social Psychology*, 317-341.

Kameda, T., Wisdom, T., Toyokawa, W., & Inukai, K. (2012). Is consensus-seeking unique to humans? A selective review of animal group decision-making and its implications for (human) social psychology. *Group Processes & Intergroup Relations*, *15*(5), 673-689.

Keil, D., & Goldin, D. (2005, July). Indirect interaction in environments for multi-agent systems. In *International Workshop on Environments for Multi-Agent Systems* (pp. 68-87). Springer Berlin Heidelberg.

Kent, J., & Leitner, M. (2007). Efficacy of standard deviational ellipses in the application of criminal geographic profiling. *Journal of Investigative Psychology and Offender Profiling*, *4*(3), 147-165.

Kennedy, J. F., Eberhart, R. C., & Shi, Y. (2001). *Swarm intelligence*. San Francisco: Morgan Kaufmann Publishers.

Kerr, N. L., & Tindale, R. S. (2004). Group performance and decision making. *Annual Reviw of Psychology.*, *55*, 623-655.

Klein, D., & Sprenger, J. (2015). Modelling individual expertise in group judgements. *Economics and Philosophy*, *31*(01), 3-25.

Klugman, S. F. (1945). Group judgments for familiar and unfamiliar materials. *The Journal of General Psychology*, *32*(1), 103-110.

Knight, H. C. (1921). *A comparison of the reliability of group and individual judgments* Unpublished master's thesis, Columbia University.

Knill, D. C., & Richards, W. (Eds.). (1996). *Perception as Bayesian inference*. Cambridge University Press.

Knill, D. C., & Saunders, J. A. (2003). Do humans optimally integrate stereo and texture information for judgments of surface slant?. *Vision Research*, *43*(24), 2539-2558.

Krause, J., Ruxton, G. D., & Krause, S. (2010). Swarm intelligence in animals and humans. *Trends in Ecology & Evolution*, *25*(1), 28-34.

Larrick, R.P., Mannes, A.E., Soll, J.B. (2012). The social psychology of the Wisdom of Crowds. In J.I. Krueger (Ed.), *Social judgment and decision making* (pp. 227 – 242). Psychology Press.

Latane, 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), 822-832.

Lehmann, E. L.; Casella, G. (1998). *Theory of point estimation* (2nd ed.). New York: Springer.

Lewis, T., & Marsh, L. (2016). Human stigmergy: Theoretical developments and new applications. *Cognitive Systems Research*, *38*, 1–3.

Libby, R., Trotman, K. T., & Zimmer, I. (1987). Member variation, recognition of expertise, and group performance. *Journal of Applied Psychology*, *72*(1), 81-87.

Lin, S. W., & Cheng, C. H. (2009). The reliability of aggregated probability judgments obtained through Cooke's classical model. *Journal of Modelling in Management*, *4*(2), 149-161.

Loftus, G. R., & Masson, M. E. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review*, *1*(4), 476-490.

Lorge, I., Fox, D., Davitz, J., & Brenner, M. (1958). A survey of studies contrasting the quality of group performance and individual performance, 1920-1957. *Psychological Bulletin*, *55*(6), 337-372.

Ma, W. J. (2010). Signal detection theory, uncertainty, and Poisson-like population codes. *Vision Research*, *50*(22), 2308-2319.

Mahmoodi, A., Bang, D., Olsen, K., Zhao, Y. A., Shi, Z., Broberg, K., ... & Roepstorff, A. (2015). Equality bias impairs collective decision-making across cultures. *Proceedings of the National Academy of Sciences*, *112*(12), 3835-3840.

Maloney, L.T. (2002). Statistical decision theory and biological vision. In D. Heyer, & R. Mausfeld (Eds.), In *Perception and the physical world: Psychological and philosophical* i*ssues in perception* (pp. 145 – 189). New York: Wiley.

Maloney, L. T., Trommershäuser, J., & Landy, M. S. (2007). Questions without words: A comparison between decision making under risk and movement planning under risk. In W.D. Gray(Ed.), *Integrated models of cognitive systems* (pp. 297-313). Oxford: Oxford University Press.

Marsh, L., & Onof, C. (2008). Stigmergic epistemology, stigmergic cognition. *Cognitive Systems Research*, *9*(1), 136-149.

Martin, R. (1998). Majority and minority influence using the afterimage paradigm: A series of attempted replications. *Journal of Experimental Social Psychology*, *34*(1), 1-26.

Martignon, L., & Hoffrage, U. (1999). Why does one-reason decision making work. In G. Gigerenzer, P. Todd., & the ABC Research Group (Eds), *Simple heuristics that make us smart*, 119-140.

Martignon, L., & Hoffrage, U. (2002). Fast, frugal, and fit: Simple heuristics for paired comparison. *Theory and Decision*, *52*(1), 29-71.

McGraw, J. J., Wallot, S., Mitkidis, P., & Roepstorff, A. (2014). Culture's building blocks: investigating cultural evolution in a LEGO construction task. *Frontiers in Psychology*, *5*. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4162372/pdf/fpsyg-05-01017.pdf

McKinstry, C., Dale, R., & Spivey, M. J. (2008). Action dynamics reveal parallel competition in decision making. *Psychological Science*, *19*(1), 22-24.

Metz, C. E., & Shen, J. H. (1992). Gains in accuracy from replicated readings of diagnostic images prediction and assessment in terms of ROC analysis. *Medical Decision Making*, *12*(1), 60-75.

Millar, S. K., Oldham, A. R., & Renshaw, I. (2012). Interpersonal coupling in rowing: The mediating role of the environment. *Jounal of Sport and Exercise Psychology*, *34*(Supp), 110-111.

Mooers, Jr., Robert, L. (1972). *Finding your way in the outdoors*. Outdoor Life Press.

Munsterberg, H. (2014). *Psychology and social sanity*. Floating Press. Retrieved from https://books.google.hu/books?id=eHlIBAAAQBAJ (Original work published 1914)

Mojzisch, A., & Krug, K. (2008). Cells, circuits, and choices: Social influences on perceptual decision making. *Cognitive, Affective, & Behavioral Neuroscience*, *8*(4), 498-508.

Moscovici, S. (1976). *Social influence and social change*. London: Academic Press.

Moscovici, S. (1980). Toward a theory of conversion behavior. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 13, pp. 209-239). New York: Academic Press.

Moscovici, S., Lage, E., & Naffrechoux, M. (1969). Influence of a consistent minority on the responses of a majority in a color perception task. *Sociometry*, 365-380.

Moscovici, S., & Personnaz, B. (1980). Studies in social influence: V. Minority influence and conversion behavior in a perceptual task. *Journal of Experimental Social Psychology*, *16*(3), 270-282.

Newell, F. N., Bülthoff, H. H., & Ernst, M. O. (2003, July). Cross-modal perception of actively explored objects. In I. Oakley, S. O'Modhrain, & F. Newell (Eds.), *Eurohaptics 2003 Conference Proceedings* (pp. 291-299). Trinity College, Dublin Ireland.

Newell, F. N., Ernst, M. O., Tjan, B. S., & Bülthoff, H. H. (2001). Viewpoint dependence in visual and haptic object recognition. *Psychological Science*, *12*(1), 37-42.

Ohtsubo, Y., Masuchi, A., & Nakanishi, D. (2002). Majority influence process in group judgment: Test of the social judgment scheme model in a group polarization context. *Group Processes & Intergroup Relations*, *5*(3), 249-261.

Oruç, I., Maloney, L. T., & Landy, M. S. (2003). Weighted linear cue combination with possibly correlated error. *Vision Research*, *43*(23), 2451-2468.

Page, S. E. (2008). *The difference: How the power of diversity creates better groups, firms, schools, and societies*. Princeton University Press.

Pasteels, J. M., Deneubourg, J. L., & Goss, S. (1987). Self-organization mechanisms in ant societies. I. Trail recruitment to newly discovered food sources. *Experientia Supplementum, 54,* 155-175.

Passino, K. M., Seeley, T. D., & Visscher, P. K. (2008). Swarm cognition in honey bees. *Behavioral Ecology and Sociobiology*, *62*(3), 401-414.

Patel, D., Fleming, S. M., & Kilner, J. M. (2012). Inferring subjective states through the observation of actions. *Proceedings of the Royal Society of London B: Biological Sciences*, *279*(1748), 4853-4860.

Pescetelli, N., Rees, G., & Bahrami, B. (2016). The perceptual and social components of metacognition. *Journal of Experimental Psychology: General*, *145*(8), 949-965.

Pete, A., Pattipati, K. R., & Kleinman, D. L. (1993a). Optimal team and individual decision rules in uncertain dichotomous situations. *Public Choice*, *75*(3), 205-230.

Pete, A., Pattipati, K. R., & Kleinman, D. L. (1993b). Distributed detection in teams with partial information: A normative-descriptive model. *IEEE Transactions on Systems, Man, and Cybernetics*, *23*(6), 1626-1648.

Preston, M. G. (1938). Note on the reliability and the validity of the group judgment. *Journal of Experimental Psychology*, *22*(5), 462-471.

Peterson, S. C. (1988). Chemical trail marking and following by caterpillars of *Malacosoma neustria*. *Journal of Chemical Ecology*, *14*(3), 815-824.

Peterson, C. R., & Phillips, L. D. (1966). Revision of continuous subjective probability distributions. *IEEE Transactions on Human Factors in Electronics*, (1), 19-22.

Raafat, R. M., Chater, N., & Frith, C. (2009). Herding in humans. *Trends in Cognitive Sciences*, *13*(10), 420-428.

Reynolds, C. W. (1987). Flocks, herds and schools: A distributed behavioral model. *ACM SIGGRAPH computer graphics*, *21*(4), 25-34.

Riccia, D. G., Lenz, H. J., & Kruse, R. (Eds.). (2001). *Data fusion and perception*. Springer.

Roessingh, P. (1989). The trail following behaviour of *Yponomeuta cagnagellus*. *Entomologia Experimentalis et Applicata*, *51*(1), 49-57.

Roessingh, P. (1990). Chemical marker from silk of *Yponomeuta cagnagellus*. *Journal of Chemical Ecology*, *16*(7), 2203-2216.

Roessingh, P., Peterson, S. C., & Fitzgerald, T. D. (1988). The sensory basis of trail following in some *lepidopterous larvae*: Contact chemoreception. *Physiological Entomology*, *13*(2), 219-224.

Rowe, G., & Wright, G. (1999). The Delphi technique as a forecasting tool: issues and analysis. *International Journal of Forecasting*, *15*(4), 353-375.

Schonbar, R. (1945). The interaction of observer pairs in judging visual extent and movement: the formation of social norms in "structured" situations. *Archives of Psychology,* No.299. NY.

Sebanz, N., Bekkering, H., & Knoblich, G. (2006). Joint action: bodies and minds moving together. *Trends in Cognitive Sciences*, *10*(2), 70-76.

Shanteau, J., & Nagy, G. F. (1984). Information integration in person perception: Theory and application. *Issues in Person Perception*, 48-86.

Sherif, M. (1936). *The psychology of social norms*. New York: Harper & Row.

Slovic, P., & Lichtenstein, S. (1971). Comparison of Bayesian and regression approaches to the study of information processing in judgment. *Organizational Behavior and Human Performance*, *6*(6), 649-744.

Smith, M. (1931). Group judgments in the field of personality traits. *Journal of Experimental Psychology*, *14*(5), 562-565.

Sniezek, J. A., & Henry, R. A. (1989). Accuracy and confidence in group judgment. *Organizational Behavior and Human Decision Processes*, *43*(1), 1-28.

Sniezek, J. A., & Henry, R. A. (1990). Revision, weighting, and commitment in consensus group judgment. *Organizational Behavior and Human Decision Processes*, *45*(1), 66-84.

Sniezek, J. A., Schrah, G. E., & Dalal, R. S. (2004). Improving judgment with prepaid expert advice. *Journal of Behavioral Decision Making*, *17*(3), 173-190.

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), 780-805.

Sorkin, R. D., & Dai, H. (1994). Signal detection analysis of the ideal group. *Organizational Behavior and Human Decision Processes*, *60*(1), 1-13.

Sorkin, R. D., Hays, C. J., & West, R. (2001). Signal-detection analysis of group decision making. *Psychological Review*, *108*(1), 183-203.

Sorkin, R. D., Luan, S., & Itzkowitz, J. (2001) Rational models of social conformity and social loafing. Paper presented at the European Association for Decision Making (SPUDM-18), Amsterdam, Netherlands, August 2001.

Sorkin, R. D., Luan, S., & Itzkowitz, J. (2004). Group decision and deliberation: A distributed detection process. In D. Koehler & N. Harvey (Eds.), *Blackwell handbook of judgment and decision making*, 464-483.

Sorrentino, R. M., King, G., & Leo, G. (1980). The influence of the minority on perception: A note on a possible alternative explanation. *Journal of Experimental Social Psychology*, *16*(3), 293-301.

Spence, C., & Driver, J. (2004). *Crossmodal space and crossmodal attention*. Oxford University Press.

Sperber, D., & Wilson D. (2001). *Relevance: Communication and cognition.* Oxford: Blackwell Publishers.

Stasser, G. (1999). A primer of social decision scheme theory: Models of group influence, competitive model-testing, and prospective modeling. *Organizational Behavior and Human Decision Processes*, *80*(1), 3-20.

Stein, B. E., & Meredith, M. A. (1993). *The merging of the senses*. The MIT Press.

Steiner, I.D. (1972). *Group process and productivity.* New York: Academic Press.

Stewart, T. R. (2001). Improving reliability of judgmental forecasts. In J. A. Armstrong (Ed.), *Principles of forecasting: A handbook for researchers and practitioners* (pp. 81–106). Boston, MA: Kluwer Academic Publishers.

Stroop, J. R. (1932). Is the judgment of the group better than that of the average member of the group?. *Journal of Experimental Psychology*, *15*(5), 550-562.

Sumpter, D. J., & Beekman, M. (2003). From nonlinearity to optimality: pheromone trail foraging by ants. *Animal Behaviour*, *66*(2), 273-280.

Surowiecki, J. (2004). *The wisdom of crowds: why the many are smarter than the few and how collective wisdom shapes business, economics, society and nations*. Little, Brown.

Tajfel, H. (1969). Social and cultural factors in perception. In G. Lindzey & E. Aronson (Eds.), *The handbook of social psychology* (Vol. 3, 2nd ed., pp. 315–394). Reading, MA: Addison-Wesley.

Taylor, J.R. (1997). *An Introduction to Error Analysis* (2nd ed.). Sausalito, CA: University Science Books.

Theraulaz, G., & Bonabeau, E. (1999). A brief history of stigmergy. *Artificial Life*, *5*(2), 97-116.

Traniello, J. F. A., & Robson, S. K. (1995). Trail and Territorial Communication in Social Insects. In R. T. Cardé & W. J. Bell (Eds.), *Chemical ecology of insects 2* (pp. 241–286). Boston, MA: Springer US.

Trömmershauser, J., Kording, K., & Landy, M. S. (Eds.). (2011). *Sensory cue integration*. Oxford University Press.

Trujillo-Ortiz, A., Hernandez-Walls, R., Barba-Rojo, K., & Cupul-Magana, L. (2007). Roystest: Royston's Multivariate Normality Test. [A MATLAB file]. Available from http://www. mathworks. com/matlabcentral/fileexchange/17811, 10.

Tummolini, L., Castelfranchi, C., Ricci, A., Viroli, M., & Omicini, A. (2004, July). "Exhibitionists" and "Voyeurs" do it better: A shared environment for flexible coordination with tacit messages. In *International Workshop on Environments for Multi-Agent Systems* (pp. 215-231). Springer Berlin Heidelberg.

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), 105-115.

Tylén, K., Weed, E., Wallentin, M., Roepstorff, A., & Frith, C. D. (2010). Language as a tool for interacting minds. *Mind & Language*, *25*(1), 3-29.

Uecker, W. C. (1982). The quality of group performance in simplified information evaluation. *Journal of Accounting Research*, 388-402.

van Beers, R. J., Sittig, A. C., & van der Gon Denier, J. J. (1996). How humans combine simultaneous proprioceptive and visual position information. *Experimental Brain Research*, *111*(2), 253-261.

van Beers, R. J., Sittig, A. C., & van der Gon, J. J. D. (1998). The precision of proprioceptive position sense. *Experimental Brain Research*, *122*(4), 367-377.

van Beers, R. J., Sittig, A. C., & van Der Gon, J. J. D. (1999). Integration of proprioceptive and visual position-information: An experimentally supported model. *Journal of Neurophysiology*, *81*(3), 1355-1364.

van der Wel, R. P., Knoblich, G., & Sebanz, N. (2011). Let the force be with us: Dyads exploit haptic coupling for coordination. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(5), 1420-1431.

van der Wel, R. P. R. D., Sebanz, N., & Knoblich, G. (2014). Do people automatically track others' beliefs? Evidence from a continuous measure. *Cognition*, *130*(1), 128-133.

Verhaeghe, J. C. (1982). Food recruitment in *Tetramorium impurum* (Hymenoptera: Formicidae). *Insectes Sociaux*, *29*(1), 67-85.

Vesper, C., Morisseau, T., Knoblich, G. & Sperber, D. (2016). Matching matchboxes: Co-actors create non-conventional communication systems for joint action. In T. Barkowsky, Z.F. Llansola, H. Schultheis, & J. van de Ven (Eds.), *Proceedings of the 13th Biannual Conference of the German Society for Cognitive Science* (pp. 47-50). Retrieved from http://elib.suub.uni-bremen.de/edocs/00105521-1.pdf

von Winterfeldt, D., & Edwards, W. (1986). *Decision analysis and behavioral research*. Cambridge, United Kingdom: Cambridge University Press.

Vul, E., & Pashler, H. (2008). Measuring the crowd within probabilistic representations within individuals. *Psychological Science*, *19*(7), 645-647.

Wallsten, T. S., Budescu, D. V., Erev, I., & Diederich, A. (1997). Evaluating and combining subjective probability estimates. *Journal of Behavioral Decision Making*, *10*(3), 243-268.

Wahn, B., Schmitz, L., Koenig, P., & Knoblich, G. (2016). Benefiting from Being Alike: Interindividual Skill Differences Predict Collective Benefit in Joint Object Control. In A. Papafragou, D.

Grodner, D. Mirman, & J.C. Trueswell (Eds.), *Proceedings of the 38th annual conference of the Cognitive Science Society* (pp. 2747-2752). Austin, TX: Cognitive Science Society.

Watson, W. E., Michaelsen, L. K., & Sharp, W. (1991). Member competence, group interaction, and group decision making: a longitudinal study. *Journal of Applied Psychology*, *76*(6), 803.

Welch, R.B. and Warren, D.H. (1986) Intersensory interactions. In K.R. Boff, L. Kaufman., & J.P. Thomas, (Eds.), *Handbook of perception and human performance: Sensory processes and perception* (pp.25.1-25.36). Wiley.

Wilson, E. O. (2000). *Sociobiology: The new synthesis*. Cambridge, MA: Harvard University Press. (Original work published 1975)

Yaniv, I. (1997). Weighting and trimming: Heuristics for aggregating judgments under uncertainty. *Organizational Behavior and Human Decision Processes*, *69*(3), 237-249.

Yaniv, I. (2004a). The benefit of additional opinions. *Current Directions in Psychological Science*, *13*(2), 75-78.

Yaniv, I. (2004b). Receiving other people's advice: Influence and benefit. *Organizational Behavior and Human Decision Processes*, *93*(1), 1-13.

Yaniv, I., & Choshen-Hillel, S. (2012). When guessing what another person would say is better than giving your own opinion: Using perspective-taking to improve advice-taking. *Journal of Experimental Social Psychology*, *48*(5), 1022-1028.

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.

Yaniv, I., & Milyavsky, M. (2007). Using advice from multiple sources to revise and improve judgments. *Organizational Behavior and Human Decision Processes*, *103*(1), 104-120.

Yeung, N., & Summerfield, C. (2012). Metacognition in human decision-making: Confidence and error monitoring. *Philosophical Transactions of the Royal Society B*, *367*(1594), 1310-1321.

Yuille, A., & Kersten, D. (2006). Vision as Bayesian inference: Analysis by synthesis?. *Trends in Cognitive Sciences*, *10*(7), 301-308.

Zarnowitz, V. (1992). The accuracy of individual and group forecasts. In Zarnowitz, V., *Business Cycles: Theory, History, Indicators, and Forecasting* (pp. 444-461). University of Chicago Press.

Zhang, H., Daw, N. D., & Maloney, L. T. (2013). Testing whether humans have an accurate model of their own motor uncertainty in a speeded reaching task. *PLoS Computational Biology*, *9*(5), e1003080. Retrieved from http://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1003080

Appendix A

Proof of the theorem of Maximum Collective Benefit

In this section the formal proof is provided for the theorem of Maximum Collective Benefit, formulated for two unbiased estimators (two different sensors or distinct judges) of a two-dimensional vector–valued quantity (see section 1.3 for its psychological formulation). We show that the variance of optimally linearly combined estimates is minimized when the eigenvectors corresponding to the larger or smaller eigenvalues of variance-covariance matrices of the two estimators are orthogonal to each other.

Notation. The transpose of any vector $\mathbf{v}$ or matrix $\mathbf{M}$ will be denoted as $\mathbf{v}'$ and $\mathbf{M}'$ respectively, and $\boldsymbol{I}$ will denote the identity matrix. Throughout this section capitalized boldfaced unsubscripted letters (e.g., $\boldsymbol{A}$) are used to refer to random vectors, and capitalized unboldfaced subscripted letters (e.g., X_A) are used to refer to random scalars. Let $\mathbf{a} = [x_a, y_a]'$ and $\mathbf{b} = [x_b, y_b]'$ be two observations of an estimated parameter $\boldsymbol{\theta}$ retrieved from two independent unbiased observers represented by bivariate random variables $\boldsymbol{A}$ and $\boldsymbol{B}$. Unbiasedness of observers implies that the two random variables share a common expected value equal to the true parameter value, that is,

$$\mathrm{E}(\boldsymbol{A}) = \mathrm{E}(\boldsymbol{B}) = \begin{bmatrix} x_0 \\ y_0 \end{bmatrix} = \boldsymbol{\theta}. \tag{A.1}$$

The two variables have possibly distinct variance-covariance matrices that characterize random error in the individual observations:

$$\boldsymbol{\Sigma}_i = \begin{bmatrix} \sigma^2_{X,i} & \mathrm{cov}(X_i, Y_i) \\ \mathrm{cov}(X_i, Y_i) & \sigma^2_{Y,i} \end{bmatrix}, \boldsymbol{\Sigma}_i \succ 0; \tag{A.2}$$

where $\mathrm{cov}(X_i, Y_i)$ is covariance between scalar random variables X_i and Y_i, $i \in (A; B)$. The independence assumption implies zero cross-covariance between $\boldsymbol{A}$ and $\boldsymbol{B}$, that is, $\rho(X_A, X_B) = \rho(Y_A,$

$Y_B) = \rho(X_A, Y_B) = \rho(X_B, Y_A) = 0$, where $\rho(i, j)$ denotes a correlation between any two scalar variables i an j.

By the spectral decomposition theorem (Halmos, 1963, p. 241-247), a variance-covariance matrix can be factored out into the form $\sum = \mathbf{VLV'}$, where $\mathbf{L} = \mathrm{Diag}(\mathbf{l}_1, \mathbf{l}_2)$ is a diagonal matrix whose entries $\mathbf{l}_i$ are eigenvalues of the variance-covariance matrix $\sum$, and $\mathbf{V} = [\ \mathbf{v}_1,\ \mathbf{v}_2\ ]$ is a matrix whose columns are eigenvectors corresponding to the eigenvalues of the variance-covariance matrix $\sum$, which satisfy the following condition: $\sum \mathbf{v}_\mathrm{d} = \mathbf{l}_\mathrm{d}\mathbf{v}_\mathrm{d}$ (Mardia, Kent, & Bibby, 1979), and $d \in (1; 2)$ is the precision dimension.

The two observations can be linearly combined to derive a more accurate estimation of $\boldsymbol{\theta}$ by taking their weighted average

$$\bar{\mathbf{m}} = \mathbf{Wa} + (\boldsymbol{I} - \mathbf{W})\mathbf{b}. \tag{A.3}$$

Our goal is to apply proper weights to the individual observations $\mathbf{a}$ and $\mathbf{b}$ to minimize the mean squared error (*MSE*) of the resulting estimates of $\boldsymbol{\theta}$. For vector estimations the *MSE* is given by the following general equation (Taboga, 2010):

$$\mathrm{MSE}(\widehat{\boldsymbol{\theta}}) = \mathrm{Tr}\left(\mathrm{Var}(\widehat{\boldsymbol{\theta}})\right) + \left\|\mathrm{Bias}(\widehat{\boldsymbol{\theta}}, \boldsymbol{\theta})\right\|^2, \tag{A.4}$$

where $\mathrm{Tr}\left(\mathrm{Var}(\widehat{\boldsymbol{\theta}})\right)$ is the trace of the variance-covariance matrix of an estimator $\widehat{\boldsymbol{\theta}}$, defined as the sum of its diagonal elements; and the $\mathrm{Bias}(\widehat{\boldsymbol{\theta}}, \boldsymbol{\theta}) = \mathrm{E}(\widehat{\boldsymbol{\theta}}) - \boldsymbol{\theta}$ term is the estimator's bias. From equations (A.1) and (A.3) it follows that the estimator $\bar{\mathbf{m}}$ derived through Eq. (A.3) is unbiased (its expected value is equal to the true value of $\boldsymbol{\theta}$), and that its *MSE* is determined solely by the trace of its variance-covariance matrix.

The combined estimator $\bar{\mathbf{m}}$ will have minimum variance exactly when weights $\mathbf{W}_i$ assigned to observation vectors ($\mathbf{a}$ and $\mathbf{b}$ in our case) are selected by taking the inverse of the variance-

covariance matrices characterizing the random error inherent in the observations, that is, when $\mathbf{W}_{\mathrm{i}} = \boldsymbol{\Sigma}_{\mathrm{i}}^{-1}$ (James, 2006; p. 324). The weighted estimate then is

$$\bar{\mathbf{m}} = (\boldsymbol{\Sigma}_{\mathrm{A}}^{-1} + \boldsymbol{\Sigma}_{\mathrm{B}}^{-1})^{-1}(\mathbf{W}_{\mathrm{A}}\mathbf{a} + \mathbf{W}_{\mathrm{B}}\mathbf{b}). \tag{A.5}$$

And the variance-covariance matrix of the combined estimator is given by

$$\boldsymbol{\Sigma}_{\bar{\mathrm{m}}}^{-1} = \boldsymbol{\Sigma}_{\mathrm{A}}^{-1} + \boldsymbol{\Sigma}_{\mathrm{B}}^{-1} = (\mathbf{V}_{\mathrm{A}}\mathbf{L}_{\mathrm{A}}\mathbf{V'}_{\mathrm{A}})^{-1} + (\mathbf{V}_{\mathrm{B}}\mathbf{L}_{\mathrm{B}}\mathbf{V'}_{\mathrm{B}})^{-1}. \tag{A.6}$$

In this case, the combined estimator $\bar{\mathbf{m}}$ is guaranteed to have lower variance than the random variables $\boldsymbol{A}$ and $\boldsymbol{B}$ realized in individual observations $\mathbf{a}$ and $\mathbf{b}$, and it will have lower *MSE* accordingly. Under the stipulated assumption that observers $\boldsymbol{A}$ and $\boldsymbol{B}$ are unbiased, the combined estimator $\bar{\mathbf{m}}$ derived through Eq. (A.5) is the best linear unbiased estimator of $\boldsymbol{\theta}$. If $\boldsymbol{A}$ and $\boldsymbol{B}$ are also Gaussian random variables, the weighted average vector $\bar{\mathbf{m}}$ derived through Eq. (A.5) is also the maximum likelihood estimate of $\boldsymbol{\theta}$.

We wish to specify the geometric relation between eigenvectors $\mathbf{V}_{\mathrm{A}}$ and $\mathbf{V}_{\mathrm{B}}$ corresponding to the variance-covariance matrices $\boldsymbol{\Sigma}_{\mathrm{A}}$ and $\boldsymbol{\Sigma}_{\mathrm{B}}$ with given eigenvalues $\mathbf{L}_{\mathrm{A}}$ and $\mathbf{L}_{\mathrm{B}}$ that minimizes the trace of the variance-covariance matrix $\boldsymbol{\Sigma}_{\bar{\mathrm{m}}}$ in Eq. (A.6), that is, when the *MSE* of optimally combined estimates is minimal.

Proposition 1. *Of any possible 2-by-2 variance-covariance matrices* $\boldsymbol{\Sigma}_{\mathrm{A}}$ *and* $\boldsymbol{\Sigma}_{\mathrm{B}}$ *with given eigenvalues* $\mathbf{L}_{\mathrm{A}} = \mathrm{Diag}(\mathbf{l}_{\mathrm{A},1}, \mathbf{l}_{\mathrm{A},2})$ *and* $\mathbf{L}_{\mathrm{B}} = \mathrm{Diag}(\mathbf{l}_{\mathrm{B},1}, \mathbf{l}_{\mathrm{B},2})$*, where eigenvalues* $\mathbf{l}_{\mathrm{i},1} < \mathbf{l}_{\mathrm{i},2}$ *correspond to the dimensions of low and high variance respectively , the trace of the variance-covariance matrix* $\boldsymbol{\Sigma}_{\bar{\mathrm{m}}} = (\boldsymbol{\Sigma}_{\mathrm{A}}^{-1} + \Sigma_{\mathrm{B}}^{-1})^{-1}$ *is minimized when the eigenvectors corresponding to the larger or smaller eigenvalues of the respective variance-covariance matrices* $\boldsymbol{\Sigma}_{\mathrm{A}}$ *and* $\boldsymbol{\Sigma}_{\mathrm{B}}$ *are orthogonal to each other, that is,* $\mathbf{v}_{\mathrm{A},1} \perp \mathbf{v}_{\mathrm{B},1}$ *, and* $\mathbf{v}_{\mathrm{A},2} \perp \mathbf{v}_{\mathrm{B},2}$.

Proof. (i) We start by reformulating the problem and expressing variance-covariance matrix $\sum$ with precision matrix $\mathbf{C} := \sum^{-1}$. Substituting for $\sum$ in Eq. (A.6) with precision matrix, we get

$$\mathbf{C}_{(\mathrm{A+B})} = \mathbf{C}_{\mathrm{A}} + \mathbf{C}_{\mathrm{B}} = \mathbf{V'}_{\mathrm{A}}\mathbf{L}_{\mathrm{A}}^{-1}\mathbf{V}_{\mathrm{A}} + \mathbf{V'}_{\mathrm{B}}\mathbf{L}_{\mathrm{B}}^{-1}\mathbf{V}_{\mathrm{B}}. \tag{A.7}$$

Denote the eigenvalues of matrices $\mathbf{C}_A$, $\mathbf{C}_B$, and of their sum $\mathbf{C}_{(A+B)}$ as λ_A, λ_B, and $\lambda_{(A+B)}$ respectively. Because the variance-covariance matrix is a positive-definite matrix (which we assume), the eigenvalues of a precision matrix are simply the inverses of the eigenvalues of the corresponding variance-covariance matrix, that is, $\lambda_d = l_d^{-1}$, $\lambda_{A,1} > \lambda_{A,2}$, and $\lambda_{B,1} > \lambda_{B,2}$. By the same property, its trace is equal to the sum of its eigenvalues. Hence, the trace of a variance-covariance matrix is equal to the sum of the inverses of the eigenvalues of the corresponding precision matrix. Then,

$$\mathrm{Tr}(\mathbf{\Sigma}_{\bar{m}}) = \lambda_{(A+B),1}^{-1} + \lambda_{(A+B),2}^{-1}. \qquad (A.8)$$

By the distributive property of the trace of a matrix, from Eq. (A.7) it follows that the trace of the precision matrix $\mathbf{C}_{(A+B)}$ is equal to the sum of eigenvalues of $\mathbf{C}_A$ and $\mathbf{C}_B$, that is,

$$\lambda_{(A+B),1} + \lambda_{(A+B),2} = \lambda_{A,1} + \lambda_{A,2} + \lambda_{B,1} + \lambda_{B,2}. \qquad (A.9)$$

Consequently, the sum of the two eigenvalues of $\mathbf{C}_{(A+B)}$ (this parameter can be read as the *total obtainable precision*) is solely determined by the eigenvalues $\mathbf{L}_A^{-1}$ and $\mathbf{L}_B^{-1}$, and is invariant to the eigenvectors $\mathbf{V}_A'$ and $\mathbf{V}_B'$. Since for two variance-covariance matrices with given sets of eigenvalues the total obtainable precision is fixed, we can normalize it for convenience,

$$\lambda_{(A+B),1} + \lambda_{(A+B),2} = 1. \qquad (A.10)$$

Then we can express the trace of the variance-covariance matrix $\mathbf{\Sigma}_{\bar{m}}$ in terms of the absolute difference between the eigenvalues, $\Delta\lambda$, of its precision matrix $\mathbf{C}_{(A+B)}$,

$$\mathrm{Tr}(\mathbf{\Sigma}_{\bar{m}}) = f(\Delta\lambda) = \lambda_{(A+B),1}^{-1} + \lambda_{(A+B),2}^{-1} = \frac{1}{0.5 + 0.5\Delta\lambda} + \frac{1}{0.5 - 0.5\Delta\lambda}. \qquad (A.11)$$

The derivative of the function described by Eq.(A.11) is

$$f'(\Delta\lambda) = \frac{8\Delta\lambda}{(\Delta\lambda^2 - 1)^2}. \qquad (A.12)$$

It can be seen that the function described by Eq. (A.12) is equal to zero if and only if $\Delta\lambda = 0$, and it monotonically increases in the range $0 \leq \Delta\lambda < 1$. Thus, Tr($\boldsymbol{\Sigma}_{\bar{\mathrm{m}}}$) is minimized when the difference between the eigenvalues of the precision matrix $\mathbf{C}_{(\mathrm{A+B})}$ is minimal.

(ii) Minimization of the difference between two eigenvalues given their sum (which we have shown to be fixed by the conditions of the problem) can be implemented by maximizing the second (smaller) eigenvalue. Now, we turn to specifying the maximum second eigenvalue, $\lambda_{(\mathrm{A+B}),2}$, of the precision matrix $\mathbf{C}_{(\mathrm{A+B})}$for given $\mathbf{L}_{\mathrm{A}}^{-1}$ and $\mathbf{L}_{\mathrm{B}}^{-1}$. By our definition, the absolute maximum value the second eigenvalue can take is as large as that of the first eigenvalue, that is, $\lambda_{(\mathrm{A+B}),2} = \lambda_{(\mathrm{A+B}),1} = (\sum_{\mathrm{i}} \sum_{\mathrm{d}} \lambda_{\mathrm{i,d}})/2$, where $i \in (A; B)$, $d \in (1; 2)$. However, a solution which satisfies this condition does not necessarily exist for two variance-covariance matrices with given sets of eigenvalues. To find the constraints, we can use the property that any covariance matrix $\boldsymbol{\Sigma}$ is a symmetric and positive-definite Hermitian matrix (Strang, 1988). Hence, the precision matrix $\mathbf{C}$, which is the inverse of $\boldsymbol{\Sigma}$, is also a Hermitian matrix, which can be seen from Eq. (A.7). Therefore, we can apply properties of a sum of two Hermitian matrices to $\mathbf{C}_{(\mathrm{A+B})}$. Then, the upper bound of $\lambda_{(\mathrm{A+B}),2}$ is subject to Weyl's dual inequality (Knutson & Tao, 2001; Weyl, 1912):

$$\lambda_{(\mathrm{A+B}),2} \leq \lambda_{\mathrm{A},2} + \lambda_{\mathrm{B},1}. \tag{A.13}$$

For any 2×2 Hermitian matrix $\mathbf{C}_{\mathrm{A}}$ there always exists a matrix $\mathbf{C}_{\mathrm{B}}$ with a set of eigenvalues $\mathbf{L}_{\mathrm{B}}^{-1}$, that for their sum $\mathbf{C}_{(\mathrm{A+B})}$ inequality (A.13) is saturated (Knutson & Tao, 2001). Without a loss of generality, assume that $(\lambda_{\mathrm{A},1} + \lambda_{\mathrm{B},2}) > (\lambda_{\mathrm{A},2} + \lambda_{\mathrm{B},1})$. In this case, from Eq. (A.9) it follows that $\Delta\lambda_{(\mathrm{A+B})}$ takes its minimum value when inequality (A.13) is an equality.

We now seek for the conditions when inequality (A.13) is an equality. Without a loss of generality assume that we choose a coordinate system aligned with the eigenvectors of $\boldsymbol{\Sigma}_{\mathrm{A}}$, so that

$$\mathbf{C}_{\mathrm{A}} = \mathbf{L}_{\mathrm{A}}^{-1} = \begin{bmatrix} \lambda_{\mathrm{A},1} & 0 \\ 0 & \lambda_{\mathrm{A},2} \end{bmatrix}. \tag{A.14}$$

The precision matrix $\mathbf{C}_{\mathrm{B}}$ is unknown, but because variance-covariance is always symmetric about the diagonal, it can be described in the following form:

$$\mathbf{C}_{\mathrm{B}} = \begin{bmatrix} a & c \\ c & b \end{bmatrix}. \tag{A.15}$$

The sum of precision matrices $\mathbf{C}_{\mathrm{A}}$ and $\mathbf{C}_{\mathrm{B}}$, accordingly, equals to

$$\mathbf{C}_{\mathrm{A+B}} = \begin{bmatrix} (\lambda_{\mathrm{A},1} + a) & c \\ c & (\lambda_{\mathrm{A},2} + b) \end{bmatrix}. \tag{A.16}$$

To find the entries of $\mathbf{C}_{\mathrm{B}}$ under which inequality (A.13) is an equality, we will use:

1) the property of a square matrix that its determinant is equal to the product of its eigenvalues (Beauregard & Fraleigh , 1973, p. 307), and therefore:

$$\left|\mathbf{C}_{(\mathrm{A+B})}\right| = \lambda_{(\mathrm{A+B}),1}\lambda_{(\mathrm{A+B}),2}; \tag{A.17}$$

2) an alternative parameterization of the determinant of a 2×2 matrix via its entries (Weisstein, 2016):

$$\det\begin{bmatrix} a & c \\ c & b \end{bmatrix} = \begin{vmatrix} a & c \\ c & b \end{vmatrix} = ab - c^2; \tag{A.18}$$

3) and the characteristic equation of a matrix (Weisstein, 2016):

$$|\mathbf{M} - \lambda \boldsymbol{I}| = 0, \tag{A.19}$$

where λ is an eigenvalue of $\mathbf{M}$. Using equations A.16 – A.19 we can assert that if $\lambda_{\mathrm{A+B}}$ is a possible eigenvalue of $\mathbf{C}_{\mathrm{A+B}}$, there must exist a solution to the following system of equations:

$$(\lambda_{\mathrm{A},1} + a - \lambda_{(\mathrm{A+B})}) \times (\lambda_{\mathrm{A},2} + b - \lambda_{(\mathrm{A+B})}) - c^2 = 0, \tag{A.20a}$$

$$a + b = \lambda_{\mathrm{B},1} + \lambda_{\mathrm{B},2}, \tag{A.20b}$$

$$ab - c^2 = \lambda_{\mathrm{B},1}\lambda_{\mathrm{B},2}. \tag{A.20c}$$

Equation (A.20a) is the characteristic equation of $\mathbf{C}_{\mathrm{A+B}}$ with determinant parameterized as in Eq. (A.18). Equation (A.20b) specifies the trace of $\mathbf{C}_{\mathrm{B}}$, and equation (A.20c) specifies its determinant being parameterized as in Eq. (A.18) on the left side of the equation, and as in Eq. (A.17) on the right side of the equation.

Substituting ($\lambda_{\mathrm{A,1}} + \lambda_{\mathrm{B,2}}$) for $\lambda_{(\mathrm{A+B})}$ in Eq. (A.19a) yields a unique solution of the form

$$a(\lambda_{\mathrm{A,2}} - \lambda_{\mathrm{A,1}}) = \lambda_{\mathrm{B,2}}(\lambda_{\mathrm{A,2}} - \lambda_{\mathrm{A,1}}). \tag{A.21}$$

Corollary, solving for b and c, the unique solution to $\mathbf{C}_{\mathrm{B}}$, under which inequality (A.13) is an equality is

$$\mathbf{C}_{\mathrm{B}} = \begin{bmatrix} \lambda_{\mathrm{B,2}} & 0 \\ 0 & \lambda_{\mathrm{B,1}} \end{bmatrix}. \tag{A.22}$$

Now we can apply eigendecomposition to $\mathbf{C}_{\mathrm{A}}$ and $\mathbf{C}_{\mathrm{B}}$. Consistently constructing the matrices of eigenvalues so that that the first dimension corresponds to the highest precision,

$$\mathbf{L}_{\mathrm{i}}^{-1} = \begin{bmatrix} \lambda_{\mathrm{i,1}} & 0 \\ 0 & \lambda_{\mathrm{i,2}} \end{bmatrix}, \tag{A.23}$$

and applying Eq. (A.7), we obtain the following eigenvectors corresponding to the smallest eigenvalues of $\boldsymbol{\Sigma}_{\mathrm{A}}$ and $\boldsymbol{\Sigma}_{\mathrm{B}}$: $\mathbf{v}_{\mathrm{A,1}} = [1\ 0]'$, and $\mathbf{v}_{\mathrm{B,1}} = [0\ 1]'$. Accordingly, $\langle \mathbf{v}_{\mathrm{A,1}} | \mathbf{v}_{\mathrm{B,1}} \rangle = 0$, which satisfies the condition of orthogonality, hence, $\mathbf{v}_{\mathrm{A,1}} \perp \mathbf{v}_{\mathrm{B,1}}$. Since eigenvectors of symmetric matrices with distinct eigenvalues are always orthogonal (Abadir & Magnus, 2005), $\mathbf{v}_{\mathrm{A,1}} \perp \mathbf{v}_{\mathrm{B,1}} \cup \mathbf{v}_{\mathrm{A,2}} \perp \mathbf{v}_{\mathrm{B,1}}$, which is what we set out to prove. ■

As a side note, one can notice that if $\lambda_{\mathrm{B,1}} = \lambda_{\mathrm{B,2}}$, the solution returned via Eq. (A.20a) and (A.20b) is $a = \lambda_{\mathrm{B,2}} = b$ (which is also the only possible value diagonal entries of $\mathbf{C}_{\mathrm{B}}$ can take in this scenario), and if $\lambda_{\mathrm{A,1}} = \lambda_{\mathrm{A,2}}$, while $a = \lambda_{\mathrm{B,2}}$ satisfies the equation (A.20a), it is not a unique solution, and trying to solve Eq. (A.20a) incurs division by zero. This makes sense, because, if this condition holds, the determinant of the sum of $\mathbf{C}_{\mathrm{A}}$ and $\mathbf{C}_{\mathrm{B}}$ is invariant to the rotation of $\mathbf{C}_{\mathrm{B}}$. Put it another way,

if a variance-covariance matrix is "circular" (has equal eigenvalues), the corresponding transformation matrix performs uniform scaling, and its eigenvectors include all possible non-zero vectors, including the eigenvectors of a second variance-covariance matrix. Therefore, the relations of orthogonality cannot be defined for this situation. However, if the two distributions are not isotropic, the condition if orthogonality is both necessary and sufficient when the overall variance of the optimally weighted average of the two estimates is minimum possible.

List of references used in the proof:

Abadir, K. M., & Magnus, J. R. (2005). *Matrix algebra.* Cambridge: Cambridge University Press.

Beauregard, R. A., & Fraleigh, J. B. (1973). *A first course in linear algebra: With optional introduction to groups, rings, and fields*. Boston: Houghton Mifflin.

Halmos, P. R. (1963). What does the spectral theorem say?. *The American Mathematical Monthly*, *70*(3), 241-247.

Horn, R. A.; Johnson, C. R. (2013). *Matrix analysis* (2nd ed.). Cambridge University Press.

James, F. (2006). *Statistical methods in experimental physics*. Hackensack, NJ: World Scientific.

Knutson, A., & Tao, T. (2001). Honeycombs and sums of Hermitian matrices. *Notices of American Mathematical Society*, *48*(2), 175-186.

Mardia, K. V., Kent, J. T., & Bibby, J. M. (1979). *Multivariate analysis*. New York: Academic Press.

Strang, G. (1988). *Linear algebra and its applications* (3rd ed.). San Diego: Harcourt Brace Jovanovich.

Taboga, M. (2010). *Lectures on probability theory and mathematical statistics*. Retrieved from https://www.statlect.com/glossary/mean-squared-error

Weisstein, E.W. (n.d.). Determinant. In *MathWorld - a Wolfram web resource*. Retrieved from http://mathworld.wolfram.com/Determinant.html

Weyl, H. (1912). Das asymptotische Verteilungsgesetz der Eigenwerte linearer partieller Differentialgleichungen (mit einer Anwendung auf die Theorie der Hohlraumstrahlung). *Mathematische Annalen*, *71*(4), 441-479.

ACKNOWLEDGEMENT

TO EXTERNAL FUNDING AGENCIES
CONTRIBUTING TO PHD DISSERTATIONS

Name of doctoral candidate: Pavel Valeryevich Voinov

Title of dissertation: *Interpersonal Information Integration in Judgment Revision and Collective Judgment Formation - The Benefits of Distributed Access to Redundant and Complementary Visual Information in a Shared Environment*

Name of supervisor(s): Gunther Klaus Knoblich, Natalie Sebanz

External funding agency: **European Research Council**

Acknowledgement:

This research has received funding from the European Research Council (ERC) under the European Union's Seventh Framework Programme (FP7/2007-2013) under **grant agreement No 616072 – JAXPERTISE.**

This research has received funding from the European Research Council (ERC) under the European Union's Seventh Framework Programme (FP7/2007-2013) under **grant agreement No 609819 – SOMICS.**