Social cognitive inference in interest-based bargaining

Mihael Ales Jeklic

Thesis submitted for the degree of Doctor of Philosophy

University College London

Declaration

I, Mihael Ales Jeklic, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

September 2021

Abstract

Negotiating rationally means ‘making the best decisions to maximize your interests’ (Bazerman & Neale, 1992). This thesis develops and tests a theoretical proposition that the quality of these negotiating decisions – and the behavior of negotiators and ultimately their bargaining outcomes – critically depends on the processes of social cognitive inference (mentalizing). Because the sole purpose of negotiation is to satisfy motivating mental states of negotiators (interests), mentalizing must underpin bargaining.

The theoretical part of the thesis undertakes a targeted review of fields of mentalizing and negotiation, creating a conceptual platform for the novel proposition that social cognitive inference (mentalizing) underpins negotiation. Three studies test two key predictions stemming from this theoretical proposal: that individual differences in the capacity to mentalize correlate with both value creating and value claiming in negotiation. The findings suggest that mentalizing predicts (i) both value creating and value claiming in multi-issue negotiation, (ii) value claiming in a single-issue distributive negotiation, and (iii) odds of settling partisan perception-driven disputes.

The second part of the thesis explores a proposition that the ‘negotiation’ task construal biases social cognition and the negotiators’ strategic choice toward competition, resulting in depressed individual and joint gain. The underlying theory is based on Friston’s active inference (free energy minimization) framework. Because negotiation situations are markedly ambiguous and uncertain, negotiators’ inference must rely heavily on priors, which in bargaining tend to be competitive. This depresses gains. We test this in four studies. The findings show that understanding a task as ‘negotiation’ (versus an alternative collaborative frame) (i) inhibits integrative and compatible aspects of joint gain in multi-issue tasks, (ii) accounts for variance not explained by manipulating trust in the negotiating partner, (iii) biases negotiators’ strategies toward contending and away from problem-solving where (iv) these strategies mediate the effect of the ‘negotiation’ construal on negotiation outcomes in a task with hidden value potential.

Impact statement

We are ineffective negotiators. While negotiation situations offer opportunities to generate value, we predictably and dramatically fail to capture it. Where agreement is possible, impasses are commonplace; optimal outcomes are rare, as low as four percent; where parties have identical preferences, up to forty percent fail to reach them. This startling inefficiency incurs significant personal and social costs, wastes resources and productivity, and increases conflict.

The present thesis has considerable practical importance because of the pervasiveness of negotiation in human affairs and the vast amount of value at stake, coupled with the evidence of value destruction and risk inherent in competitive negotiation.

Our current situation is particularly precarious. SARS-CoV-2 has damaged economies and livelihoods, and widened the existing inequalities between nations and individuals, particularly the gaps based on gender, race and income. The global economy is at its lowest level it has ever been in peacetime. The geopolitical challenges of the tensions between the East and the West, the divided America and post-Brexit United Kingdom, the rising neo-statism (or G-zero mentality) and nationalism, cyber risks, and the ever more urgent climate change, suggest that we will have to negotiate many aspects of our lives. These challenges will involve negotiation.

At the practical (non-academic) level, our research ought to benefit commercial activity, social enterprise and professional practice. It highlights the social cognitive capacities in individuals that predict negotiation performance, and our lessons will inform process interventions and negotiation education. Our research also shows that we are limited by the competitive mindset evoked by the ‘negotiation’ frame, which results in increased contending and depressed problem-solving, causing poor outcomes. We provide the tools for naming, confronting, and perhaps changing these phenomena.

We hope that the impact will occur locally, regionally, nationally and internationally, and benefit individuals, communities and organizations. The benefits should be incremental as the knowledge seeps into practice through training and education, and as the research evolves further.

Finally, we expect the findings of our research to contribute to the negotiation scholarship. We provide a novel comprehensive theory that combines the fields of negotiation and social cognition and has the potential to explain (and predict) both negotiation processes and outcomes. The implications ought to spur further research. Understanding the social cognitive capacities that drive outcomes and the nature of the human competitive predilection in negotiation settings are worthwhile tasks.

In sum, our research carries considerable impact. At the academic level, the novel theory contributes to the negotiation scholarship and should generate new research. At the practical level, our research ought to facilitate improvements in negotiation processes and outcomes.

Table of contents

List of Tables

List of Figures

Acknowledgements

I have been privileged and humbled to have the support of some remarkable people on this research journey. My PhD would have never happened without:

Bruce Patton, who taught me negotiation.

Prof. Jan Dalhuisen, who trusted me with his students, let me set up the first negotiation course at KCL, and supported me in academia ever since.

Prof. Peter Fonagy, my supervisor, who with endless patience guided my research, was always there for me, nudged me closer to thinking like a scientist, and provided a role model that I aspire to. In many ways, this PhD is his work. Thank you, Peter.

Prof. Asim Khwaja and his team at UCLH, who saved my life.

Katia and *Ethan,* my everything. To you I dedicate this thesis.

[page intentionally left blank]

In Thompson's mind was this thought: Khrushchev's gotten himself in a hell of a fix. He would then think to himself, 'My God, if I can get out of this with a deal that I can say to the Russian people: "Kennedy was going to destroy Castro and I prevented it."' Thompson, knowing Khrushchev as he did, thought Khrushchev will accept that. And Thompson was right. That's what I call empathy. We must try to put ourselves inside their skin and look at us through their eyes, just to understand the thoughts that lie behind their decisions and their actions.

McNamara, The Fog of War (2004)

The ruler who is ignorant of the designs of neighboring princes, cannot treat with them.

Sun Tzu, The Art of War (5th century BC)

I have had a philosophy for some time in regard to SALT [Strategic Arms Limitation Talks], and it goes like this: the Russians will not accept a SALT treaty that is not in their best interest, and it seems to me that if it is in their best interest, it can't be in our best interest.

Senator Spence Floyd,
cited in Bazerman and Neale (1992, p. 19)

Introduction

Negotiating rationally means ‘making the best decisions to maximize your interests’ (Bazerman & Neale, 1992). The theoretical proposition developed and tested in this thesis is that the quality of these negotiating decisions, and consequently the behavior of negotiators and ultimately negotiation outcomes, crucially depend on the processes of social cognitive inference ('mentalizing'; Fonagy, Gergely, Jurist, & Target, 2004; Heyes & Frith, 2014). The argument is as follows. Negotiators’ interests and their beliefs (policies) about optimal bargaining behavior are mental states. The choice of action in negotiation depends on understanding these mental states in oneself and in the counterparty, and this understanding requires mentalizing. Social cognitive inference underpins both value claiming and value creating. There is no negotiation without mentalization.

The empirical studies in the present thesis test key aspects of this theory: first, that individual differences in the capacity to mentalize predict negotiation success in value creating and value claiming, and second, that the construal of a task as ‘negotiation’ biases negotiators’ social cognition and strategic choice toward competition, and consequently inhibits optimization and depresses dyadic value.

Part one of the thesis contains the general theoretical proposition that mentalizing underpins bargaining and three studies that test the impact of mentalizing on value generating and value claiming in different negotiation tasks. Part two of the thesis introduces the impact of situational construal as a modifier of social cognition and its impact on strategic choice and outcomes in negotiation, and tests this in three studies. The thesis concludes with a general section that discusses the model of mentalizing in negotiation, integrates the findings of the studies, and outlines directions for future research.

Part I:

Social-cognitive inference in negotiation

The theoretical part of this thesis starts with a targeted review of negotiation and mentalizing, two conceptually alien fields heretofore not comprehensively considered together. This is necessary to establish a conceptual platform for the original contribution of this thesis: the proposition that mentalizing underpins negotiation.

The first chapter introduces mentalizing as perceiving and thinking about behavior in terms of mental states and explores its key evolutionary, developmental and neurobiological aspects. It starts by a brief overview of the evolutionary underpinnings of social cognitive inference and continues with the key features of mentalizing organized around the four polarities between implicit and explicit, cognitive and affective, self- and other-focused, and internally and externally focused mentalistic inference. It continues with an overview the developmental stages of mentalizing, and concludes with the phenomenology and neurobiology of poor mentalizing.

The second chapter discusses negotiation as an interdependent, mixed-motive interaction aimed at satisfying what the negotiation theory calls 'interests': whatever negotiators care about in a particular bargaining setting. It presents the building blocks of the negotiation theory including distributive and integrative negotiation, interests and value in negotiation, and continues with the tension between creating and claiming value, the key aspect of mixed-motive negotiation.

In the third chapter we turn to the key proposition of this thesis. The main claim is that social cognitive inference - making sense of behavior in terms of mental states – facilitates the parties' understanding of the motivators of negotiation and beliefs about optimal action (bargaining strategies) and thus underpins negotiators' strategic choice. This pertains to both value creating and value claiming. In other words, if negotiation is about taking action that satisfies interests, and interests and beliefs (policies) about optimal action are mental states, and further if mentalizing is the process required to represent and manipulate such mental states, mentalizing must underpin bargaining. The general model states that negotiators mentalize to infer their own and (hidden) counterparty's mental states from their (observable) behavior, and to decide their own action by inferring the likely reaction of the counterparty (based on their modelled mental states). We then discuss the theoretical implications of this theory, specifically how prescriptive advice for distributive and integrative negotiation is essentially a prescription to mentalize, and the how specific types of impaired mentalizing ought to interfere with the bargaining process.

Chapters four, five and six are empirical studies. Our theory that mentalizing underpins bargaining generates a specific testable prediction that the capacity to mentalize predicts both value creating and value claiming in negotiation. This hypothesis is tested in a multi-issue negotiation task (Study 1), in a zero-sum distributive task (Study 2) and in a partisan perception-driven dispute (Study 3).

MENTALIZING: HOLDING MIND IN MIND

Processes and pervasiveness of mentalizing

We navigate the social world by keeping others' minds in mind. Constant consideration of mental states of other people – what they think, feel, hope, fear, love, hate – is a fundamental component of social existence. Predicting how other people are likely to behave is a critical capacity in any social group and depends on understanding of others' minds. Basic social interactions – e.g., telling somebody something so they understand what you mean, pretending you like the gift you've received when you do not, taking turns in a discussion, navigating a walk on a busy pavement full of pedestrians, waving a hand to stop a taxi, understanding a train sign stating 'unattended luggage may cause delays' – requires social cognitive inference. The more complex social interaction is, the more sophisticated mentalizing tends to be required: marketing executives consider customers' preferences, lawyers make arguments with a judge's mind in mind, military brass forge strategies pre-empting enemy moves, and state leaders make ultimatums anticipating the behavior of their counterparts. We are capable of high-order reflection on our own thoughts (first order intentionality, e.g., 'I think the keys are there'), as well as on thoughts of others (second order intentionality. e.g., 'I think you think the keys are there'), and their thoughts about our thoughts (third order intentionality, e.g., 'You think that I think that you think the keys are there') and so on, all the way to fifth or sixth order of such reflective thinking.

Perhaps as the result of the priority granted by evolution, mentalizing is pervasive to the point of being excessive. We tend to ascribe intentions not just to people, but also, across the board, to things that do not have a mind, such as groups ('immigrants come to England to get free healthcare'), teams ('Arsenal want to win the Premiership), nation-states ('Greece does not want to pay its debt'), animals ('Bobby, don't bark, I told you it annoys people') and even inanimate objects such as vehicles ('stupid car won't start'), household objects ('where are my keys hiding'), fluffy toys ('poor tattered teddy') and computer-generated images (Heider & Simmel, 1944). We are particularly prone to perceive agency in things that are self-propelled and of humanoid form, and can do so as early as 9 months of age (Gergely, Nádasdy, Csibra, & Bíró, 1995). Presumably this approach has an evolutionary advantage in ensuring that we do not miss a mind if there is one (Lieberman, 2013).

Most of this social mindreading is performed without conscious awareness and does not draw attention, except in cases of severe impairment such as autism (Baron-Cohen, 1997). This, largely implicit nature as well as the core and natural character of the mentalizing processes are probably the reason it took us so long to understand how pervasive and essential it is for any kind of social navigation, and to articulate its main features in a coherent way. For example, the current understanding of the theory of mind is only 45 years old (Dennett, 1971).

The following section introduces the processes, neurobiology and development of mentalizing.

Understanding behavior in terms of mental states

Our brains constantly construe our reality by matching the (bottom-up) sensory inputs with the (top-down) expectations (Friston & Frith, 2014; Friston et al., 2013; Moutoussis, Fearon, El-Deredy, Dolan, & Friston, 2014; Moutoussis, Trujillo-Barreto, El-Deredy, Dolan, & Friston, 2014). The resulting sensations, perceptions, thoughts, feelings, impulses, ideas, dreams, wants, concerns and so on, are mental states. They belong to the mind and have a representational nature (e.g., a perception that there may be a snake in the grass is an inferred model of the real 'snake' in the hidden real world out there).

The relationship between mental states and action is causal. Mental states cause behavior. The physical action with which we interact with the world is a direct result of the intangible contents of our mind. We may be explicitly aware neither of mental states nor of such causal influence, but the motivators are present and determine behavior nonetheless (Bargh, 2013; Bargh & Chartrand, 1999; Bargh & Morsella, 2008; Decety, Cacioppo, Morsella, & Bargh, 2011; Nisbett & Wilson, 1977).

Contents of the mind of another are not directly observable. This observational opacity means mentalizing others is restricted to creating mental models of what their minds must be like, given what we *can* observe – their behavior – through social cognitive inference (mentalizing). For example, a teenager's tiptoeing into his home in early morning can be understood as being intended to avoid detection and motivated by fear of consequences of breaking the curfew, although those mental states are opaque to the observer.

Representing the contents of one's own mind recruits the same meta-representational capacity that is required for representing the contents of the mind of another. Self-awareness and awareness of mental states of others are closely linked in terms of the recruited brain areas (Frith & Frith, 2003). This is somewhat counterintuitive as we seem to believe that we have unmediated direct access to our thoughts and feelings (Cartesian 'first person doctrine'), the idea that has historically dominated the research in social sciences and has only recently begun to shift, after cognitive neuroscience and modern philosophy of the theory of mind built a robust case that internal modelling of our own and others' mental states and action may not be conscious at all, or only partly so (Fonagy et al., 2004, p. 205).

The neurobiology of mentalizing

The neural networks responsible for mentalizing are heavily involved during mental operations that involve *people*, but not during general reasoning tasks, unless these include content about other people (Van Overwalle, 2011). The mere physical or imagined presence of an agent with a mind seems to trigger these mentalizing networks, which shape decision making and behavior (e.g., Kovács, Téglás, & Endress, 2010).

The analysis of the neural networks that involve mentalizing shows that it is a four-dimensional capacity (Fonagy & Luyten, 2009). It is operating on continua between four sets of extremes (polarities): implicit and explicit, cognitive and affective, self and other, and internal- and external-focused (Bateman & Fonagy, 2012). Each dimension is based on relatively different neural systems (for an outline of the features of the four polarities together with the activating brain networks refer to **Table 1**).

Table 1. Four polarities of mentalizing (based on Luyten & Fonagy, 2015)

Polarity	Features	Neural circuits
Automatic / Implicit	Unconscious (outside of awareness) Parallel and fast processing Reflexive (rather than reflective) No effort or focus required No intention required Subject to implicit bias and distortion, particularly in complex interpersonal interactions	Amygdala Basal ganglia Ventromedial prefrontal cortex (VMPFC) Lateral temporal cortex (LTC) Dorsal anterior cingulate cortex (dACC)
Controlled / Explicit	Conscious (explicit) Mediated by language Sequential and slow processing Reflective (rather than reflexive) Effortful Intentional Requires focus	Lateral prefrontal cortex (LPFC) Medial prefrontal cortex (MPFC) Lateral parietal cortex (LPAC) Medial parietal cortex (MPAC) Medial temporal lobe (MTL) Rostral anterior cingulate cortex (rACC)
Internal-focused	Understanding one's own mind and that of others through a direct focus on the mental interiors (e.g., thoughts, feelings) of self and others	Medial frontoparietal network (more controlled)
External-focused	Understanding one's own mind and that of others based on external features (such as facial expressions, posture, and prosody)	Lateral frontotemporoparietal network (more automatic)
Self–focused and other-focused	Shared networks underpin the capacity to mentalize about the self and others	Shared representation system (more automatic) versus mental state attribution system (more controlled)
Cognitive–Affective	Mentalizing may focus on more cognitive features (more controlled), such as belief-desire reasoning and perspective-taking, versus more affective features (more automatic), including affective empathy and mentalized affectivity (the feeling and thinking-about-the-feeling).	Cognitive: several areas in prefrontal cortex Affective: VMPFC

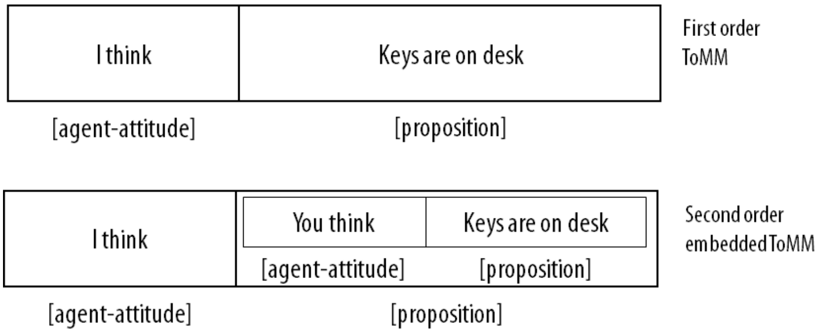
First, mentalizing can be *implicit and explicit*. The process of mindreading is not necessarily conscious. While we are capable of and often do engage in explicit mentalizing, most mentalizing cognition takes place automatically and outside of awareness. The distinction between explicit (controlled) and implicit (automatic) social cognition is consistent with the dual process theory dominant in social and cognitive psychology (Chaiken & Trope, 1999; Evans & Stanovich, 2013). The key proposition is that mentalizing operates through two distinct parallel types of processes: (1) the automatic processes, which are implicit, effortless, associative and uninterruptible, and are our default cognitions unless intervened upon by (2) explicit processes, which are controlled, slow, effortful and subject to interruption (Evans & Stanovich, 2013; Kahneman, 2011). The implicit-explicit mentalizing has specific neurocognitive correlates, the reflexive (automatic, System 1, Type 1) network, including the amygdala, basal ganglia, VMPFC, LTC, dACC, and the more frontally located reflective (controlled, System 2, Type 2) network that includes the LPFC, MPFC, LPAC, MPAC, MTL, rACC (Lieberman, 2007).

Mentalizing that governs ordinary day-to-day interaction is usually automatic and relies on a set of assumptions about self and others. This conserves energy by reserving the expensive reflective processes for situations where such explicit mentalizing is necessary. This is partly achieved by the general reluctance to engage explicit cognition whenever possible (the 'cognitive miser' phenomenon; Kahneman, 2011; Toplak, West, & Stanovich, 2014) and partly by automatization of cognitive skills (Bargh & Chartrand, 1999): with development, mindreading that initially requires controlled reflection becomes automatic and slips out of conscious awareness (Satpute & Lieberman, 2006). In other words, in a non-eventful, non-surprising ‘benign’ environment (Kahneman & Klein, 2009), there is no need to mentalize explicitly. In fact, excessive mentalizing may be out of place and a sign of at-risk status for personality disorder (Fonagy & Luyten, 2012; Sharp et al., 2011).

A mark of good mentalizing is the capacity to activate explicit reflection when the situation requires it (J. Allen & Fonagy, 2008). Imagine you are having a relaxing morning coffee with a usually reserved friend, when he unexpectedly breaks down in tears. The difference between the effortless implicit mental activity before the breakdown and the directed, focused, effortful thinking about the friend’s mental state is due to the involvement of explicit mentalizing. The adaptive flexibility serves to subject the automatically generated mental content to conscious scrutiny and iterative enrichment of the controlled inference (Van Overwalle & Evandekerckhove, 2013). An inability to do so results in the overreliance on inflexible, primitive and often wrong implicit assumptions that are likely inadequate given the situational demands (e.g., by failing to create a probable model of the mental states of others).

Second, *cognitive and affective mentalizing* involves two relatively independent neural processing systems. The first of these, the Theory of Mind Mechanism (ToMM), was proposed by Baron-Cohen. These specific sets of neural processes, commonly associated with the prefrontal cortex as a whole, are in charge of belief-desire reasoning. They process mental content in terms of ‘M-representations’: succinct unitary formulae linking agents, attitudes and propositions along the lines of ‘Jack (agent) thinks (attitude) keys are on the desk (proposition)’ (Baron-Cohen, 1994; Baron-Cohen, Golan, Chakrabarti, & Belmonte, 2008). An M-representation can be embedded in another M-representation, thus creating higher-order theory of mind cognitions (see **Figure 1**).

Figure 1. M representations in first and second order of ToMM cognitive mentalizing

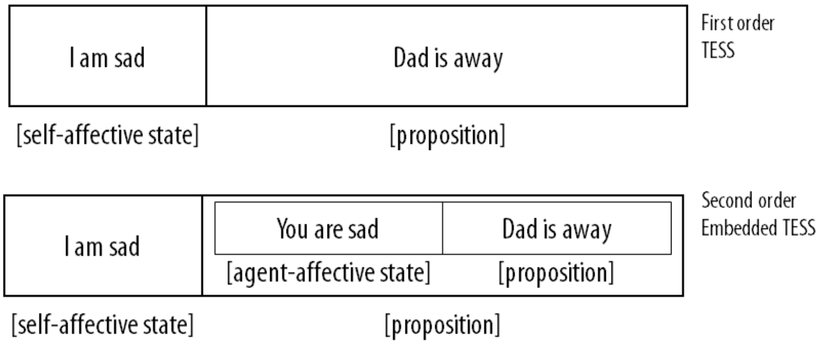


Note. M-representations can be embedded in higher order M-representations thus facilitating reflective theory of mind thinking (Chakrabarti & Baron-Cohen, 2006).

For example, ‘Jack believes Mary thinks keys are on the desk’ is an example of second-order thinking. ‘Mary thinks Jack believes Mary thinks keys are on the desk’ an example of third-order thinking, and so on. Theoretically, the only constraint on the limit of such reflection is the computational power of the agent. A mature human can aspire up to 4th or 5th level of such intentionality.

While ToMM deals with *cognitive* content, affective experience is processed by a second, separate neural empathizing system (TESS). It represents *affective* states as E-representations which link the self with affective states. Similar to ToMM, E-representations can be embedded in other higher-order E-representations thus allowing empathic reactions to another’s emotional states: ‘I-am happy (self-affective state) that you-are happy (agent-affective state proposition). It is a basic emotional processing system, associated with the inferior prefrontal gyrus and is available from as early as 3 months of age (see **Figure 2**).

Figure 2. E representations in first and second order TESS affective mentalizing



Note. E-representations can be embedded in higher order E-representations to represent empathic reactions to another's affective states (Chakrabarti & Baron-Cohen, 2006).

TESS differs from the perspective-taking ToMM not so much in terms of content as in the nature of processes: ToMM capacity to represent agent attitudes (e.g., I *think, believe, pretend* that [proposition]) allows provisional reasoning, doubt and degrees of probability. TESS, on the other hand, is far more bivalent (e.g., I *am* [affect-state] that [proposition]). You either are in the state of affect or not (you either *are* sad or *not*).

Third, mentalizing can be *focused on self or on other*. Neuroimaging shows that two distinct neural networks are shared in understanding self and other. The first, phylogenetically older, system mediates a direct affective experience of another's state of mind and has been traditionally associated with the term 'empathy' (Lieberman, 2007; Rameson & Lieberman, 2009). It is located in the frontoparietal mirror-neuron system (Rizzolatti & Craighero, 2004), which is responsible for understanding embodied self and others through motor-simulation mechanisms (Fonagy & Luyten, 2009; Keysers & Gazzola, 2006). This system gets activated both when body-related events happen to ourselves as well as when we observe them happen to another (e.g., observing a hairy black spider crawl up your colleague's arm). These processes facilitate an immediate affective understanding of what it is like to be the other and may be the key evolutionary mechanism underpinning social empathy (Lieberman, 2007). The second, phylogenetically younger cortical midline system processes information in more abstract and symbolic ways (Fonagy & Luyten, 2009) and is instrumental in distinguishing observed experience of others from one's own through inhibition of imitative action.

The last polarity relates to the focus of mentalizing that can be *internal or external*. The processes facilitating inference about mental states can be focused on exteriors (e.g., facial expressions, body posture, tone of voice and physical action) or on mental interiors (thoughts and feelings). Correct internally-focused mentalizing of self involves the capacity to label emotional states, which is underpinned by a system of second order representations of mental states, achieved through interaction with marked-mirroring mature minds during development (Fonagy et al., 2004; Fonagy & Target, 1996, 1997). As mental interiors in another can be thought of as second-order representations of the exteriors, the internal-external polarity applies to other-focused mentalizing as much as to the self-focused one.

The original developers of mentalizing theory (J. Allen, 2006, 2013; J. Allen & Fonagy, 2008; Bateman & Fonagy, 2012; Luyten & Fonagy, 2015) suggested that mental health problems arise from counterproductive dominance of any pole within the four polarities (i.e., automatic versus deliberate processing, over-heightened attention to feelings versus cognitions, mis-attention to self versus others, and mis-attention to internal experiences versus external actions). Further, the connection with supportive others in times of need is underpinned by an ability to process social information and move fluidly along each of the mentalizing dimensions. Finally, mentalizing might be restricted and produce unwarranted certainty (psychic equivalence), unsubstantiated presumptions about feelings and beliefs (hypermentalizing) and insistence on actions to demonstrate subjective experience and intention (teleological stance).

Development of mentalizing

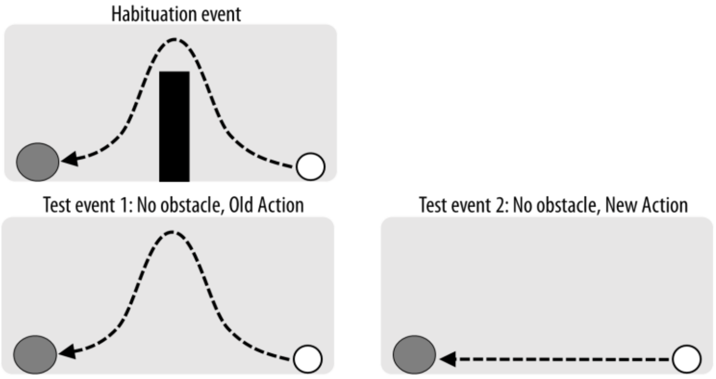
The capacity to mentalize is not a genetic given. Development of the capacity to propositionally reason about mental states as causes of behavior is the product of a complex developmental process and is closely linked to the emergence of subjectivity (Fonagy, Gergely, & Target, 2007).

Evidence suggests that the development of mentalizing takes place in roughly three stages. Infants before 9 months of age seem to be prewired for a very limited processing of physical and social agency that allows them to understand that they are in an action-effect relationship with the environment and to engage in species-specific interactions with their caregivers (Stern, 2007).

At about 8-9 months, the infants become capable of *teleological understanding* of agency. Teleological reasoning involves understanding action in terms of the outcomes it pursues (e.g., observing a person throwing a rock at a window is understood as aimed at breaking the glass; he is throwing the rock *in order to* break the window). The ability to teleologically understand behavior also presupposes a theory of rational action that includes the ability to process the causal relationships between action, outcome, and reality constraints.

In a classical set of experiments, Csibra, Gergely and collaborators (1999; see also Gergely & Csibra, 2003) measured the surprise of 9-month old babies to changing behaviors of animated circles. Initially, they repeatedly presented babies with a short video clip where a small circle approached a bigger circle by jumping over a rectangular obstacle. After the infants got habituated to that stimulus, the testing was conducted by showing a different recording. The barrier was now removed and the path between the two circles was clear. The small circle then behaved in one of two ways: it either went directly to the big circle in a straight line, or it 'jumped' over where the barrier used to be, as if it were still there, before approaching the big circle (**Figure 3**). Infants were significantly more surprised by the latter behavior although the small circle's path itself (involving jumping) was identical to the one taken in the first stage of the experiment.

Figure 3. Teleological reasoning in infancy


Habituation event
Test event 1: No obstacle, Old Action
Test event 2: No obstacle, New Action

Note. In the absence of obstacle, the 'old action' event generates significantly more surprise to the experimental group than the 'new action' although the movement of the small circle is different from the habituation event (Gergely & Csibra, 2003).

The behavior of the small circle in the experiment can be explained teleologically only if the impact of the presence of the impenetrable obstacle is understood. Jumping only makes sense if there is a barrier in your way. If there is none, the small circle can (and should) approach the big circle in a straight line. This explains the infants' surprise at the irrational jumping action in the absence of the barrier, as well as the relative lack of surprise at the different, but more rational direct-line path of the small circle.

An action is teleologically explained if the outcome justifies it as rational given the reality constraints. This kind of reasoning ignores considerations of anything prior and intangible – including *prior mental states* such as desires and beliefs about reality constraints – that initiate the action before it takes place (Csibra & Gergely, 1998; Csibra et al., 1999). Although teleological understanding may be an effective way of interpreting human behavior in instances where beliefs correspond to reality, it fails when they do not. This can happen in instances of information asymmetry, pretense or deception. In addition, as teleology only considers the action-outcome relationship, the only way of changing anticipated outcome is by physically interfering in this relationship as it takes place (e.g., by physically stopping the person from throwing the rock, or intercepting the rock's flight before it reaches the window).

To overcome these limits, children develop the understanding of mental causation in two final qualitative leaps. As the first step, at some time during the second year, infants become able to think in terms of intentions (that is, that people are motivated by mental states) but not yet in terms of beliefs (that is, that people are motivated by beliefs that can be false). A child understands that action may be motivated by prior intentions that people can, but do not have to, act upon (Fonagy et al., 2004), but still fail Sally-Anne tests of false beliefs (Baron-Cohen, 1985).

At around the age of four, children begin to appreciate the nature of beliefs as mental states and become able to comprehend that beliefs can be false in both themselves and others. This is reflected in the capacity to imagine intangible mental states as motivators of behavior, and to propositionally reason about them (that is, children develop the 'Theory of Mind'; Dennett, 1978; Perner & Wimmer, 1985; Premack & Woodruff, 1978).

A smooth transition through these developmental stages depends on interactions between the infant and mature, attuned and sufficiently reflective minds. Repeated interactions involving ‘marked mirroring’ of the infant’s mental states facilitate the development of a symbolic representational system for thoughts and feelings (Fonagy et al., 2004). Caregivers epistemically engineer communications with the infants to facilitate such learning (Heyes & Frith, 2014; Meins et al., 2003).

Mature mentalizing involves the capacity for balanced cognitive and affective appreciation of mental states, the reflective ability to distinguish inner from outer reality and internal emotional processes from interpersonal events, as well as the capability to use mentalizing for relational and affect regulative purposes (Bateman & Fonagy, 2012). Serious disturbances of mentalizing, on the other hand, accompany severe psychopathology such as borderline personality disorder (Fonagy, Luyten, & Strathearn, 2011) and psychoses (Lysaker et al., 2013), and mentalizing may be at the core of most psychotherapeutic interventions (Fonagy & Allison, 2014; Fonagy, Luyten, & Allison, 2015; Fonagy, Luyten, Allison, & Campbell, 2017).

The description of the ascent of mentalizing in the hands of a sufficiently attuned and benight caretaker mirrors Winnicott’s good-enough ‘holding’ (Winnicott, 1960), a notion that for an infant’s healthy development the mother needs to be capable of metaphorically ‘holding the child’, making his or her spontaneous actions recognized and attended to. By doing that, the mother simultaneously gives meaning to the infant’s behavior and signals that such behavior, which is an expression of the infant’s impulses, is acceptable. This is particularly critical in the earliest developmental period.

In the Kleinian psychoanalytic tradition, the concept of the development of mentalizing is similar to Bion’s notion of the birth of the thinking apparatus (Bion, 1988). Bion says that ‘thinking is a development forced on the psyche by the pressure of thoughts and not the other way round’, and that the psychopathology may result not only from ‘a breakdown in the development of thoughts,’ but also from ‘the breakdown in the development of the apparatus for “thinking” or dealing with thoughts’ (p. 179). Bion’s intention was to describe the formation of the very earliest of concepts and processes in the first year of life, but a similar sequence (of content driving the development of a cognitive process) likely applies to the development of mentalizing as the result of the pressure of thoughts and feelings that need to be mentalized to make sense of the behavior of oneself and others (see also Target & Fonagy, 1996, p. 469). Further comparisons between the theory about mentalizing and other psychoanalytic thinkers, including Matte-Blanco (Matte-Blanco, 2003; Matte-Blanco & Rayner, 2018), Segal (Segal, 1957), Steiner (Steiner, 1992), and Britton (Britton, 1995) have been made in a comprehensive review by Target and Fonagy (1996).

Phenomenology and neurobiology of poor mentalizing

In population without marked mental health problems dramatic losses of mentalizing are not uncommon, but tend to be temporary affairs, highly dependent on interpersonal and situational factors. More importantly, impaired mentalizing shows itself as a reemergence of developmentally early, pre-representational modes of social cognition.

They come in two basic varieties: first, as a virtual absence of mentalizing, resulting in non-intentional understanding of self and other, and second, as inaccurate mentalizing, where mental causation is represented, but faultily so. Neurobiologically, impaired mentalizing entails an imbalanced activity of the brain centers responsible for the self-other, explicit-implicit, affective-cognitive and inner-outer focused mentalizing discussed above (Fonagy & Luyten, 2009, 2012).

Teleological reasoning

From the age of around 9 months, and up until year three or so, children reason teleologically, until such thinking is superseded by considering intentions (see p. 28, above). In adult life, teleological reasoning is maladaptive as it leaves out of focus mental causation: actions of agents are explained and predicted in terms of goals they lead to rather than in terms of the mental states that generate them. The cognitive focus is on external determinants of mental states and scotomisation of putative internal markers.

In practice, teleological reasoning involves focus on physical properties of actions; sentences like ‘we will see what they do’, ‘I will believe it when I see it’ and ‘actions speak louder than words’ are indicative of teleology. It involves a sort of mind-blindness to intangible mental states as generators of behavior. Only physical action counts. The computation-intensive effort to identify internal states that could account for or in many instances at least complement the interpretation of behavior is given up in the interest of the rapid processing of immediately available data.

Teleological reasoning exerts relatively low demands on cognitive resources and is thus an efficient model for understanding and predicting physical events in simple scenarios where mental states are irrelevant (e.g., helping an old lady with her heavy bag), but is lacking in more complex situations. Finally, it is most clearly marked by excessive confidence in attributions, which contradicts what we know about mental states being naturally opaque.

Teleological reasoning may be the result of an acute breakdown of mentalizing, or of the parties being oblivious or indifferent to the need to mentalize, or being intentionally obtuse. Mental states are assumed to be isomorphic with what can be accessible through behavioral observation (e.g., the other side is suing to win in court, making an offer to get us to accept a bad deal). The behavior that accompanies concrete understanding is often angry, over-reactive, blaming, and prescriptive.

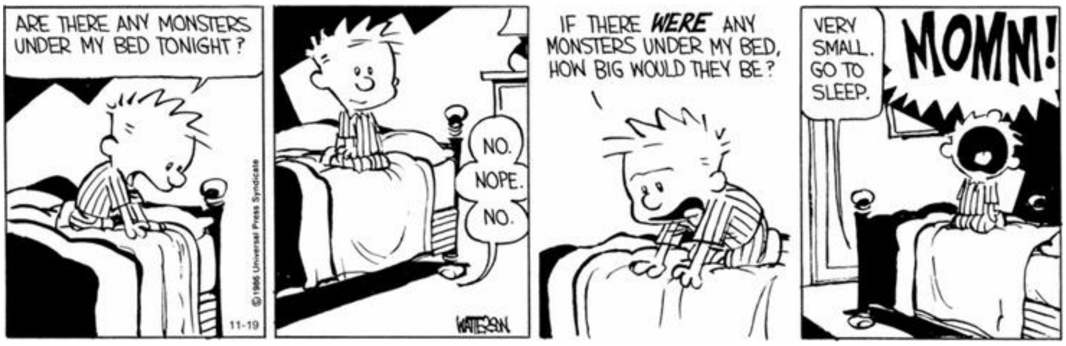
Pseudomentalizing: psychic equivalence and pretend modes

Unlike teleological reasoning, pseudomentalizing modes of social cognition treat mental states as determinants of behavior but fail to appreciate these as representations meaningfully linked with reality. This is common in children between two and five; their social cognition takes place in one of two pseudomentalistic modes: the 'psychic equivalence' mode, where ideas are felt to be reality, rather than the reality's representational correlates, and the 'pretend' mode, where all mental states are unhinged from reality, and felt as 'just thoughts' without any relevant links to reality (Fonagy & Target, 1996, 2000, 2007; Target & Fonagy, 1996).

In these pseudomentalistic modes, the relationship between the mental states and reality is heavily polarized: ideas either *are* reality, or have nothing to do with it. This is consistent with evidence that the processing of affect antedates the processing of cognitions. The processes of affective mentalizing dominate over cognitive ones as the probabilistic ToMM fails to moderate the bivalent affective nature of TESS. Pseudomentalizing involves a relative dominance of implicit over explicit and externally-focused over internally-focused mentalizing. On the self-other polarity, pseudomentalizing is associated with a weaker influence of the belief-desire (MPFC/ACC) inhibitory system on the imitative mirror neuron system, which causes a certain degree of fusion between perceptions of self and other (Fonagy & Luyten, 2009).

The *psychic equivalence* mode involves a mind-world isomorphism: mental states are felt to be what physical reality is. In the agent's mind, they are not representations, but *replicas* of the world. This implies that psychic equivalence construal of reality is vulnerable to projection. For example, a small child's fear of a monster under the bed means to that child that the monster really is under the bed, and not that he is afraid of something he thinks is under the bed. The child may otherwise – in bright daylight and hanging out with his father – very well know that monsters do not exist. In the bedroom at night, however, he is not able to mentally disconnect the affect-laden fantasy from what he feels the reality is. He is unable to consider that his fear is the result of his belief that the monster may be under the bed; the monster *is* under the bed (**Figure 4**).

Figure 4. A small child experiences elaborated fantasy as reality


ARE THERE ANY MONSTERS UNDER MY BED TONIGHT?
NO.
NOPE.
NO.
IF THERE *WERE* ANY MONSTERS UNDER MY BED, HOW BIG WOULD THEY BE?
VERY SMALL. GO TO SLEEP.
MOMM!
© 1986 Universal Press Syndicate
11-19
WATERSON

Note. Part of the humor in the comic derives from the fact that Calvin's non-mentalistic, psychic equivalence interaction with the imagined monster contains instances of sophisticated (third and fourth-order) mentalizing: the monster attempts to deceive and Calvin understands that the monster is trying to deceive. Image from Calvin & Hobbes by Bill Waterson. Used with kind permission of Universal Press Syndicate.

While this is a normal developmental stage and is naturally superseded by more mature forms of mentalizing, temporary relapses are not uncommon. Dreaming and episodes of Post-Traumatic Stress Disorder (PTSD) are examples of extreme psychic equivalence: their projective reality is understood only after they are mentalized upon awakening or when the PTSD episode passes (J. Allen, 2013, p. 152). When inferring mental states of others, psychic equivalence causes conflating one’s own mental states with the mental states of others (e.g., fear of being taken advantage of must correspond to their intent that they really are out to get you). There is no intersubjectivity in psychic equivalence; while an individual may feel that such mental states belong to another, they are really his or her own projective derivations.

The corollary of ‘mind equals reality’ is that reality must also equal the mind: everything out there is felt to be contained in the mind and, more importantly, everybody knows reality in the same way, a phenomenon known as the ‘curse of knowledge’ (Camerer, Loewenstein, & Weber, 1989). This underappreciation of subjective differences in representations of reality is a logical derivation of the non-representational nature of psychic equivalence. In the agent's mind, mental states are not subjective representations of reality, but its identical replicas, and equally perceived by everybody: everybody knows the same things. Psychic equivalence is thus the common root of a host of reasoning errors that persist in adulthood such as the hindsight bias, spotlight effect, adult egocentrism, reality bias, epistemic egocentrism, illusion of transparency, spotlight effect and the false consensus effect (Birch & Bloom, 2007).

Psychic equivalence functioning shows itself in the intolerance of alternative perspectives and in an unwarranted conviction about the ascribed mental states. Clichés such as ‘that’s just how things are’, ‘that’s reality’, ‘people take advantage of you’ are not uncommon and are used to defend rigid internally generated ideas that are projected outwards.

Pretend mode of mentalizing generates mental states that are too representational in the sense that they are disconnected from cognitive and affective reality of self and other. Instead of feeling too real, they are *un*real. The mode involves preoccupation with hypotheticals about one’s own mind or the mind of another; however, these pseudomentalistic narratives form no bridge between inner and outer reality and are disassociated so much they are meaningless. While people in psychic equivalence conflate, in pretend mode they fabricate. The output is a nonconsequential prattle, called ‘psychobabble’ in therapy, ‘bullshit’ colloquially, and hypermentalizing, extramentalizing or excessive theory of mind in the social cognition literature (Bo, Sharp, Fonagy, & Kongerslev, 2017; Dziobek et al., 2006; Luyten & Fonagy, 2015): a social-cognitive inference that generates representations about people’s mental states that ‘go so far beyond observable data that others may struggle to see how they are justified’ (Sharp et al., 2013). Such overattribution of mental states leads to obfuscated mental causation and is an endless source of interpersonal problems; people who hypermentalize tend to lack stable interpersonal relationships.

The pre-mentalizing modes are not mutually exclusive. In situations of poor mentalizing, the psychic equivalence often overlaps with the pretend mode. The more removed from reality the results of social cognitive inference are, the more convoluted and detailed the pseudomentalizing tends to be. For example, Bateman and Fonagy report that patients with the most severely impaired social cognition seem to use the most distorted mentalizing to maintain denial or to manipulate or control others (Bateman & Fonagy, 2012).

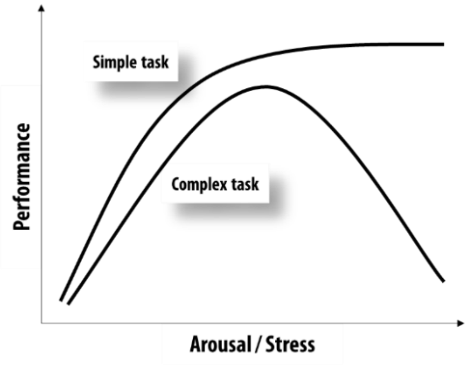
Situational inhibitors of mentalizing: arousal and cognitive load

The quality of mentalizing depends on both dispositional and situational factors. Serious personality disorders involve various impairments in mentalizing; however, there is marked variance in reflective function even within the nonclinical population. Individual mentalizing profiles reflect differences in functioning with respect to each mentalizing polarity and studies show these differences may have roots in epigenetic influences and developmental vicissitudes (Luyten, Fonagy, Lowyck, & Vermote, 2012). Mentalizing is also highly context dependent. Dispositional factors – the individual's mentalizing profile – can be thought of as a set of strengths and vulnerabilities of effective mentalizing under varying situational pressures.

Arousal

The first potential inhibitor of explicit mentalistic inference is excessive arousal (stress). Arousal involves activation of the central and the peripheral nervous systems, accompanied by changes in the heart rate and blood pressure, skin conductivity and muscle tonus, as well as less visible but fairly complex alteration in brain functioning that has adverse impact on the quality of cognition. At the beginning of last century, Yerkes and Dodson (1908) proposed a model of the relationship between task performance and arousal (**Figure 5**), stating that the impact of arousal on performance depends on how cognitively demanding the task is. If the task is easy, arousal improves performance. If the task is difficult, the quality of performance is an inverted U-function of arousal; initial increases of arousal stimulate performance until it reaches its peak, after which any further intensification of arousal starts to impair it in a progressive fashion.

Figure 5. Yerkes-Dodson arousal-performance model

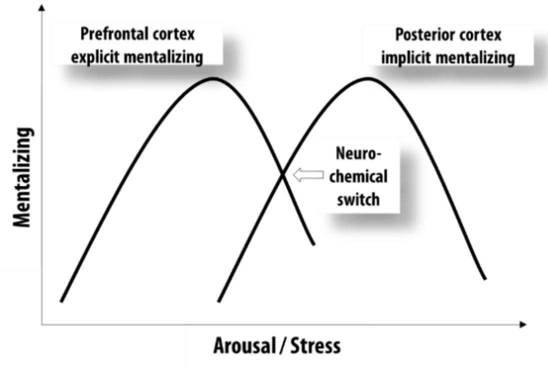


Modern neurological accounts demonstrate that as arousal levels rise, the increasingly weaker cognitive performance is due to the reduction of activity in the areas of the prefrontal cortex that are critical for higher order cognition, which includes mentalizing. At the same time, activity is increased in the posterior cortex, amygdala and hippocampus, the areas that facilitate automatic responses.

With sufficient arousal, prefrontal cortex areas get knocked off-line, while implicit fight-or-flight and vigilance functions take over (Mayes, 2000, 2006). This has clear evolutionary logic: in the face of imminent danger, basic rapid automatic responses carry superior survival value than slow, effortful cognition.

The same is the case with social cognition (**Figure 6**). If arousal exceeds a certain threshold, we lose the capacity to engage controlled mentalizing, and the automatic processes run rampant.

Figure 6. Fonagy-Luyten model of biobehavioral switch in mentalizing



The arousal systems activate depending on an individual's perception of the situation they are in. If this construal involves a danger to something important – even if only imagined – the stress levels rise and put the explicit cognitive capacity is at risk. Everyday life is full of instances where explicit cognition is inhibited in stressful situations: students 'choking' on exams, job interviews going badly, being unable to deliver a witty rebuke to a rude stranger on the tube (although oh so many come to mind after the fact), and finally, an inhibited ability to 'think' in important negotiation, the topic of this thesis.

This starts a vicious circle. One of the effects of this arousal, which itself might be a consequence of erroneous construal, is that the reduced capacity to mentalize explicitly exacerbates the severity of the misconstrual, by way of either projection (psychic equivalence mode) or hypermentalizing (pretend mode). Psychic equivalence can color the 'other' in projections that have little to do with them and are instead indicative of the person who is inferring (e.g., own hostile intentions are perceived to be a part of the other), and hypermentalizing results in a slew of possible mental states, each unmoored from reality. Such poor mentalizing further increases arousal, which impairs mentalizing even more, and so on and on.

Cognitive load

Cognitive load has been shown to impact explicit mentalizing (Spunt & Lieberman, 2013) as well as the subjective experience of empathy (Morelli & Lieberman, 2013). The implication is that the situations involving significant cognitive demands present a risk that controlled social cognition, necessary in complex environments (Kahneman & Klein, 2009), may not be available, which can result in pseudomentalizing or teleological interpretation of action. Basically, if the demands on the cognitive apparatus are too extensive, the sophisticated social cognition gives way to automatic processes, vulnerable to error.

Just like arousal, excessive cognitive load can trigger a vicious circle. As cognitive load inhibits mentalizing, failed mentalizing generates increases in cognitive load. Namely, automatic mentalizing – i.e., social cognitive inference that is exclusively implicit and lacks explicit checks – entails either psychic equivalence, which triggers attribution errors, or hypermentalizing (pretend) that generates perceptions that increasingly complicated and unmoored from reality.

Mentalizing as an evolutionary adaptation

There is a debate in natural sciences about why the human brain grew five times larger than expected in a placental mammalian of human size (Martin, 1981) in spite of the associated metabolic cost (Aiello & Wheeler, 1995). For the last 40 years, the social intelligence hypothesis has dominated these explanations. The key idea is that human cognition evolved to deal with social challenges involving intra-group and inter-group collaboration and competition: the individuals who were able to cooperate, manipulate and avoid manipulation, and form coalitions, were selected over their less competent peers (Alexander, Mellars, & Stringer, 1989; Byrne & Whiten, 1989; Dunbar, 1992, 1998, 2011; Dunbar & Shultz, 2007; Waal, 1989).

The social brain hypothesis

The crowning achievement of human evolution, the neocortex, may have developed for the very purpose of mentalizing. Across the species endowed with forebrains, the size of the neocortex (but not other parts of the brain) correlates with the size of the social group and the mentalizing capacity of its members: the larger the group, the larger the forebrain and the more robust the mentalizing ability (this is particularly clear in monkeys and apes). These relationships seem to be linear (Dunbar, 2014; Dunbar & Shultz, 2007).

The *social brain hypothesis* (Humphrey, 1976) suggests that the brain grew not in response to any challenges from the environment, but to the fact that our ancestors lived in progressively larger groups (Dunbar, 2007). The leaps in size of the neocortical tissue took place after our species had already mastered the 'hostile forces of nature' and our exceptional intelligence is hypothesized to have developed to tackle the challenges of cooperation and competition by our conspecifics (Alexander, 1989; Flinn, Geary, & Ward, 2005; Humphrey, 1976; Moll & Tomasello, 2007).

While all species face such competition, in humans the decisive evolutionary advantage turned out to be a combination of competition and collaboration: bonded social groups were the most effective organizational structures for fostering the reproductive success of their members – both in terms of protecting themselves from other, competing groups (e.g., versus homo *Neanderthalensis*), and in terms of collaboratively providing for their own members (e.g., cooperative childcare, cooperative foraging, and cooperative communication and teaching). A hypothesized 'evolutionary arms race' between societies was the evolutionary pressure spurring development of brain growth (Flinn et al., 2005).

For a group to exist, its members must collaborate. That involves constant processing of information about the ever-changing state of the group for the purposes of coordination, compromise and conflict management. Competition with out-groups requires communication and mental simulations of the competitors' likely moves and strategies. Both competition and collaboration demand imagining perspectives of others, which presupposes a capacity to symbolically represent their mental states. Mentalizing thus became a major reproductive advantage (Fonagy, 2006).

While some contemporary studies highlight the need to account for brain growth by combining the social hypothesis with the ecological (DeCasien, Williams, & Higham, 2017; Rosati, 2017) and cultural explanations (Moll & Tomasello, 2007; Street, Navarrete, Reader, & Laland, 2017; van Schaik & Burkart, 2011), there is little doubt that the demands for mentalistic inference played a significant role in brain growth.

Tomasello's Vygotskian cultural evolutionary account

Created by competition, fine-tuned by collaboration

In 2007 Moll and Tomasello offered evidence supporting what they call the 'Vygotskian intelligence' hypothesis (2007). The key of such thesis is that the social intelligence hypothesis – that primate cognition was primarily driven by social competition, elaborated further by accounts of primate politics and Machiavellism (Byrne & Whiten, 1989; Waal, 1989) – needs to be fine-tuned along the lines of the theoretical proposals of Vygotsky (1978). Vygotsky, unlike his peers who emphasized the 'arms race' argument, exclusively focused on the cooperative aspects of human interactions as drivers of uniquely human forms (and sheer capacities) of cognition.

Moll and Tomasello suggest that the accounts stressing collaboration and competition complement each other. Humphrey's competition driver thesis (Humphrey, 1976) applies to primates and Vygotsky to humans: nonhuman primates' cognitive spurs were driven by social competition, and the unique aspects of human cognition were driven, and are constituted, by social cooperation (Moll & Tomasello, 2007, p. 639).

What is particularly important for the purposes of the theory and studies reported in this thesis, particularly in Part II (pp. 181-263), is Tomasello's account of the evolution of intentionality (Moll & Tomasello, 2007; Tomasello, 2006, 2008, 2012, 2014; Tomasello, Carpenter, Call, Behne, & Moll, 2005; Warneken & Tomasello, 2006). Below we outline it briefly.

Individual intentionality

The key aspect of great apes' sociality is the motivation to live in a social group (Tomasello, 2014, p. 135). Living inside a group entails some level of competition – mainly for food and mates – and shows itself in dominance and affiliation. To deal with this aspect of social life, great apes developed sophisticated social cognition, including the ability to represent intentions and mental states, understand goals and desires, and manipulate intentional states in others. However, their social cognition is limited to competition.

Great apes' working together does not quite 'collaboration' make. First of all, they seldomly pursue any kind of joint action. When they do, while their actions may be coordinated, there is a qualitative difference between such co-action and collaboration proper: every monkey is in it for himself. Tomasello offers an interpretation of chimpanzees' 'co-action' in hunting the colobus monkey (2014, p. 35) as follows:

> *'What happens prototypically is that a small party of male chimpanzees spies a red colobus monkey somewhat separated from its group, which they then proceed to surround and capture. Normally, one individual begins the chase, and others scramble to the monkey's possible escape routes, including the ground. One individual actually captures the monkey, and he ends up getting the most and best meat. But because he cannot dominate the carcass on his own, all participants (and many bystanders) usually get at least some meat, depending on their dominance and the vigor with which they beg and harass the captor (Gilby, 2006). The social and cognitive processes involved in chimpanzee group hunting could potentially be complex, but they could also be fairly simple. The "rich" reading is a human-like reading, namely, that chimpanzees have the joint goal of capturing the monkey together and that they coordinate their individual roles in doing so (Boesch, 2005). But more likely, in our opinion, is a "leaner" interpretation (Tomasello et al., 2005). In this interpretation, each individual is attempting to capture the monkey on its own (since captors get the most meat), and they take into account the behavior, and perhaps intentions, of the other chimpanzees as these affect their chances of capture. Adding some complexity, individuals prefer that one of the other hunters capture the monkey (in which case they will get a small amount of meat through begging and harassing) to the possibility of the monkey escaping totally (in which case they get no meat). In this view, chimpanzees in a group hunt are engaged in a kind of co-action in which each individual is pursuing his own individual goal of capturing the monkey (what Tuomela, 2007 calls "group behavior in I-mode").'*

What is important here is that the great apes are not – and the pre-*Heidelbergensis* homo was not – able to represent joint goals and different perspectives for the purposes of collaboration; there can be no cooperative communication for coordinating action.

This is not to say that they are unable to perform rudimentary social inference in competitive settings. For example, when chimpanzees observe an experimenter approach many buckets (one of which containing food) and reach for one of them, they are able to infer that the food is in *that* bucket. This is quite a remarkable feat of social cognition. What is even more remarkable is that the chimps can do this only in competitive settings. A virtually identical experiment in a collaborative environment does not yield the same result. No matter how much the experimenter was helpfully pointing at the food containing bucket, chimps remained completely oblivious as to which cup contained the food (Bullinger, Wyman, Melis, & Tomasello, 2011; Hare & Tomasello, 2004).

This makes evolutionary sense. The computationally expensive nascent (rudimentary) social cognition was usefully employed only in social interactions of crucial importance, which were predominantly competitive: fighting for food and mates. In such contexts, it simply knowing what the competitors want and predicting simple competitive action is sufficient. There is no need to coordinate, or communicate with cooperative intentions, or take perspective of the other side for cooperative purposes.

Joint intentionality

At some point around 150 thousand years ago, early humans were forced – probably by changes in ecology – to start sourcing food by cooperation (Tomasello, 2014). What resulted was *interdependence*: individuals became dependent on each other for survival, which made the ability to coordinate an evolutionary advantage and an immediate interest of each individual.

Rousseau's stag hunt is a paradigm of social coordination (Bullinger et al., 2011) that early humans encountered in their hunting outings. In the pure game theoretic model, each of the two players must decide whether to hunt hare or stag. An individual will always catch a hare no matter what their counterpart chooses to hunt. However, stag can only be hunted down if both players decide to do so (i.e., if the other player decides to hunt rabbit, the stag hunter will be unsuccessful and go hungry). The assumption is that stag is a much larger prey and therefore food of both higher quality and quantity.

We can see the dilemma here: stag is obviously superior to hare but hunting it in the absence of a credible commitment or at least some coordination communication of the other side entails a risk. This is reflected in two Nash equilibria: the risk dominant ‘hare, hare’ strategy and the payoff dominant ‘stag, stag’. What exactly is needed for both parties to choose the optimal stag strategy? At the minimum, both players need to be able to (on their own) represent not only that it is better to catch a stag than a hare, but that their *counterparty* can represent that same thing too and, more importantly, that they can represent that the focal player thinks the same. In other words, to decide stag is the better choice I need to know that you think that, too, and that you think that I think that, too. This is mentalizing.

Understanding the dilemma and managing to end up in the payoff-dominant ‘stag, stag’ quadrant of the game requires a complex representational capacity, and the ability to recursively infer mental states (i.e., at least a third-order theory of mind). While early humans did not encounter a game theoretic Stag Hunt with all its formalities and assumptions, it is easy to imagine the challenges posed to human cognition by its real-life variants. At the very least, the hunters needed to be able to represent a joint task (to catch a stag) and individual perspectives in the context of such a joint task (e.g., I will wait here in ambush while you chase the prey, so it runs toward me to catch it). Tomasello (2014, 2019) calls this reasoning second-personal thinking that underpins joint intentionality: it includes cognitive representations that are perspectival and symbolic, and recursive social inference (i.e., I think that you think that I think).

Evolutionary pressures then selected for such joint intentionality, a cognitive mode that underpins joint attention, representation of common goals and formation of complex perspectival representations (e.g., simultaneously keeping in mind one’s own and the other’s perspectives on various issues).

Collective intentionality

Around 100 thousand years ago, modern humans faced new social challenges, partly due to increased group sizes, and partly due to the inter-group conflict (Tomasello, 2014). The competitive advantage was then not only the number of group members, but their ability to operate effectively as a collective. Because coordination within a large group is as difficult as coordination with strangers, an advantage was to have some sort of common ground, a mutual understanding every member could rely on. Tomasello suggests that the solution to this was the conventionalization of and conformity to cultural practices. This had an impact on communication and perspective-taking because a sort of group, 'objective' perspective became available (i.e., 'this is how this is done'). It was shared within a group, but not with other groups. This led to the development of complex languages, rational discourse and decision making (Tomasello, 2014, p. 138), and what is called collective intentionality: cultural conventions, norms, institutions and language. Human cultural and scientific achievements are all based on shared intentionality and the cognitive leaps it triggered.

Overall, Tomasello's evolutionary argument is that humans evolved by creating new forms of cooperation that entailed new forms of communication, which have led to new types of cognitive representation and inference. This took place in two leaps, the first one from individual to joint intentionality, and the second one from joint to collective intentionality (see **Table 2**).

Table 2. Individual, group and joint intentionality (based on Tomasello, 2014)

Individual intentionality	**Joint Intentionality**	**Collective intentionality**
Competition	Two-person collaboration	Large-scale, group level collaboration
Competitive mentalizing	Cooperative and competitive mentalizing	Group-level mentalizing

For the purposes of this thesis, the difference between individual and joint intentionality is particularly important. These two modes coexist in the mental repertoire of modern humans, reflecting a history where an individual's interests in interpersonal settings were satisfied in different ways at different times. Both modes are thus available for the individuals to operate under, and different circumstances seem to trigger different modes.

The default mode seems to be the cooperative, joint-intentionality, 'we' mode. Children are generally cooperative in mixed-motive games (that is, in games that involve cooperation and competition simultaneously). For example, 18-month old children happily help non-kin achieve their goals (Warneken & Tomasello, 2006) and 5 year-olds actively coordinate to overcome a prisoner's dilemma (Sánchez-Amaro, Duguid, Call, & Tomasello, 2019). Adults, on the other hand, default to cooperation but can revise this approach if they think about it. Rand, Greene, and Nowak (2012) demonstrated an implicit cognitive preference for cooperation (also see Tomasello, 2012), which was reversed if people had the time or were instructed to think about the mixed-motive situation carefully, suggesting that the automatic response is collaborative, but can be reversed by a controlled cognitive intervention if the individual deems it to be a superior strategy in a given situation (Chaiken & Trope, 1999; Evans & Stanovich, 2013; Kahneman, 2011). Also, priming affective responses triggers collaboration while priming reason triggers competition, suggesting that automatic outputs tend to be cooperative but can be revised by a deliberate cognitive intervention (Levine, Barasch, Rand, Berman, & Small, 2018).

Conclusion

This selective review of mentalizing and negotiation literature focused on the aspects that are critical for establishing the foundations for the theoretical model of mentalization-based bargaining that this dissertation is developing testing. Social cognitive inference is essential to making sense of people in any context, but is particularly relevant in situations where collaboration and competition are present. This capacity can vary depending on the individual differences and situational constraints, and we know a fair bit about the neurobiology and the consequences of poor mentalizing.

We now turn to an overview of the social phenomenon of negotiation, before moving to a theoretical outline of the role of mentalizing in interest-based bargaining.

NEGOTIATION

We now turn to an account of selected aspects of negotiation theory relevant for the present thesis. The chapter starts with the interest-based negotiation framework and continues by exploring how ‘interests’ of the parties are the main motivators of negotiation and by considering the notion of subjective value versus objective value in negotiation. We continue with the tension between creating and claiming value, a game theoretic framework that assists in understanding cognitive processes and choices in mixed motive negotiation, and conclude with the key lessons about negotiator cognition.

The review of negotiation literature in this chapter, jointly with the preceding chapter that discusses mentalizing, is essential for developing a conceptual platform for integration of these two unrelated fields and the development of the theory that mentalizing underpins bargaining, which we put forward and test in subsequent chapters.

What is ‘negotiation’?

For most people the word ‘negotiation’ conjures images of bankers in smoke-filled rooms haggling over the construction of an international oil-pipeline, lawyers arguing over a settlement amount in a billion-dollar class action, or leaders of countries concluding a trade deal or discussing an end to military conflict. And yet negotiation is a much more common and basic human interaction. Negotiating a curfew time with a child (or with the parent, if you are a child), discussing ending the probationary period with an unreceptive boss, spousal decision-making whether to go to a social event or stay in and watch Netflix on a kids-free night, discussing whose parking spot is under the tree with a neighbor, setting deadlines for employees, all of these are instances of negotiation. You negotiate every day. When you need someone else’s consent to get what you want, you need to negotiate: negotiation is an interpersonal decision-making process required whenever we cannot achieve our objectives on our own (Thompson, 2005, p. 22).

Negotiating with other people is a part of what makes us human. What drove the development of the social brain and cultural evolution (Dunbar, 1992, 1998, 2014; Moll & Tomasello, 2007; Tomasello, 2014, 2019) discussed in the first chapter (see *Mentalizing as an evolutionary adaptation*, p. 40), were situations that very much fit the definition of 'negotiation': I will catch that stag with you but only if we share equally, even though you are bigger; I will help you defend your den if you help me defend mine; and we will jointly sit in ambush to fight the alien others from the opposite side of the mountain. Human cooperation entails creating value, but also its distribution (i.e., creating and allocating societal resources) and the tension between creating and claiming in interdependent contexts is the key and defining feature of negotiation.

In the following section we present the key aspects of the negotiation process that are relevant for our thesis. An abridged glossary of negotiation terms is in the Appendix (p. 280).

Distributive negotiation

The second image that commonly comes to mind when thinking about 'negotiation' is what negotiation scholars call single-issue distributive bargaining (also, 'zero-sum' or 'price' negotiation), where the parties haggle over a resource they value *equally*. Equal valuations imply that a gain for one party corresponds to a loss of exactly the same value for the other. An archetype is a bazaar bargaining over the price of a rug. A dollar more for the seller is an exact same dollar less for the buyer; if I pay $900 rather than $1,000, I save the same $100 that the seller does not earn.

Pure zero-sum situations are extremely rare. First, there are always subjective elements that matter to parties that are outside of the distributive issue (see the discussion on subjective value in negotiation, p. 50., below). Second, even where the only issue is indeed the price, the disparities between the parties in wealth, risk preferences and differently construed gain-loss frames make an identical marginal utility of money extremely unlikely (i.e., the utility of a dollar to an individual with a high net worth is not the same as its utility to a heavily indebted student). However, understanding distribution and possessing skills claim value have considerable practical importance because most people assume, and consequently behave and expect their parties to behave, as if each negotiation is exclusively a zero-sum affair (Bazerman, Magliozzi, & Neale, 1985; Bazerman & Neale, 1986, 1991; De Dreu, Koole, & Steinel, 2000; Neale & Bazerman, 1992b; Pinkley, Griffith, & Northcraft, 1995; Thompson, 1991; Thompson & Hastie, 1990). This is an important issue to which we will return later in this chapter and, more comprehensively, in part two of this thesis. In addition, while not identical, the utility functions of the parties can be similar-enough to make some negotiations zero-sum for all practical purposes (e.g., a tenant and landlord negotiating the return of the rental deposit where the money matters to both, for different reasons, relatively equally). Finally, any value that is created in negotiation *must* be distributed, and the coming discussion on the bargaining range and claiming strategies matter greatly. While negotiation is (normally) not zero-sum, it also (normally) entails some distribution of the value between the parties.

Non-distributive (integrative) negotiation

More common than pure distributions are negotiations that carry the potential to create value (Deutsch, 1973; Pruitt & Rubin, 1986; Walton & McKersie, 1965). This value is available in situations that are not zero-sum, that is, when the benefit of a concession on an issue for one party does not equal the loss to the other. The first party's gain can correspond to either *a non-equal loss,* in which case the issue is *integrative*, or it can correspond to a *gain*, in which case the issue is *compatible*.

Compatible issues

In compatible issues, the parties' interests are perfectly aligned (e.g., both me and my counterparty desire to settle this lawsuit rather than litigate, and wish this settlement to be finalized before the disclosure process is triggered because that entails a major legal expense). There is no conflict at all. The compatible option is the best outcome from the parties' joint and individual perspectives (choosing any other option is a 'lose-lose' agreement; Thompson & Hrebec, 1996). Because there is no conflict, optimizing compatible gain is a mild coordination problem that can be resolved by a one-sided disclosure of general preferences on the issue and without both parties knowing the issue is compatible (Thompson & Hastie, 1990). In other words, it is sufficient if one party says, 'I would prefer to work in San Francisco than anywhere else'. The counterparty, whose preferences are identical (i.e., she wants the employee to be in San Francisco as well), will accept. The informing party may not know the preference of her counterparty is identical, and the counterparty may not be truthful regarding their own preferences. Instead, it may portray the acceptance as a concession, thus extracting extra value on another issue (Loschelder, Swaab, Trötschel, & Galinsky, 2014).

Integrative issues

Integrative value in negotiation stems from the *differences* between the parties' interests and preferences, rather than from the parties' identical opinions or preferences. This counterintuitive notion holds true in any contract. Value is found in differences (Jackson, Kaplow, Shavell, Viscusi, & Cope, 2011; Lax & Sebenius, 2006). For example, a property developer and a local authority are negotiating the construction and development of a new A&E center. The issues under negotiation are the price and the completion time. These two issues may be treated as zero-sum by the parties because every day and dollar more for the developer is a day and dollar less for the local authority. However, if the parties care about these issues differently, e.g., the developer cares mainly about its corporate objective of profit, whereas the local council needs the emergency center as soon as possible, a value creating trade of more money against quicker completion may be possible. Improved (or at least not harmed) relationships, procedural fairness and trust may further add to the value generated by the parties.

Differences enable logrolling trades and generation of integrative gain and are the most potent source of value in negotiation (Jackson et al., 2011; Lax & Sebenius, 2006). Perhaps counterintuitively, each sale and purchase deal is based not on agreement (that is necessary for a legal contract, but is not the economic reason for the deal), but on a *dis-agreement* about the value: if I am happy to sell you an item for a specific price, and you are happy to buy it at that price, that is so only because I value the item less than the price and you value it more. What we have here is a tradeoff between the value of money and the value of the item. These differences may lie in valuation (e.g., sale and purchase contracts), competencies and capabilities (e.g., joint venture agreements), funding ability and need (e.g., debt, mezzanine and equity funding deals), risk preferences (e.g., insurance contracts), timing concerns (e.g., short and long term leases) and expectations of future states (e.g., trades in futures and uncertain outcomes).

From the perspective of value creating and claiming, negotiators' interests in integrative issues are in conflict (e.g., the employer prefers a short holiday and a small bonus, the employee prefers a long holiday and large bonus). However, because the payoffs are not zero-sum, the parties have different *relative* preferences between the issues (e.g., the employer cares more about the holiday than the bonus, the opposite is true for the employee). The parties can create value by trading high value for low value options (e.g., high bonus against short holiday is better for both parties than splitting both issues 'down the middle'), a process known as logrolling (Froman & Cohen, 1970; Pruitt, 1983a). Compared to optimizing compatible gain, capturing integrative gain requires a much more nuanced information exchange on multiple issues (not only whether the parties prefer a high or low bonus, but how much more or less they prefer a high or low bonus relative to a short or long holiday). More importantly, to design an optimal integrative solution, a negotiator must engage in a deliberate cognitive effort simultaneously considering perspectives of both parties on multiple issues (e.g., if they care more about the bonus than about the holiday, and my preferences are the opposite, perhaps we could pay a high bonus and they would be willing to accept a shorter holiday?). A quantified example of dyadic and individual gains in a multi-issue negotiation is in the Appendix (p. 283).

Interests in negotiation

Value in negotiation corresponds to the degree the negotiation process and agreement satisfy negotiators' *interests*. Understanding this aspect of negotiation theory in detail is important for our thesis: to the extent interests are raison d'etre for negotiation, and interests are mental states, mentalizing must underpin bargaining.

In the following section we discuss how negotiation theory conceptualizes interests as subjective motivators of negotiation behavior, before turning to the concept of value in negotiation and how it is captured within negotiated agreements by settling integrative and compatible issues.

Interests as motivators of negotiation

We negotiate to achieve objectives that we cannot achieve on our own (Thompson, 2005, p. 22). Objectives in this context mean satisfaction of what the negotiation theory calls 'interests': the full set of concerns and desires of the negotiator (Fisher, Ury, & Patton, 1991), or 'whatever you care about that is potentially at stake in negotiation' (Lax & Sebenius, 1986a). To negotiate in a rational fashion then means 'making the best decisions to maximize your interests' (Bazerman & Neale, 1992).

Interests are the sole motivators of negotiation. It makes sense to negotiate inasmuch as you expect to serve your interests better by negotiating than by pursuing your no-deal alternatives. This primacy of interests and the interdependent nature of bargaining means that all negotiation cognitions and behaviors are – or ought to be, if one aspires to rationality – subordinate and instrumental to maximizing negotiators' interests, which includes making sure that the counterparty's interests are also met better than by their no-deal alternatives. In other words, all decisions and actions in negotiation are made in light of negotiator's own interests and with the interests of the counterparty in mind.

What are 'interests'? Researchers in economics and social and experimental psychology have long known that people are not rational utility maximizers (e.g., Kahneman, 1992, 2003; Kahneman & Tversky, 1984; Simon, 1955, 1956; Tversky & Kahneman, 1974, 1981, 1992). In negotiation, people care about more than just objective outcomes. One attempt at providing a taxonomy of interests was a conceptual paper by Carnevale and De Dreu (2006). Posing the question 'what is it that negotiators strive for?' (p. 56), the authors distinguished four relatively independent types of interests: (i) aspirations: negotiators preferences for specific outcomes such as how much they wish to achieve, or their reservation values; (ii) social motives: the negotiators' social orientation on a scale between pro-self (completely self-focused negotiator motivated solely by own gain) and pro-social (an altruistically oriented negotiator who cares about joint outcomes); (iii) identity motivation: a negotiator's wish to maintain a particular self-image in the negotiation (e.g., face-saving), and (iv) the epistemic motivation: a negotiator's desire to understand the intricacies of the bargaining situation, including the payoff structure, the counterparty's priorities and the potential for joint gain.

Subjective value in negotiation

Value in negotiation is the extent negotiators' interests are satisfied (Patton, 2005). Since interests are subjective (what negotiators 'care about'), it is negotiators' subjective perception of their gain that matters. In a series of experiments, Curhan and colleagues (Curhan, Elfenbein, & Xu, 2005; Curhan, Elfenbein, & Eisenkraft, 2010; Curhan, Elfenbein, & Xu, 2006; Curhan, Neale, & Ross, 2004) investigated the subjective value in negotiation. They found, first, that objective outcomes do matter: negotiators cared about the benefit received under the terms of the negotiated agreement, the balance between her own gain and the gain of the counterparty, whether they felt they 'lost' and whether the agreement was consistent with the principles of legitimacy (fairness). In other words, issue-specific, concrete success of parties on negotiation issues (such as, for example the amount of money they received in a sale of goods negotiation) correlated with their subjective satisfaction with the agreement. However, in addition to these outcomes, negotiators also cared about how the agreement made them feel. They cared about not losing face, behaving in line with their own principles and values, and whether they felt, during the negotiation, as a competent or incompetent negotiator. They also cared about the process: did the counterparty consider their point of view and interests, was the counterparty open to their concerns, and whether it was difficult to reach an agreement. Finally, negotiators cared about the relationship: did the counterparty make a positive or a negative impression, did they trust or mistrust their counterparties after the negotiation was done, and whether they were happy with the resulting relationship.

Summary

The concept of interest is a key aspect of negotiation theory. Interests define what is the goal of each negotiation: to have your interests met, whatever they may be. People negotiate only as long as they believe that there is a chance their interests will be better served in the ongoing negotiation than elsewhere. The value in negotiation is a derivative of the satisfaction of interests. The fact that *subjective* interests are the key driver of negotiation is critical for this thesis.

Tension between creating and claiming value

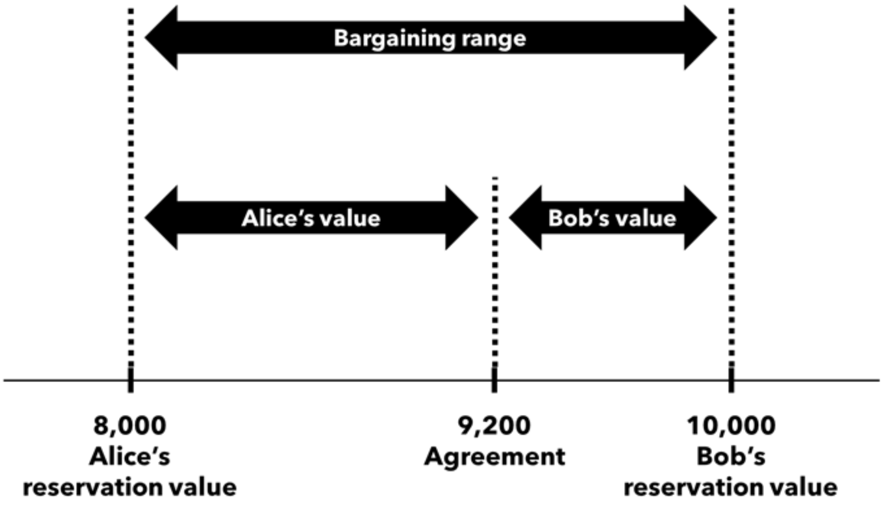
A key aspect of the negotiation process – the process negotiators engage in to satisfy their interests – is the tension between creating and claiming of value stemming from satisfied interests. We start with the basic theoretical notions required to understand this tension, including the bargaining range, information asymmetry and the value creating and claiming tactics. We then outline the tension between the simultaneous collaborative generating of joint value and competitive claiming aimed at maximizing an individual share in the joint value.

Bargaining range

The bargaining range in distributive negotiation is the difference between the parties' reservation values, which are determined by how well each party's interests are served by the courses of action they can take outside of the negotiation (Lax & Sebenius, 2006). For example, if Alice wants to sell a car because she is moving to another country (and if the price is the only issue in negotiation), the minimum price she is willing to accept depends on her no-deal alternative. If she has a neighbor who made an offer of £8,000 for the car, that is her reservation value: any deal above £8,000 is economically beneficial (e.g., better than selling to the neighbor) and any deal below that makes no sense (she would rather sell to the neighbor).

Let's further assume that Bob is interested in buying a car just like the one Alice is selling. If Bob's no-deal alternative is buying the same model from the dealer for £10,000, that is Bob's reservation value: paying less means savings and paying more makes no sense. From the joint perspective of a fully informed observer, the bargaining range (or the 'zone of possible agreement'; Raiffa, 1982), amounts to £2,000, the difference between Alice's reservation value of £8,000 and Bob's of £10,000. Any deal between Alice and Bob for any price between these two numbers is a good deal for both parties: they capture £2,000 of joint value that is available. For example, if they agree a price of £9,200, Bob pays £800 less than his reservation value, and Alice earns £1,200 more than hers; this is how they split the joint value they created: £1,200 to Alice and £800 to Bob. For a graphic representation see **Figure 7**.

Figure 7. Example of a distributive negotiation



Note. The agreement at £9,200 splits the £2,000 of generated value (bargaining range) between Alice (who earns £1,200 more than what she would have got from the neighbour) and Bob (who pays £800 less than he would have paid to the dealer).

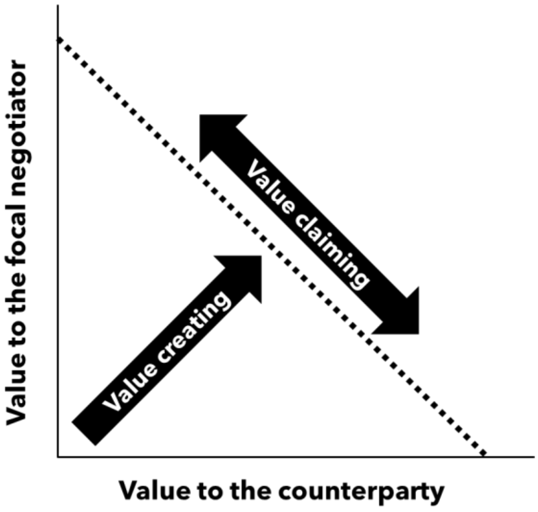
If they fail to reach an agreement, they waste an opportunity to create £2,000: reverting to their alternatives, Alice will sell it to the neighbor for £8,000 and Bob will buy it at the dealer for £10,000.

An agreement is superior to an impasse whenever the bargaining range is positive. In such cases, the negotiation has the potential to satisfy interests of both parties better than they could on their own. For this to happen, however, both parties need to agree. Withholding consent by any one party is an effective veto on an agreement. This highlights the interdependent and collaborative aspects of bargaining even in pure distributive issues: in terms of capturing gains that are unique to the bargaining situation, the parties are 'in the same boat'; their individual gains depend on the collective consensus.

If negotiation involves multiple issues where the parties have different or compatible preferences, the bargaining range is flexible rather than fixed, and the parties can widen or narrow it by identifying compatible issues and trading on the issues that have integrative potential.

A common conceptualization of the value and bargaining range involves a Pareto graph with negotiators' gains or satisfaction represented on the axes (see **Figure 8**).

Figure 8. Value creating and value claiming in dyadic negotiation

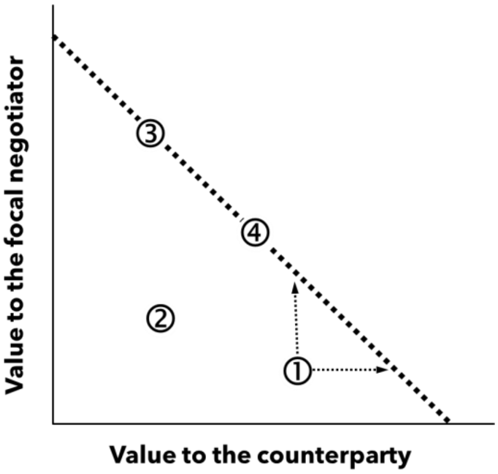


Note. The axes indicate the value of possible deals to the parties (the focal negotiator and her counterparty).

For example, if Alice's and Bob's negotiation also included a set of custom-made speakers (worth £100 to Alice but £300 to Bob), the right to park in the neighborhood (worth £200 to Alice but £1,000 to Bob) and a one-week lease clause so that Alice can run errands before she leaves the country (worth £500 to Alice but £100 to Bob), the bargaining range would be a number between £2,000 (if the parties do not include any of these additional issues) and £3,400 (if the parties include all of them).

The individual gains would depend on both how much joint value the parties create and how they distribute it. Refer to **Figure 9** for examples of contracts varying in efficiency and equity. For a detailed example of calculating gains in multi-issue negotiation please refer to the Appendix (p. 283).

Figure 9. Contracts with different levels of value and equity



Note. Contract 1 is unfair and inefficient; the counterparty is claiming more value than the focal negotiator, and the value to either negotiator could be improved without decreasing the value to the other (dashed arrows). Contract 2 is fair but inefficient; the value to both parties could be improved, but the value the negotiators had created was distributed equally. Contract 3 is efficient but unfair; all value has been captured (i.e., no gains can be made by any party without decreasing the value from the other), but the focal negotiator has claimed more than the counterparty. Contract 4 is both efficient and fair.

Information asymmetry, strategic interdependence and distributive moves

The above comes with an important caveat: the parties negotiate under *information asymmetry*. While each negotiator knows his or her reservation value, they do not know the counterparty's, and consequently cannot know for certain what the bargaining range is or, in fact, whether there is one at all. This has two consequences. First, negotiators claim value for themselves. Even if they do not focus on it, while they create value negotiators are simultaneously distributing it. For example, if Alice and Bob strike a deal at £8,500, they generate the £2,000 of joint value (value creating), but they also claim what they created: Alice gets £500 (i.e., receives £500 more than she would from her neighbor) and Bob gets £1,500 (i.e., pays £1,500 less than he would to the dealer).

Most value-claiming moves in negotiation, such as anchoring, converging concessions, non-offer offers, are based in more or less explicit misinforming about the negotiator’s reservation value (see the chapter on distributive tactics aptly titled 'Shaping perceptions to claim value' in Lax & Sebenius, 2006). For example, Bob may think that because Alice does not know he is willing to pay £10,000, perhaps she will accept less, and makes an aggressive opening offer: ‘this car has clearly not been well taken care of; I cannot possibly pay more than £5,000 for it’. Second, because both parties will engage in this kind of claiming and be wary of the other side engaging in it, capturing the available joint gain is not a given. Focused on claiming, the parties can simply overlook the possibility to create value (for example, Alice and Bob do not even consider including the speakers, parking place and lease as issues). They can also manage to convince each other that their reservation values are really such that the bargaining range is negative and that there is no deal to be made (e.g., Alice replies that she in fact does not have to sell a car at all, and the minimum she could possibly accept is £12,000). This is the essence of the tension between creating and claiming value that we discuss below.

Creating and claiming value: Negotiator’s dilemma

The tension between creating and claiming is a critical aspect of negotiation. Negotiators have simultaneous goals of maximizing joint and individual gain. Because information about interests and preferences is asymmetrically distributed, maximizing joint gain (value creating) depends mainly on effective exchange of information regarding the parties’ interests and preferences. The behavioral strategies thus include honest information exchange, (epistemic) trust and clear and efficient communication. Maximizing individual gain (value claiming), on the other hand, requires the practice of secrecy, deception and mistrust, behaviors diametrically opposed and mutually exclusive to the behaviors needed for value creation (Lax & Sebenius, 1986b).

Strategic interdependence adds another layer of complexity to this dilemma. A negotiator who behaves cooperatively and discloses key information about her preferences is at risk that the counterparty will reciprocate with an attempt to claim. For example, if Bob attempts to create value (by e.g., disclosing he really does need a parking permit and that it’s worth £1,000 to him), Alice can respond either cooperatively (honestly disclosing that the parking permit, to her, is not worth much at all), or competitively (that £1,000 is exactly the offer she already has elsewhere, but can sell the car to Bob for that amount if he so wishes). A competitive counterparty can exploit any disclosed information, so providing information is risky (Lax & Sebenius, 1986b; Mnookin, 2000; Murnighan, Babcock, Thompson, & Pillutla, 1999). Relying on imparted information is risky too, because from the perspective of the focal negotiator, the counterparty's attempts at deception-based value claiming are indistinguishable from value creating. For example, if Bob says something like 'I honestly cannot pay more than £8,000 for the car', it is virtually impossible for Alice to ascertain with any degree of certainty whether Bob is creating value (by informing of a real constraint) or claiming it (by attempting to shape her perception of what is possible).

This tension between creating and claiming has been compared to the prisoner's dilemma with attempts at value-creation mapped onto the move of cooperation and value-claiming on the move of defection (Lax & Sebenius, 1986b; Mnookin, 2000). The payoff structures between the negotiator's dilemma and the prisoner's dilemma are similar. A game theoretic analysis suggests that in the negotiator's dilemma, absurdly and just like in the prisoner's dilemma, the individually rational choice is to claim, which leads to a jointly suboptimal outcome. However, unlike the prisoner's dilemma, the negotiator's dilemma allows communication, credible commitments and multiple rounds of interaction. Therefore, the negotiator's dilemma implies a tendency for both parties to choose to claim (instead of create), but not that this choice is an irrevocable equilibrium (Lax & Sebenius, 1986b).

In the next section we turn to the key lessons from social and cognitive psychological research in negotiation. The emphasis is on biased information processing of negotiators, resulting from the implicit assumptions that negotiators make about each other and the negotiation situation.

Negotiator cognition and irrationality

The asymmetrically normative/descriptive approach combining advice and research pioneered by Howard Raiffa (Raiffa, 1982) resulted in the dialogue between prescriptive and descriptive researchers in negotiation (Neale & Bazerman, 1992b). This interaction ultimately led to behavioral decision-perspective to negotiation, largely based on the findings of Nobel laureate Kahneman and Tversky, that highlights the bounded rationality of negotiators whose decision making, predominantly based on heuristics, often results in suboptimal agreements.

A number of systematic negotiation-relevant biases have been highlighted by decision-perspective research, including the excessive power of framing and anchoring, overreliance on available information, irrational escalation of commitment, blind spots and reactive devaluation (Bazerman & Neale, 1992; Ross & Ward, 1995).

For our purposes, the key aspects of this research are the fixed-pie and incompatibility biases. They are the anticipated key features of negotiation and serve to distort the perception of the interdependent task coloring it in competitive overtones, which in turn results in negotiators employing their cognitive capacity for competitive rather than collaborative purposes. We discuss this extensively in part two of the thesis.

Fixed pie bias

That negotiators assume, often unconsciously, that the value in negotiation is fixed (i.e., we both want the same thing equally and a dollar more for you is a dollar less for me), is one of the most robust findings in negotiation research. The fixed-pie bias, or the zero sum assumption as it is sometimes called, is well documented in negotiation literature (Bazerman et al., 1985; Bazerman & Neale, 1986, 1991; De Dreu, Koole, et al., 2000; Menkel-Meadow, 1983; Neale & Bazerman, 1992b; O'Connor & Adams, 1999; Pinkley et al., 1995; Thompson, 1991).

The result of the fixed-pie bias is that the negotiation is thought to be purely distributive (see *Distributive negotiation*, p. 50). The zero-sum frame leads to a predictable loss of value as negotiators behave as if the only worthwhile effort in negotiation is value distribution. They miss out on the value potential inherent in integrative issues, where they could logroll the issues based on their relative preferences (for a detailed discussion of integrative issues see p. 51).

Incompatibility bias

A related bias is that negotiators' interests must be incompatible (i.e., if I want this, you cannot possibly want it too). This bias is associated with lose-lose agreements, where in situations where the parties have compatible interests (e.g., both employer and employee prefer the employee to work in San Francisco) somehow end up with an agreement where they are both worse off (e.g., the employee works somewhere else, see *Compatible issues*, p. 52). In the introductory chapter we showed that failing to agree a compatible issue is a staggering coordination failure and reflects, from the perspective of social cognition, a fundamental failure of mentalizing. In a classic set of experiments that first highlighted this propensity of negotiators to seal lose-lose deals, Thompson and Hastie (1990) directly measured the expectations negotiators brought to the table and found that they assumed their interests to be directly and completely opposed to those of their counterparty. Some, but not all, learned about the potential for joint value during the process of bargaining, and most of this learning occurred at the very beginning of the negotiation. However, a substantial majority of negotiators failed to realize that their interests in compatible issues were perfectly in tune with the interests of their counterparty.

Expected salient features of 'negotiation'

What do people think is required in 'negotiating'? O'Connor and Adams (1999) asked novice negotiators to list the behaviors required for a successful negotiation. They found that people share beliefs about a 'correct' sequence of behaviors in negotiation, but that those behaviors are not the behaviors associated with good outcomes. In particular, negotiating beginners expected (i) their interests to be incompatible with the interests of their counterparties, rather than potentially compatible or relatively-different, thus allowing for compatible gain or logrolled integrative gain, (ii) that the correct sequence of negotiating calls for issues to be settled sequentially, which prevents identification of relative differences (that facilitates integrative gain) and cements the conflict and zero-sum aspect of each issue, rather than simultaneously, which has the potential to reverse that, and finally (iii) that the negotiation is likely to result in an impasse (again emphasizing the conflict aspects of the bargaining process).

Later studies replicated and extended these findings. For example, Van Boven and Thompson (2003) examined the relationship between negotiators' mental models and the outcomes they achieved in mixed integrative-distributive tasks. They found that negotiators who were able to reach optimal outcomes understood the payoffs and information exchange (enabling them to spot relative differences that enable integration) and the trading (logrolling) processes (e.g., simultaneous rather than sequential issue settlement) far better than their peers who reached suboptimal outcomes.

In the second part of this thesis, we develop a theory that the persistently suboptimal outcomes are potentially the result of the way parties construe ambiguous and uncertain interactions we call 'negotiations', and that the above outlined expectations critically bias such construals and consequently drive competitive choices resulting in poor outcomes (pp. 181-210).

Conclusion

This chapter provided an overview of the critical aspects of the negotiation process required for the conceptual platform that we need to develop the idea of how mentalizing underpins bargaining.

Negotiation is an intensely interpersonal affair, involving parties who attempt to get their interests (key motivators) satisfied via bargaining better than they could on their own. Rarely, the interests of the parties are diametrically opposed and the negotiation is purely distributive. More commonly, negotiation involves issues where the parties' preferences are identical (compatible issues), and issues where their preferences are different (integrative issues).

Negotiation is also a complex affair. It takes place under information asymmetry regarding the key aspects of value and under the condition of strategic interdependence. More importantly, negotiation is a mixed-motive game (Schelling, 1980), involving simultaneous motives (and hence behaviors) of negotiators to create and claim value. At the same time, the behaviors pursuing these two goals are mutually exclusive: claiming impedes creating.

This tension is present in both distributive negotiation and negotiation that has (hidden) value potential. In distributive settings, the creating aspect is captured in the value represented by the bargaining range (i.e., do the parties close the deal or not), and the value claiming aspect is in how much of the bargaining range goes to each party. The task is to get as much as you can (value claiming) while making sure the other side gets enough to still say ‘yes’. In integrative negotiation, the tension between creating and claiming is more pronounced. The negotiators are both creating and claiming value at the same time. The value creating moves involve providing accurate information and trusting the information imparted by the counterparty with a view to identify any compatible issues and logroll any integrative ones. The value claiming moves are similar to the ones used in distributive negotiation and aimed at maximizing a negotiator’s share in joint value, whatever that may be.

Finally, negotiator rationality is bounded by cognitive constraints. While negotiators may attempt to optimize, they frequently rely on heuristics that result in biased decision making. The key biases relevant for our work are the fixed-pie bias, leading the negotiators to assume that negotiation is a strictly distributive affair, and its cousin the incompatibility bias, leading negotiators to enter into lose-lose outcomes.

In the next chapter, we introduce the idea that because negotiation is about interests (and interests are mental states), and because value creating and value claiming strategic moves are made with the other side’s mind in mind, negotiation must involve mentalizing.

MENTALIZING NEGOTIATION: SOCIAL COGNITIVE INFERENCE IN INTEREST-BASED BARGAINING

The next section of the thesis introduces a novel model of mentalizing-based negotiation. The key part is understanding the negotiation process in terms of causation between mental states in negotiators and their choices of behavior. Because negotiators' interests and beliefs about what is possible are mental states, mentalizing must underpin negotiation.

This is a novel proposition. To the best of our knowledge based on a comprehensive literature review outlined below (pp. 92 - 103), no existing study or paper considers mentalizing as underpinning the critical representational and decision making processes in negotiation by allowing negotiators to infer own and others' mental states in order to determine their bargaining choices.

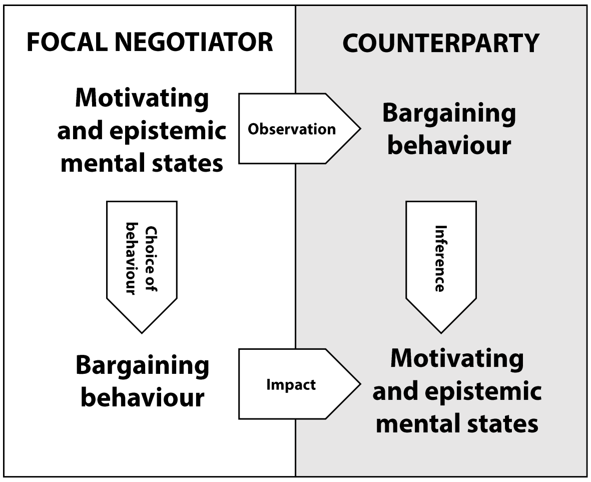
We first develop and articulate this model and apply lit to the two key aspects of negotiation, value creating and value claiming, and outline the two testable hypotheses that we examine in the chapters that follow: that mentalizing predicts both value creating and value claiming in negotiation. We continue with the theoretical implications of the model, particularly regarding the phenomenology of poor mentalizing in bargaining. We conclude by discussing the impact of cognitive load and stress, the two moderators highly likely to be present in negotiation, on the mentalizing and bargaining processes.

The model

Mentalizing in negotiation

Critical aspects of decision making in negotiation depends on mentalizing. Before and during negotiation, a negotiator engages in self-focused mentalizing to explicate their goals (interests) and to plan the optimal bargaining behavior (strategy and tactics) to achieve these goals. In addition to negotiator's interests, these choices of behavior critically depend on the contingent behavior of the counterparty (e.g., an action will be effective if they agree to whatever the focal negotiator is trying to achieve). To effectively predict the counterparty's behavior, the focal negotiator must construct, using mentalistic inference, a mental model of the relevant mental states of the counterparty (see **Figure 10**).

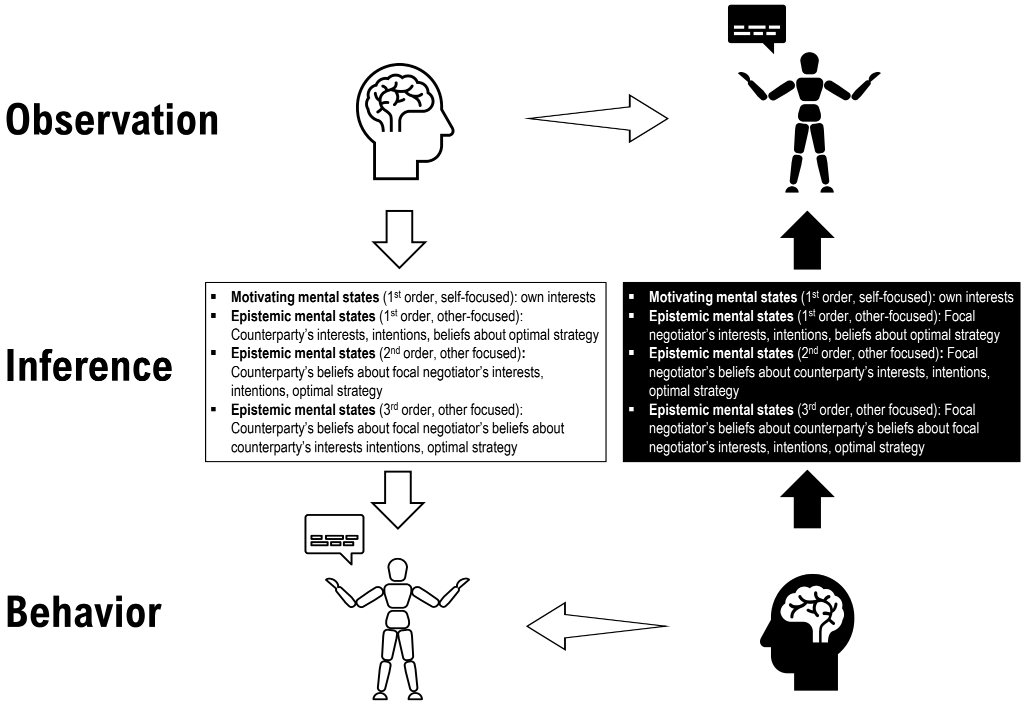
Figure 10. Mentalizing in negotiation from the perspective of the focal negotiator



Note. Motivating and epistemic mental states include awareness of own interests and beliefs about the counterparty's mind. They lead to decisions about optimal course of action (bargaining behavior).

The mental states essential to decide negotiators' behavior include, inter alia, (1) the focal negotiator's own interests, inferred via self-focused mentalizing (i.e., what is it that they wish to achieve); (2) a first order interpersonal representation of the counterparty's interests and their beliefs of optimal action (i.e., what the counterparty wishes to achieve and how they believe they ought to go about it), (3) a second-order interpersonal representation of the counterparty's representation of the focal negotiator's interests and bargaining strategy (i.e., what does the counterparty think we want and how we are trying to achieve that); and so on and so forth (**Figure 11**). While theoretically there is no limit to the order of reflection, representations of the third-order (what the focal negotiator thinks what the counterparty thinks that negotiator thinks that the counterparty thinks) and above quickly become too abstract to contribute to decision making (see Weerd, Verbrugge, & Verheij, 2017 for analysis of how theory of mind assists computer agents).

Figure 11. Detailed model of mentalizing in negotiation



- **Motivating mental states** (1st order, self-focused): own interests
- **Epistemic mental states** (1st order, other-focused): Counterparty's interests, intentions, beliefs about optimal strategy
- **Epistemic mental states** (2nd order, other focused)**:** Counterparty's beliefs about focal negotiator's interests, intentions, optimal strategy
- **Epistemic mental states** (3rd order, other focused): Counterparty's beliefs about focal negotiator's beliefs about counterparty's interests intentions, optimal strategy

- **Motivating mental states** (1st order, self-focused): own interests
- **Epistemic mental states** (1st order, other-focused): Focal negotiator's interests, intentions, beliefs about optimal strategy
- **Epistemic mental states** (2nd order, other focused)**:** Focal negotiator's beliefs about counterparty's interests, intentions, optimal strategy
- **Epistemic mental states** (3rd order, other focused): Focal negotiator's beliefs about counterparty's beliefs about focal negotiator's interests, intentions, optimal strategy

At each of the above points of contact between mentalizing and negotiation ineffective mentalizing will impair (1) metacognition, (2) first-order and (3) second-order inference, preventing the parties from engaging in negotiation in an efficient mode where the focal negotiator and the counterparty act together to create added value deriving from the appropriate identification of compatibility and differences between their perspectives.

Mentalizing in value creating

Mary Follet's (Follet, 1925) anecdote about a quarrel over a window in the library is a case in point (reported and applied to interest-based negotiation in Fisher et al., 1991, p. 40):

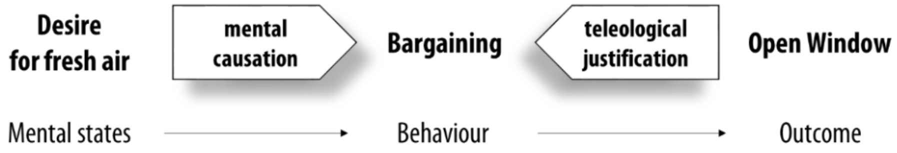
> *'Consider the story of two men quarrelling in a library. One wants the window open and the other wants it closed. They bicker back and forth about how much to leave it open: a crack, halfway, three quarters, all the way. No solution satisfies them both. Enter the librarian. She asks one why he wants the window open: "To get some fresh air." She asks the other why he wants it closed: "To avoid the draft." After thinking a minute, she opens wide a window in the next room, bringing in fresh air without a draft.'*

The key insight of the vignette is in the cognitive leap that occurs when the quarrelling men's behavior starts to make sense in terms of mental states. The men are not really after an open or closed window. They want air and no draft. These mental states (interests) motivate the parties' bargaining behavior, but are not made explicit. Because the discussion – which can quickly escalate into a heated argument – is about the *window*, the men cannot find the solution. But it ought not to be about the window; their quarrel reflects a failure to mentalize. The men are preoccupied with tangible action and behaviors and do not realize they are motivated by observationally opaque mental states – interests. The librarian's intervention is effective exactly because it triggers mentalizing in men when she asks *why* they *want* the window open or closed.

Bargaining of the quarrelling men is teleological (see **Figure 12**). Their behaviors are naively rational because the pursued outcomes justify them. Demanding and insisting on opening the window may lead to it being open and rejecting these demands will keep it shut (which will instrumentally satisfy the opaque underlying motivators of fresh air and keeping the draft away). The parties are engaged in a tug of war with words; they may as well be physically pushing and pulling the window.[1]

[1] Going beyond the facts of the vignette, the men's behavior could also be the result of pseudomentalizing (the other's mental states are projectively infused with their own fears, concerns, thoughts, e.g., the counterparty is 'aggressive', 'obnoxious', or 'crazy') or by hypermentalizing

Figure 12. Mental causation and teleology in Follet's anecdote about the library window



Note. Teleological reasoning explains the negotiator's behavior in terms of goals it justifiably leads to given the reality constraints (he is bargaining *to* get the window open), which leaves mental causation out of focus. Mentalistic reasoning accounts for prior generating interests and beliefs about reality constraints (the negotiator is bargaining *because* he *wants* fresh air and *believes* opening the window is the best way of achieving that). The opacity of mental states means that modelling of mental causation needs to be inferential.

The library example shows that the competition about the window that triggers, or rather creates, anxiety in both parties which undermines the ability of either to show curiosity about the other's mental states. They make presumably automatic assumptions about what the other is thinking that are superficial and inaccurate. This rapidly regresses to a teleological representation of complex mental states where the freedom to open or close the window can be measured in inches. The teleology is an extreme instance of attending to external rather than internal cues: cognition is dominated by emotion and the perspective of the self takes a position of undisputed prominence compared to the putative experience of the other. As the result, neither metacognition, first-order mentalizing, nor any higher levels of inference can take place. The librarian, who is not incapacitated by the emotional upheaval, can find a solution that meets the stated needs of both parties.

We are proposing that interest-based, value creating negotiation requires negotiators to assume the joint intentionality mode (see p. 44) when approaching negotiation. In this context the negotiating dyad (i.e., the focal negotiator and the counterparty or counterparties) at least partially evaluate the successes and failures from the perspective of collaborative activities. For that to take place, the individuals in the dyad must assume intentional mental states that entail a degree of de-individuation. To some degree, self-identification must be subsumed into that of a negotiating unit, which enables distinctive ways of functioning. It is worth stressing that there is no mysterious

(creating exceedingly specific but improbable mental states, e.g., he wants the window closed because he is agoraphobic). We will return to this later in the chapter.

leap into the mystical interpersonal space of “we-ness”. Simply, the ‘I’ now sees itself and the interpersonal landscape as at least partially “we-structured”: such ‘we-mode’, which corresponds to Tomasello’s joint intentionality, is an individual state. A negotiator perceives themselves as a part of the unit and sees joint action and collaboration as a goal. We say ‘partially’ because while being a part of the negotiating unit, negotiators are also individuals with interests that do not fully overlap with the interests of the unit, resulting in the tension discussed in the literature (see pp. 56-63). Looked at in this way it is hardly surprising that fully integrative negotiation in ‘we-mode’ is rarely achieved. It requires the content of individual intention to be transformed. Each negotiator needs see themselves and the other as intending to play a mutually collaborative role in the action of negotiation.

According to our model, the imperative to ‘focus on interests, not positions’ (Fisher et al., 1991) can be understood not as a call for a clear differentiation of interests from positions, which has been criticized as arbitrary (Provis, 1996), but as a process-focused suggestion to look beyond behavior toward motivating mental states. In fact, much negotiation prescriptive advice relating to discovery of interests (e.g., Fisher et al., 1991; Lax & Sebenius, 2006; Thompson, 2005) is remarkably similar to techniques of mentalization-based psychotherapy (e.g., Bateman & Fonagy, 2016). For example, the authors of Getting to Yes stress that ‘[t]he ability to see the situation as the other side sees it, as difficult as it may be, is one of the most important skills a negotiator can possess. It is not enough to know that they see things differently. If you want to influence them, you also need to understand empathetically the power of their point of view and to feel the emotional force with which they believe in it’ (p. 23). They recommend explicit mentalizing of perceptions and affect (‘make emotions explicit and acknowledge them as legitimate’; p. 30) and explicit self-mentalizing (‘Look at yourself during the negotiation. Are you feeling nervous? Is your stomach upset? Are you angry at the other side?’; p. 29). Furthermore, inquiring ‘why’, ‘what is the reason they may be saying no’, and tools such as the Currently Perceived Choice are exactly the questions that elicit explicit social cognition.

Interests in negotiation are mental states that need to be inferred. The motivating force and the subjective importance of such mental states derive from their affective charge. Current neuroscience research suggests that affects are states of the subject, representing its internal (emotional) responses to the situation it finds itself in, with a purpose of signaling how well it is 'advancing the cause of staying alive to reproduce' (Solms, 2015). If we are doing well, the affects are positive and feel ‘good’; if not, they feel ‘bad’ (Solms, 2013). These rudimentary emotions originate in the subcortical brain regions as basic emotion-action sequences linking affect with a prepotent behavioral response (e.g., fear with fight-or-flight action-sequence). These can be downregulated as well as embedded in and elaborated on by our secondary (largely upper limbic) and tertiary (largely neocortical) higher-cognitive processes into a complex cognitive-emotional experience. For example, a desire for achieving a high pay in a salary negotiation contains a cognitively elaborated-embedded emotional desire for safety, which is itself a cognitive elaboration of the basic activation of the seeking emotion-action system (Panksepp, 2004).

Negotiators concurrently hold a multitude of bargaining-relevant motivating mental states at different levels of cognitive elaboration. Sometimes these are congruent, such as the desire for high pay, success, and respect, and sometimes they are in conflict with each other, such as a long term versus short term profit, fairness versus economic gain, and honesty versus impression management (Patton, 2005). These mental states are not necessarily explicit. A clear understanding of motivating mental states is not a given. A directed effort is often required for their explication (Fisher et al., 1991). This effort is mentalizing.

Mentalizing is thus necessary to explicate, or at the very least get an intuitive understanding of, interests in oneself and the counterparty. Mentalizing will also assist in grasping the interests that drive any principals in agency relationships and those of any constituency negotiators represent. This is particularly important because while negotiators in agency relationships must represent the interests of their principals and constituencies as if they were theirs, they obviously do not know them by default any better than those of their counterparties. As interests are the critical element of negotiation, its *raison d'être* so to speak, mentalizing is a critical capacity for negotiators.

Fairness, for example, is an interest. Research in the Ultimatum Game investigates this interest. The setup is as follows. In a two-player, one shot game, one player (proposer) suggests a specific split of a potentially available amount (e.g., £20) to her counterparty, the responder (e.g., I get £18 and you get £2). If the responder accepts, the parties receive the amounts according to the proposer's suggestion. If the responder rejects, nobody gets anything. Evidence in such experiments unequivocally shows that we possess a deep, implicit aversion to inequity: the mode offer is an even split, the average demand by the proposer is for less than 70% of the amount, and proposals offering less than 20% are routinely rejected (Güth, Schmittberger, & Schwarze, 1982).

Aversion to inequity is common in mammals and seems to be evolutionarily driven; comparing ourselves with our peers in terms of how we fare is a reliable predictor of reproductive success (Brosnan, 2008) and as such an implicit source of affect and motivation. What is particularly interesting is that experiments comparing responses to offers respondents thought were made by people versus identical offers they thought were made by computer algorithms (in reality all were made by computers) show significantly higher agitation and rejection rates if the offers were thought to come from humans (Sanfey, Rilling, Aronson, Nystrom, & Cohen, 2003). There may be good evolutionary reasons for treating actions of people differently from mindless 'events' such as algorithm outcomes or natural occurrences. For the purposes of this thesis, the key implication is that it is the process of mentalizing that differentiates offers based on authorship, and that such process takes place only if the agents feel the decisions involve other minds.

Now we turn to compatible and integrative issues (see pp. 52-54, above). The preferences that drive joint gains in such issues – e.g., we both prefer the job to be in San Francisco, and I care more about the bonus and you care more about the holiday – are mental states subject to self-focused and other-focused mentalizing. While identifying compatible issues does not require mentalizing *stricto sensu* (one party may disclose their preference that is compatible with the other side's, and the counterparty can accept this option without disclosing their preference was identical, see Loschelder et al., 2014), it is without doubt helpful because identification of compatible issues is not a given (up to 40% of negotiators fail to correctly settle them; Thompson & Hrebec, 1996). However, for trading integrative issues, mentalizing is critical. We explained above (p. 52) how logrolling requires a deliberate cognitive effort simultaneously considering both parties' perspectives on multiple issues. Such cognitive effort presupposes a clear picture of one's own and the counterparty's preferences, which requires mentalizing.

In summary, interests are mental states and must be inferred. For value creating, a clear understanding of one's own interests (self-focused mentalizing) and the interests of the counterparty (other-focused mentalizing) is necessary. This gives us a testable hypothesis: the quality of value creating should correlate with the quality of mentalistic inference.

Mentalizing in value claiming

While less obviously than integrative negotiation, distributive bargaining also benefits from social cognitive inference. Successful value claiming involves behaviors that shape perceptions of the counterparty about what is possible (Lax & Sebenius, 2006, p. 183). The intention behind the tactics that 'shape perceptions' is to influence the mental states of the counterparty, specifically their beliefs about the focal negotiator's reservation value and the negotiation's bargaining range, which should then result in increased odds that the counterparty accepts a proposal favorable to the focal negotiator. Two counterparty's mental states are relevant: the counterparty's own reservation value and the counterparty's belief about the focal negotiator's reservation value. Put simply, to claim successfully, a negotiator must convince the counterparty that his reservation value is much closer to the counterparty's than it really is. For example, a prospective car buyer's[2] claiming tactic will be effective if it convinces the seller that the buyer is able to accept no less than £9,800 for the car because it is much closer to the seller's reservation value of £10,000 than to the buyer's of £8,000. For that, the negotiator must, first, accurately estimate the counterparty's reservation value and, second, convince the seller that the buyer's value is close to theirs. If successful, the counterparty will be left with a choice between the focal negotiator's offer and their alternative, and if the offer is superior to the alternative, this will be the deal. Therefore, a successful execution of virtually any value claiming tactic – e.g., ultimatums, first

[2] We are using the example of Alice and Bob negotiating over a car, see above, p. 57.

offers, converging concessions, lock-ups – requires estimating one mental state of the counterparty and manipulating another. Both require mentalizing.

Furthermore, most prescriptive advice on value claiming is based on sophisticated predictions of how the majority of people behave, so on a type of expected-value mentalizing. For example, first offers: Lax and Sebenius (2006, p. 187) suggest: 'In cases when you are not hopelessly uninformed, seriously consider going first'. This simple one-line heuristic is a result of high-level abstract mentalizing: you ought to go first because people, on average, are susceptible to anchoring (Furnham & Boo, 2011) if their estimates are malleable, and in negotiation the parties' expectations about what is possible are very flexible during the beginning stages of the interaction. The go-first advice takes advantage of that. First offers anchor the negotiation (Huber & Neale, 1986; Neale & Northcraft, 1986; Northcraft & Neale, 1987) and the effect has proven to be stable across cultures and multi-issue negotiations (Gunia, Swaab, Sivanathan, & Galinsky, 2013). The recommendation to go first is the result of explicit probabilistic inferential reasoning based on an average susceptibility to anchoring grounded in empirical evidence.

The second part of this prescription – the 'unless uninformed' qualification – is also based on such expected-value, abstract mentalizing. Bazerman (2013, p. 55) provides an evocative example:

> *'Imagine that while traveling in a foreign country, you meet a merchant who is selling a very attractive gemstone. Although you have purchased a few gems in your life, you are far from an expert. After some discussion, you make the merchant an offer that you believe, but are not certain, is on the low side. He quickly accepts. How do you feel?'*

The feeling of being taken advantage of – rather odd given the fact that a second ago you made this offer expecting it to be rejected – comes from inference of what the quick acceptance means in terms of the counterparty's mental state (they would accept a lot less!). Falling victim to this so called 'winner's curse' was due to insufficient consideration of the counterparty's cognitions (they likely know the exact value of the stone). Research showing similar phenomena has been replicated over and over (Bazerman & Chugh, 2006; Chugh & Bazerman, 2007) and demonstrates that people tend to leave contingent decisions of others out of focus, which is likely a result of poor mentalizing.

Conversely, recipients of first offers who engage in explicit mentalizing are less prone to anchoring effects. Galinsky and Mussweiler (2001) demonstrated that while first offers were a strong predictor of the final deal prices, when the recipient of the first offer thought about the counterparty's perspective by considering their alternatives to the negotiation, reservation price, or even his or her own target – that is, if she mentalized – the anchoring effect was eliminated. The findings were quite robust and the interpersonal influences were controlled for; the results were present in both negotiations in person and by email.

Finally, it is worth emphasizing that no matter how abstract the mentalizing that underpins the advice to go first (unless uninformed), the execution of such anchoring will require very concrete, situation-specific mentalizing as described in the first paragraph of this section: the negotiator will need to infer the counterparty's reservation value, and continue inferring their beliefs about what the focal negotiator's bottom line is as they try to manipulate them. This results in our second testable hypothesis: the effectiveness of value claiming should correlate with the quality of mentalistic inference.

Mentalizing and negotiation outcomes

While there is currently no research on how mentalizing itself affects bargaining outcomes there is evidence that selected capacities that stem from mentalizing do have such effect. For example, Neale and Bazerman (1983) investigated whether the capacity for perspective taking benefits the parties in negotiations under conventional and final-offer arbitration and found that the perspective taking ability positively affects the concession rate, number of issues resolved, and quality of outcomes in negotiation. Galinsky and collaborators (2008) showed that perspective taking increases the ability to both create and claim value in negotiation. The studies showed that the integrative and distributive gains of the subjects correlated with their assessed perspective taking capacities (first study) and with the priming of perspective taking (second and third study). The follow-up set of experiments (Gilin, Maddux, Carpenter, & Galinsky, 2013) differentiated the cognitive perspective taking and empathic social competency to show that each capacity is useful in different types of bargaining involved. Another study showed that psychopathic traits may predict success in competitive bargaining and losses in cooperative bargaining (Ten Brinke, Black, Porter, & Carney, 2015), which given the impairments psychopaths show in emotional empathy and their strong performance in cognitive mentalizing (Blair, 2005) supports the conclusion of varying importance of these two aspects of mentalizing in bargaining.

Artificial intelligence bargaining experiments (de Weerd, Verbrugge, & Verheij, 2013a) provide strong support that the theory of mind helps negotiators meet their interests. The computer agents played multiple rounds of a game (Colored Trails) that featured both integrative and distributive aspects. The negotiators had a confidential set of preferences and the interaction was limited to making and accepting or rejecting offers. The hypothesis that the theory of mind reasoning would allow the agents to inferentially reason (mentalize) the counterparty's interests was confirmed; the agents that had the theory of mind enabled benefited greatly as they were able to reason about their own interests, the interests of the other party, and to communicate them effectively.

The processes of mentalizing support the understanding of bargaining behavior as caused by mental states. It is effective mentalizing that facilitates perspective taking and empathy with another, as well cognitive and affective understanding of our own mental states, which then facilitates rational interest-based action. When social cognition fails, the above processes are disrupted.

Ineffective mentalizing and negotiation processes

We seem to be rather ineffective negotiators, which incurs significant personal and social costs, wastes society’s resources, productivity and creative opportunity, and increases society’s conflict and self-destructiveness (Pruitt & Rubin, 1986; Raiffa, 1982). Irrational failures of conflict resolutions are common, even in cases where there is little uncertainty about the consequences of impasses such as divorces, industrial actions, international conflicts and litigation (Ross, 1995). For example, a meta-study of 32 experiments showed that fewer than 4% of negotiators reach optimal (win-win) outcomes. The incidence of worst (lose-lose) outcomes in compatible issues can be as high as 40% (Thompson & Hrebec, 1996). Ineffectiveness is not limited to naïve negotiators; it affects expert negotiators equally (Neale & Northcraft, 1986).

The above studies show that negotiators do not achieve the outcomes they could, even though these are possible and presumably preferred by all parties involved. A part of the reason is that negotiation takes place under the condition of information asymmetry where the parties do not know each other's intentions, preferences, interests and reservation values. Information asymmetry in mixed-motive interactions – involving a 'mixture of mutual dependence and conflict, of partnership and competition' (Schelling, 1980, p. 89) leads to the tension between creating and claiming discussed above (*Tension between creating and claiming value*, p. 56). The harmful effects of this phenomenon come in gradients. At the extreme end, aggressive pursuits of claiming frustrate negotiations with a positive bargaining range and as a result, an agreement that should take place never does. More likely is the scaled-down variant where the parties refrain from fully engaging in creating moves, justifiably anxious about the risk of exploitation, which results in sharing too little information for the parties to work out what value they can create (Lax & Sebenius, 2006, p. 131).

The information asymmetry and the mixed-motive interdependence exist because the parties' mental states are opaque. Pareto-efficient negotiation outcomes would be a matter of mathematical calculation if all relevant parties' beliefs and motivating mental states were explicit and symmetrically shared. Because that is not the case, the information asymmetry needs to be addressed by interpersonal interaction and mentalistic inferential modelling.

It is important to stress that these dynamics of negotiation present a set of barriers to efficient agreements. Good mentalizing improves the odds of overcoming them. Poor mentalizing, by contrast, increases the odds of misunderstanding the relevant asymmetrically shared bargaining-relevant mental states; these end up being ignored, implicitly assumed to be something they are not, or explicitly misunderstood.

The first chapter explored the ways mentalizing can go awry (see *Phenomenology and neurobiology of poor mentalizing*, p. 32). The next section explores the psychology and neural correlates of poor mentalizing, and ends with a discussion of situational moderators that impair social cognition in bargaining.

Teleological mode

Failures of mentalizing have an impact on negotiation. The *teleological mode* disregards mental causation and understands behavior in terms of goals rather than mental states. In negotiation, teleological meaning-making leaves the generating mental states (interests) out of focus altogether. It may be an efficient way to predict the effects of certain bargaining behavior where the mental states do not matter, for example, the dirty situational bargaining tactics such as personal attacks, creating stressful situations, the good-guy bad-guy routine (Fisher et al., 1991, pp. 140-141). However, teleological reasoning in situations involving motivating mental states (interests) results in positional bargaining that disregards interests (see pp. 70-75).

Teleological reasoning may be the result of an acute breakdown of the capacity to mentalize, or of the parties being oblivious or indifferent to the need to mentalize, or being intentionally obtuse. The mental states are obscured, not noticed or considered unimportant and conflated with behavior (e.g., the other side is suing *to win* in court, making an offer in order *to get us to accept* a bad deal). The behavior that accompanies concrete understanding is often angry, over-reactive, blaming, and prescriptive.

Patton (2005) provides an example of a small local magazine that ran a damaging article on a political candidate whom they had confused with a convicted criminal with the same name and hometown. The politician filed a libel suit demanding an amount of money that would likely bankrupt the paper. Instead of compensation, the settlement involved a promise of the newspaper to publish a retraction and a glowing front cover biography, which met the damaged party's interest better than money. This would have not been possible if teleological social cognition obscured the generating mental states of the plaintiff: the desire to repair reputation and the belief that this could be best achieved by winning in court.

Pseudomentalizing

Unlike teleological reasoning, *pseudomentalizing modes* of social cognition do appreciate mental states as causes of human behavior. However the inferred mental states do not correspond to reality in any meaningful way. They are either felt to be reality (psychic equivalence), or they are detached from it (hypermentalizing). For example, Abe calls a long-standing business customer Bella and asks for a one-day extension on the delivery of goods due under a contract. Bella politely refuses and quickly hangs up. Pseudomentalizing, Abe thinks Bella is ‘out to get him’ (psychic equivalence) or, as an alternative scenario, that Bella refused because she was upset that Abe disagreed with her opinion a while ago that caused Bella, in Abe’s opinion, some embarrassment (hypermentalizing). This is not an instance of teleology because Abe does use mental states to explain Bella’s behavior. However, Abe is probably pseudomentalizing because his attributions are likely incorrect and more indicative of Abe’s mental states than Bella’s.

Psychic equivalence

If a negotiator operates in the *psychic equivalence mode*, mental states are felt to correspond to reality perfectly. In the negotiator’s mind, the inferred mental states are not representations of the counterparty’s mind, but reality itself (see pp. 34-37). In negotiation, psychic equivalence means the focal negotiator’s understanding of the counterparty is not based on genuine thinking about probable states of the counterparty’s mind, but on implicit assumptions derived from the mental states of the focal negotiator. For example, an anxious negotiator may attribute her anxiety to be caused by the counterparty’s pseudomentalized intentions that she feels are reality (‘if I am anxious, it’s for a reason: you are out to get me’). This is particularly likely when the counterparty’s plausible alternative mental states feel incompatible with the emotional charge of the situation. In bitter disputes, for example, it is difficult to imagine anything vulnerable as a motivating mental state of the counterparty (e.g., in divorces, high child support and alimony demands are unlikely to be seen as anything else but motivated by vengeance, ignoring the possibility of other contributing interests such as financial security, respect or recognition of the spouse's contribution during the marriage and the desire to provide for the child). The mixed-motive nature of bargaining makes negotiators prone to mis-ascribing adversarial intentions to the counterparty as well as unwilling to test the implicit attitude of epistemic hypervigilance (e.g., 'strangers are not to be trusted'), which despite good evolutionary reasons may not be entirely adequate in modern bargaining situations.

The zero-sum assumption – a bias that value in negotiation is a fixed-sum, implying an exclusively distributional nature of bargaining (Bazerman & Neale, 1992) – and the related incompatibility bias – the tendency to interpret interests of the counterparty as a logical negative of one's own interests (Thompson & Hastie, 1990) – both have roots in psychic equivalence. An illustrative example is the statement congressman Spence reportedly made during the Strategic Arms Limitation Talks (SALT) negotiation: 'I have had a philosophy for some time in regard to SALT, and it goes like this: the Russians will not accept a SALT treaty that is not in their best interest, and it seems to me that if it is in their best interest, it can't be in our best interest' (Bazerman & Neale, 1992, p. 19). The folly of this statement is not in (probably adequately) appreciating the gravity of the conflict between the superpowers, but in the complete failure to conceive of a scenario in which a reduction of nuclear warheads – which haven't laid waste to the world on numerous occasions only by sheer luck – may be in *everybody's* interest.

This is not uncommon. Thompson and Hastie (1990) in their elegant research on the incompatibility bias showed that in situations where there is nothing to negotiate (i.e. the parties have identical interests) up to 40% of bargaining dyads fail to reach the mutually preferred outcome and, even when the two sides do reach such optimal agreement, in more than half of cases neither party realizes that the other party has also benefitted (i.e. they assumed they themselves had 'beat the other party' on the bargaining issues).

Similarly, psychic equivalence possibly contributes to the tendency of negotiators to revise downwards their assessments of proposal desirability once they learn that the author of the proposal is the counterparty, a phenomenon known as reactive devaluation (e.g., both Israeli Jews and Israeli Arabs similarly devalued a Palestinian plan when it was ascribed to the "other side" (Maoz, Ward, Katz, & Ross, 2002)).

The key tenets of naïve realism – the rigidity of one's own convictions and the perception of exaggerated bias in others due to their cognitive errors and ulterior motivations (Pronin, Gilovich, & Ross, 2004) – are based on the psychic equivalence between the mind and reality (Bateman & Fonagy, 2012). Such pseudomentalizing underlies the intractability of conflict that is caused by the mis-mentalized (invented) intentions of the parties rooted in one's own mental states, rather than the counterparty's. A naïve realist misattributes behavior to the values (or lack thereof) held by the observed actors under the psychic equivalence assumption that beliefs are universal (e.g., 'I would never behave that way because I care about gender equality; if you do this, it must mean you don't'). There is research demonstrating that egocentrism drives misunderstanding in negotiation, linking the biases discussed above (Chambers & De Dreu, 2013). The important implication is that these biases are caused by a relapse to a psychic equivalence mode of social cognition that predates mature representational mentalizing. We discuss naïve realism and how it is linked to poor mentalizing in detail below.

Pretend mode

The *pretend mode of mentalizing* involves representations that are disconnected from the cognitive and affective reality. People in pretend mode are preoccupied with generating numerous affectively worthless hypotheses about one's own and others' minds (see pp. 34-37). Negotiators operating in the pretend mode imagine numerous plausible reasons and causes of behavior, but these narratives lack concrete explanatory and predictive power. While some meaningless negotiation spiel is the usual part of the initial stages of bargaining, excessive use leaves the parties feeling misunderstood and may result in the inability to explicate the underlying interests. A less acute version of pretend hypermentalizing may interfere with prioritization of interests and impair value creation through logrolling and dovetailing differences in forecasts, risk attitudes and time preferences. At the level of contractual commitments, pretend mode cognitions underlie overregulation in contracts.

In summary, while teleological failures in bargaining overlook mental causation, the pseudomentalizing modes ascribe mental states incorrectly. In the psychic equivalence mode, negotiators assume that their mind perfectly reflects reality and thus conflate their own mental states and external reality, part of which are minds of other people. Pretend mode, on the other hand, involves inconsequential ruminations about mental states which are disassociated from the cognitive and affective reality of the negotiators.

Naïve realism

Three tenets of naïve realism

People decide how to act based on mental representations of a reality they construct using prior knowledge and situational input, rather than on direct access to objective reality (Griffin & Ross, 1991; Lieberman, 2005; Ross & Nisbett, 1991). Perception of inherently ambiguous social situations is constructive and affected by goals, needs and expectations (Bruner, 1957), for example temporarily (Förster & Liberman, 2007; Higgins, Rholes, & Jones, 1977) and chronically accessible constructs (Bargh, Bond, Lombardi, & Tota, 1986; Bargh & Thein, 1985; Higgins, 1996), exposure to material objects (Kay & Ross, 2003; Kay, Wheeler, Bargh, & Ross, 2004), scripts, schemas and knowledge structures (Ross & Nisbett, 1991), among others. Although actively construed, social perception is typically not felt as self-generated, but as a direct experience of reality. For example, we experience an aggressive person, not a person we decided is aggressive (Hastorf & Cantril, 1954). This contributes to obdurateness of bias if there is any, because a review by controlled cognition (Evans & Stanovich, 2013) does not feel necessary. Maladaptive consequences, e.g., the fundamental attribution error, have been explored under the heading of *naïve realism* in the context of social conflict and misunderstanding (Griffin & Ross, 1991; Robinson, Keltner, Ward, & Ross, 1995; Ross, 1995; Ward, Ross, Reed, Turiel, & Brown, 1997).

Naïve realism theory essentially states that we cannot divorce ourselves from our own take on reality. There are no impartial 'facts.' Data do not have a logic of their own that results in identical perceptions and cognitions for all people. Instead, people perceive and interpret (construe) the information available to sensory perception and as prior knowledge in terms of the perceiver's own personality structure, previous experience, conscious and unconscious needs, thoughts and fears (Griffin & Ross, 1991; Ward et al., 1997). Modern accounts in computational neuroscience show that much of what we take to be perception is in fact the result of selective sampling of the sensorium in line with prior expectations (Friston, 2005; Friston et al., 2013; Friston et al., 2014).

The first two key tenets of naïve realism are, *first*, that subjective construal matters greatly and, *second*, that the subjective belief in the veracity of one's own construal causes people to 'fail to make adequate allowance for the variability and/or impact of subjective construal' (Ward et al., 1997, p. 108). The evidence for these two tenets of naïve realism includes the 'false consensus effect' (Botvin, Botvin, Baker, Dusenbury, & Goldberg, 1992; Gilovich, 1990, 2008), the musical tapping study (Ward et al., 1997) and the belief that others' beliefs will converge toward one's own ('belief in a favourable future'; Rogers, Moore, & Norton, 2017), among others.

The final tenet of Ward and Ross's (1997) naïve realism theory is *social hostility*. A consequence of the first two tenets, the social hostility principle suggests that people have a (commonly unquestioned) assumption that others share their subjective perception, judgment and experience. Because a naïve realist feels they see the world 'as it really is', all other people must see it the same way and consequently believe and think the way the naïve realist believes and thinks. If they fail to do so, this can only be because they are ignorant (uninformed), lazy, intellectually inadequate, or – what in the context of dispute resolution proves to be by far the most damaging attribution – that they are biased by self-interest, ideology, or idiosyncratic personality characteristics (Ward et al., 1997, p. 116). While in such thinking the naïve realist is quite correct about other people's bias – at least of its existence, perhaps not of its magnitude – they remain remains blissfully unaware of the presence of the same bias in themself. This third tenet, also called the *fundamental attribution error*, has wide empirical support, such as the classic study by Hastorf and Cantril (1954) that shows the diverging emotional partisan perceptions of a football game, and its replications (e.g., Griffin & Ross, 1991; Keltner & Robinson, 1993; Lord, Ross, & Lepper, 1979; Pronin et al., 2004; Robinson et al., 1995; Ross, 2014, among others). In conflict, naïve realism underpins partisan perceptions of fairness and equity (e.g., Diekmann, Samuels, Ross, & Bazerman, 1997).

Naïve realism is also at the core of the *false polarization effect*, an attribution error fueling the subjective appearance of the width of the gulf between the participants of opposing ideological camps. For example, the participants who think of themselves as pro-life and pro-choice overestimated the magnitude of disagreement between themselves and their ideological opponents and thought 'the other side' was considerably more extreme in their views than they really were (Keltner & Robinson, 1993, 1996, 1997; Robinson et al., 1995). Finally, naïve realism underpins the paradoxical phenomenon of *reactive devaluation*, where a negotiation proposal decreases in value to the recipient simply because an opponent (rather than the focal negotiator) has made such proposal (Maoz et al., 2002; Ross, 1995, 2014; Ross & Ward, 1995).

Naïve realism is a consequence of poor mentalizing

A theoretical proposition developed and tested in this thesis is that naïve realism in negotiation – like may reasoning biases (e.g., the hindsight bias; Wertz & German, 2007) – can be seen as a consequence of poor mentalizing. Developmentally, young children under the age of about four possess excessive confidence in the objectivity of their subjective experience of the world (see *Development of mentalizing*, p. 28). They do not feel that whatever they are experiencing is a representation, or an affective reaction to such a representation, that only *corresponds to reality* to a smaller or larger degree. Instead, they feel that what they experience *is* reality itself (e.g., for Calvin in the cartoon on p. 35, the monsters really *are* under the bed). A corollary of this is that because what they experience is reality itself, this experience must be identical for all people (as all people obviously experience the same reality). Any and all knowledge must then be shared; what I know, you must know also, and *vice versa* (Birch & Bloom, 2007; Fonagy et al., 2007; Fonagy & Target, 2007). It is only later in development, once mentalizing matures, that children are able to disassociate beliefs from reality and consequently represent false beliefs ('Sally-Anne test'; Baron-Cohen, 1985). They can then begin to appreciate that their subjective experience is based on representations of reality rather than on a direct access to the reality, that such representations are unique to each person, and that they may be correct or false to various degrees. In this context, Fonagy and Luyten (2009) suggest that naïve realism may be seen as a consequence of poor mentalistic inference, more precisely from the sense of 'oneness' that stems from an undifferentiated experience of shared knowledge, rather than from the overconfidence in an individual's own knowledge (p. 1363). Such reemergence of a primitive mode of mentalizing is commonly associated with severe personality disorders such as the BPD but is not uncommon to occur temporarily in non-clinical populations, particularly under conditions of stress and excessive cognitive load.

Situational inhibitors of mentalizing in negotiation

The quality of mentalizing depends on both dispositional and situational factors. There is marked variance in the reflective functioning even between and within people who suffer no marked mental health impairments. Individual mentalizing profiles reflect differences in functioning with respect to each mentalizing polarity and studies show these differences may have roots in epigenetic influences and developmental vicissitudes (Luyten et al., 2012).

Mentalizing is also highly context dependent. Dispositional factors – the individual's mentalizing profile – can be thought of as a set of strengths and vulnerabilities of effective mentalizing depending on situational pressures. Two facets of negotiation are particularly unconducive to mentalizing: complexity and subjective relevance.

Complexity of negotiation and cognitive load

Negotiation situations are highly ambiguous and uncertain (De Dreu, Beersma, Steinel, & Van Kleef, 2007). The parties do not know each other's interests nor the overall payoff structure. The interaction is complex because of the opacity of the mental states and the ambiguity of negotiation behavior. As outlined above (*Tension between creating and claiming value,* p. 56), information exchange is necessary to generate value (Pruitt & Rubin, 1986; Raiffa, 1982), but highly problematic because of the tension between creating and claiming value (the negotiator's dilemma): it is risky to both provide accurate information or rely on the information provided by the counterparty (Lax & Sebenius, 1986b; Mnookin, 2000; Murnighan et al., 1999). The key reason for the obdurateness of the negotiator's dilemma is the inability to mentalize the counterparty's bargaining behavior and infer their collaborative (value creating) or competitive (value claiming) intent with any degree of certainty because the parties' behaviors are markedly ambiguous. In other words, when a party intends to deceive, that party is highly motivated to hide her intent, and the behaviors are, form the perspective of the other side, indistinguishable from the behaviors that create value. We are incapable of recognizing deception based on non-verbal behavior (Bond & DePaulo, 2006; Ekman, 2009; Etcoff, Ekman, Magee, & Frank, 2000) and the counterparty's statement that she cannot pay more than a certain number will appear exactly the same when stated with either intention: it can be motivated by an intention to honestly inform of a real constraint and facilitate a deal, or to mislead about the walk away point to claim value. We discuss this extensively later in the thesis.

In addition to the information that calls for social cognitive computation, non-social cognition (reasoning) is also in high demand in negotiation. Checking the quality of commitment and the value of the deal on the table, tracking time and possible deadlines, discussions with constituencies, generating options to satisfy the parties interests (e.g., if one party wants fresh air and the other no draft) are some of examples of such demands. Some of them can be mitigated through careful advance preparation, but many persist during the negotiation.

A high level of cognitive load has been shown to impact explicit mentalizing (Spunt & Lieberman, 2013) as well as the subjective experience of empathy (Morelli & Lieberman, 2013). The implication is that the situations involving significant cognitive demands present a risk that controlled social cognition may be replaced with automatic mentalizing, possibly resulting in pseudomentalizing or teleological interpretation of action. Basically, if the demands on the cognitive apparatus are too extensive, the sophisticated social cognition gives way to automatic processes, vulnerable to error.

Ideally, these would be available for consideration through controlled and balanced mentalizing. If that is prevented due to excessive cognitive load or for other reasons (e.g., stress and switch, below), implicit mentalizing does the inferring, which subjects the negotiators to the risk of bias. More importantly, as we will develop later in the thesis in more details, the zero-sum assumption and the incompatibility bias (Bazerman & Neale, 1986, 1992; Thompson, 1991; Thompson & Hastie, 1990; Thompson & Hrebec, 1996) are also likely to operate as heuristics, as well as the expectations of competition (O'Connor & Adams, 1999) and excessive epistemic vigilance (Sperber et al., 2010), manifesting themselves in common-sense wisdoms such as ‘don’t trust strangers’ and ‘better safe than sorry’.

The more complex the negotiation, the more it calls for explicit cognitive efforts, including controlled mentalizing. At the same time, complexity increases cognitive load, which reduces the controlled component of mentalizing. The default implicit mentalizing heuristics are likely to result in suboptimal, risk-averse bargaining, and zero-sum and incompatibility assumptions (Thompson, 1991; Thompson & Hastie, 1990; Thompson & Hrebec, 1996). In addition, and as indicated above (p. 39), automatic mentalizing can manifest itself as hypermentalizing, which paradoxically increases the cognitive load as the focal negotiator needs to deal with an excess of possible inferences about mental states of the counterparty, all of them unmoored from reality. This leads to a failure to hold a stable and plausible image of the counterparty’s mind, and instead to endless elaboration of increasingly complex models of what might be motivating the counterparty and what the counterparty’s intentions might be.

Arousal and the ‘stress-and-switch’ paradox

To the extent that the situation involves interests that are important to the parties, negotiating is bound to increase the levels of negotiators’ physiological arousal. Arousal involves the activation of the central and the peripheral nervous systems, accompanied by changes in the heart rate and blood pressure, skin conductivity and muscle tonus, as well as less visible but complex changes in the functioning of the brain that have an adverse impact on higher-level cognition.

Yerkes and Dodson’s (1908) model of the impact of arousal on task-performance predicts that the performance in higher-complexity tasks such as negotiation is vulnerable to increases of arousal. Similarly, explicit mentalistic inference, driven by neocortical brain networks, is at risk if arousal reaches the biobehavioral switch (Luyten & Fonagy, 2015). The implicit cognitions that replace explicit social cognitive reasoning consist mainly of fight-or-flight and vigilance functions (Mayes, 2000, 2006), hardly helpful in important negotiations.

Negotiators may be facing a paradoxical situation: the more important the negotiation, the more it calls for reflective, explicit and balanced mentalizing; at the same time, such a situation – particularly if not going well – is likely to increase arousal levels. If the stress reaches the biobehavioral switch-point, the areas of the prefrontal cortex that are critical for controlled thinking get turned off. That shuts down the explicit, balanced mentalizing and implicit modes take over, presenting a risk of un-auditable teleological reasoning and pseudomentalizing, as well as a reduction of the capacity for problem solving and for sustained attention (Hockey, 1970). In other words, the more important the negotiation, the more likely the increased arousal results in shutting down the processes of social cognition that the situation needs most.

At the same time, as outlined above (p. 37), arousal may trigger a snowball effect because the impairments of mentalizing – the reduction in the involvement of explicit social inference – increases the likelihood of hypermentalizing or psychic equivalence errors. The resulting models of the counterparty are unmoored from reality or colored by projections, which further increases arousal, which impairs mentalizing even more, and so on and on.

Conclusion

Because of the strategic interdependence and its mixed-motive nature, bargaining involves simultaneous cooperation and competition. In these, mentalizing is always present, even if it goes unnoticed. The processes may be implicit, but are taking place nonetheless. Like seeing the fighting triangles as having intentions, beliefs and desires (Heider & Simmel, 1944), we cannot help but automatically ascribe mental states to our bargaining counterparts. Equally, our own behaviors are based on our implicit mental states as much as our explicit ones.

The theory of bargaining underpinned by mentalizing offers explicit modelling of how negotiators consider mental causation in negotiation. It describes the focal negotiator's understanding and predicting of the counterparty's behavior in terms of their (inferred) mental states.

Interests, the motivating mental states of negotiators, are a paragon of something that is observationally opaque and needs to be inferred. Value creating depends on correctly understanding the interests of the negotiating parties, and value claiming depends on correctly estimating the counterparty's valuations and manipulating their beliefs about what is possible. This results in two hypotheses that this thesis tests in the empirical studies that follow in subsequent chapters: the quality of mentalizing should correlate with both effective value creating and value claiming.

The theoretical proposal developed in the first three chapters also provides insight into the cognitive intrapersonal and interpersonal processes in bargaining from the perspectives of neuroscience and cognitive psychology. We know a lot about impaired mentalizing and how it disrupts cognitions. Poor social inference involves reemergence of primitive, developmentally early pre-mentalizing modes of social cognition. Applying these to bargaining provides insights into the nature of impairments of negotiation-relevant cognitions as well as an explanation of bargaining phenomena such as the zero-sum assumption, incompatibility bias, reactive devaluation and naïve realism (Curhan et al., 2004; Maoz et al., 2002; Pronin et al., 2004; Robinson et al., 1995; Ross, 1995, 2014; Ross & Ward, 1995; Ward et al., 1997). Poor mentalizing begets poor interest-based bargaining and inefficient outcomes.

Perhaps most importantly, dramatic losses of mentalizing in the non-clinical population tend to be highly dependent on situational moderators such as complexity and stress. This may give rise to a paradoxical situation where the complexity and importance of the negotiation create levels of arousal and cognitive load that biobehaviorally inhibit the social cognitive functioning they demand (the cognitive load and the 'stress and switch' paradox).

Present studies

Existing Research

We conducted a comprehensive search to identify the existing literature linked to our thesis. We used the *APA database* (Journals, PsycArticles, PsycBooks, PsycExtra, PsycInfo), *Web of Science* and *ProQuest* (Journals, Conference Papers & Proceedings, Dissertations & Theses and Wire Feeds).

We first searched using the terms *negotiat**, *bargaining*, *bargainer* and *dispute* in the title and keyword fields to identify negotiation and mediation studies. We combined that with a search for mentalizing and related constructs, using *mentaliz**, *social cognition*, *perspective taking, empathy*, *theory of mind*, and *emotion recognition* as search terms in the titles, subject terms, and keywords of the papers.

The search returned 136 results in total. We included the papers in our review if they focused on interpersonal negotiations (not an ultimatum or dictator game, trust game, any kind of game theoretic task, group negotiation or an organization-level negotiation). Further, they have had to examine mentalizing or related constructs (perspective taking, empathy or theory of mind) as predictors of either the negotiation process or the outcomes in negotiation. After applying the above inclusion criteria, 53 articles did not focus on negotiation or mediation or negotiation outcomes (e.g., they dealt with mentalizing in therapy or development), 30 were duplicates, in 9 the designs did not test whether the relevant mentalizing-related social cognitive capacities were directly associated with the outcomes, six featured game theoretic tasks rather than negotiation tasks, four were prescriptive and descriptive magazine articles or book chapters, for two we were not able to obtain a full text and one was a book review. In the end, 32 papers were relevant.

We will discuss these papers in four sections below. First, we will present existing studies and theoretical papers that consider mentalizing-related variables in negotiation, namely perspective taking, empathy and emotion recognition. We will provide an overview of the field and focus in detail on the studies that are most relevant for the purposes of our thesis. In the second section, we will discuss the three papers that have considered mentalizing in the context of negotiation and discuss how our research advances and extends our understanding of social cognition in negotiation. Third, we will continue with a brief outline of the papers that considered the theory of mind in negotiation involving artificial intelligence agents. And fourth, we conclude with an overall discussion that will lead to the presentation of the studies in this thesis.

Perspective taking and empathy studies

The majority of existing literature investigates the impact on negotiation outcomes of two interpersonal capacities associated with mentalizing: perspective taking and empathy. Perspective taking accounts for much of this research and the consensus seems to be that it facilitates bargaining outcomes. Perspective taking has been found to enhance joint gains (Kemp & Smith, 1994), to minimize incidences of impasse and mitigate negotiators' egoism impediments in integrative contexts (Trötschel, Hüffmeier, Loschelder, Schwartz, & Gollwitzer, 2011), to enhance logrolling (Moran & Ritov, 2000), to benefit parties in negotiations under conventional and final-offer arbitration (Neale & Bazerman, 1983), and to encourage compromise (Marietta et al., 2013). It also assists law enforcement officers in dealing with difficult, dangerous, and disordered individuals (Vecchi, Wong, Wong, & Markey, 2019), and helps negotiators in existing relationships reach agreement (Ramirez-Fernandez, Ramirez-Marin, & Munduate, 2018).

Taking perspective assists value claiming, too. It helps negotiators resist the anchoring effect of first offers (Galinsky & Mussweiler, 2001) and negotiators who take cultural perspectives are better at distributive aspects of bargaining (S. Lee, Adair, & Seo, 2013). Epley, Caruso, and Bazerman (2006) found that considering others' thoughts and perspectives increases egoistic behavior such that people actually take more of the available resources. The same was found by Galinsky et al. (2008) and Gilin et al. (2013), studies we discuss in detail below.

Białaszek, Bakun, McGoun, and Zielonka (2016) warned that perspective taking may come at a cost: imagining the counterparty's perspective can inadvertently backfire as the imagined perspective might be more competitive due to misattribution. Also, Gehlbach et al. (2015) investigated whether different ways to put one ‘in someone else’s shoes’ cause different outcomes in negotiation. In a computer-based simulation participants took perspective in five different ways ranging from simple information revelation to experiential perspective taking, finding that more immersive perspective taking leads to feeling more positive about the relationship and making greater concessions during the negotiation.

Regarding empathy, Rahim et al. (2002) found that it correlates with motivation, which in turn positively correlates with problem solving and negatively with competitive bargaining. Empathy also mediates the relationship between anger and deception (Yip & Schweitzer, 2016), but does not seem to have a role in facilitating cooperation in virtual negotiation (Marchi, Targi, Liston, & Parlangeli, 2020).

A series of authors considered perspective taking and empathy concurrently. Ku and Galinsky (2006) found that, in a competitive interaction, empathy made one's negotiation partner feel more satisfied with how they were treated. However, only cognitive perspective-taking facilitated the understanding of the other side’s interests and the development of integrative and creative options. Both perspective taking and empathy were found to mediate conflict (Betancourt, 2004). In two studies, Cohen (2010) found that empathy, but not perspective taking, discouraged misrepresentation, inappropriate information gathering and feigning emotions to manipulate counterparties, suggesting that unethical bargaining might be better deterred by empathic intervention rather than perspective taking.

This brings us to the two leading sets of studies in the field that we will discuss in detail. In the first set of studies, Galinsky et al. (2008) investigated to what extent these two distinct social capacities, perspective taking and empathy, predict success in strategic social interactions (negotiations). The hypothesis was that perspective taking would be more useful than empathy because ‘securing economically efficient outcomes, cognitive appreciation of another person’s interests is more important than an emotional connection with that person.’ In the first study, 70 MBA students were tested using the IRI (Davis, 1983) and represented either a buyer or a seller in a negotiation of a gas station sale (Texoil), where no deal was possible unless the negotiators uncovered joint underlying interests and used creative options (lower price for the station and a management position for the seller). More than two thirds (68.6%)

of the dyads concluded the agreement. Logistic regression showed that perspective taking (but not empathy) predicted whether a successful deal was reached both on the dyadic level with marginal significance (p = .050) and on the individual level (p = .023). In the second study, 152 MBA students negotiated the same task in two experimental groups. The empathy condition participants received special instructions to 'try to understand what [the counterparties] are feeling, what emotions they may be experiencing'. In the perspective taking condition participants were instructed to 'try to understand what [the counterparties] are thinking, what their interests and purposes are.' The proportion of successful deals varied as a function of the experimental condition (p = .03). The dyads with a perspective-taking buyer achieved a deal more frequently (76%) than the control dyads (39%, p = .01). The empathizers, however, were not more successful in deal making than the controls (p = .30). In the final study of the paper, 146 MBA students were primed in the same way as in the second study but negotiated an 8-issue negotiation paradigm (the New Recruit). On the dyadic level, the experimental condition was predictive (p = .014): perspective taking condition dyads reached higher joint gains than dyads in the control condition. On the individual level, perspective taking recruiters also achieved higher individual gains than controls. In the candidates, there was no difference between the conditions' individual performance. However, the candidates achieved higher gains than controls when they negotiated with both empathizing recruiters and perspective taking recruiters than controls. The overall conclusion was that in strategic social interactions the relative benefits of perspective taking outstrip the benefits of empathy.

The follow-up set of experiments (Gilin et al., 2013) explicitly differentiated between cognitive perspective taking and empathic social competency, and tested how these two capacities contribute to success in strategic interactions. The prediction was that they are 'likely to be differentially useful in competitive tasks based on the task–competency match' (p. 4). In other words, where cognitive appraisal of strategic concerns and motivations is important, perspective taking ought to help, and where the task calls for emotional sensitivity, empathy should. In the first study, undergraduate students (N = 90) were assessed with the IRI (Davis, 1983) before participating in a 10-round, computer-based specially designed Disarmament Game. In each round, the players privately and simultaneously decided how many bombs they disarmed and/or whether they attacked the counterparty. If any one party attacked, the game terminated. Alternatively, if no-one attacked for 10 rounds, the benefits were calculated based on the number of bombs disarmed (joint benefit). There were two individually lucrative strategies that could be derived from the rules: a competitive ‘refuse to disarm and attack’ and a collaborative ‘disarm without attack’. Which strategy was optimal depended on the choices of both players (interdependence). Only perspective taking related to players net distributive gains (p = .04). At the dyadic level, joint gain was predicted by dyadic perspective taking (p = .03). Dyadic empathy was marginally negatively predictive of such gains (p = .05). In the second study, 135 undergraduate students, again assessed with the IRI, played a social coalition game. They were grouped in triads and allowed a brief meeting before the game started to conduct small talk. After that, each participant had to decide privately which of the other two participants they wished to match with. If two participants in a dyad identified each other, the match was declared, and they had a chance at winning a cash prize. In three quarters (75.6%) of the groups a successful match took place. Matching was significantly associated with empathy (*p* = .02). Perspective taking was not significant and a contrast hypothesis test in HLM indicated the two effects were not significantly different from each other (p > .05). The authors suggest that the overall analysis confirms that empathy is a stronger predictor of success than perspective taking. In the third study, 84 undergraduates engaged in the same coalition game as in the second study and received ‘performance tips’ priming them into empathy or perspective taking (experimental conditions), or not (control condition). HLM analysis showed that empathizers matched about twice as frequently as perspective takers (*p* = .02). The authors concluded that perspective-taking and empathy are conceptually distinct and differentially useful in mixed-motive negotiation tasks. In the final study, 75 students participated in a Ultimatum Game and had to choose their partner based on appeals they received. The appeals were either cognitive or affective, and the participants were primed the same way as in study 3, above. Consistent with the studies above, perspective taking primes were found to be more effective in cognitive appeals and empathy primes in affective appeals (*p* = .025). The overall conclusion was that perspective taking and empathy both promote understanding that can be helpful in negotiation tasks, but only when ‘the underlying structure or content of the task requires that particular social competency' (p. 13). In other words, different competencies are called for by different types of tasks.

The research presented above investigates perspective taking and empathy as two conceptually distinct capacities that assist negotiation outcomes because they allow understanding the counterparty and establishing an emotional connection with the counterparty, respectively. We have asserted that mentalizing – a complex social cognitive capacity to infer opaque minds from observable data – underpins virtually all decision-making in negotiation and consequently determines negotiation outcomes. Understanding the counterparty – cognitively (perspective taking) or affectively (empathy) – forms part of such decision making and is underpinned by mentalizing. In other words, both perspective taking and empathic responses are underpinned by social cognitive (mentalistic) inference. For example, the counterparty's perspective is taken by explicitly and cognitively inferring the mental states of the other based on observable data (i.e., it corresponds to other-focused, cognitive and explicit mentalizing). Mentalizing generates the taking of perspective and empathic responses.

The findings of the reported studies are therefore consistent with the predictions of our model. To the extent mentalizing facilitates perspective taking and empathic responses, we should expect to see the correlations between these two sub-capacities and negotiation outcomes to also manifest themselves for mentalizing. We are also not surprised by the differential value of empathy and perspective taking in different tasks; negotiators who can mentalize effectively can adaptively use the type of social cognitive inference the situation requires. Finally, while critically evaluating these studies is beyond the purpose of this review, it is worth noting that the two key sets of studies reported above are likely to be on the verge of being underpowered (based on the post-hoc power analysis considering the effect sizes and sample size).

Emotion recognition studies

Three studies considered the ability of negotiators to discern the emotions in others as a predictor of negotiation performance. We discuss them in turn.

Elfenbein, Foo, White, Tan, and Aik (2007) had 164 undergraduate students negotiate a four-issue task (containing a distributive, compatible and two integrative issues) and complete a measure of facial emotion recognition accuracy (ERA), developed by the authors. The subjects were split at median into high and low ERA groups. Multiple regression showed that high ERA in sellers predicted greater creation of value ($p < .01$) and a marginal edge in claiming value ($p < .06$). In buyers, no such relationship was found.

In a more recent study, Schlegel, Mehu, van Peer, and Scherer (2018) investigated how general mental ability (GMA) and emotional intelligence (EI) predict objective and relational outcomes in a dyadic negotiation. The findings showed that the GMA of the participants ($N = 130$) was unrelated to negotiation outcomes. Higher scores of the recruiters on some of the EI instruments (the motional understanding branch) were related to higher joint gains ($p < .05$). This effect was more pronounced in female dyads. Second, Higher total EI scores were not related to one's own individual gains as predicted, but to higher individual gains of the counterparty ($p < .05$).

In a follow up study, Schlegel (2021) investigated whether training in emotion recognition can assist negotiation processes and outcomes. In the study ($N = 166$), the relevant hypotheses were that the dyads that received ERA training would reach higher joint gains than dyads in the control condition, that these dyads will distribute the pie of gains more equally, that they would rate themselves and their partners as more cooperative and that they will use less forcing and more problem solving. Trained dyads did not reach significantly higher joint outcomes ($d = .05$), but their gains were more equitably distributed ($d = -.44$), they rated themselves and their counterparties as less competitive ($d = -.42$ and -.47), they used less forcing ($d = -.29$) but not more problem solving ($d = -.01$).

Like the capacity to take perspective and feel an empathic connection with another, the ability to accurately recognize emotions is an aspect of mentalizing (based on other-focused, explicit and affective social cognitive inference). Emotions are mental states that are observationally opaque and must be inferred from the available sensorium. In fact, one of the most recognized tests for assessing mentalizing is the *Reading the mind in the eyes* test (Baron-Cohen, Wheelwright, Hill, Raste, & Plumb, 2001), designed to test the ability of participants to infer the emotion from pictures of eyes.

The findings of these studies are consistent with the predictions of our model. The ability to discern others' emotional states correctly is based on the capacity to mentalize, but (as before) is only a part of it.

Papers considering mentalizing in negotiation

Mentalizing in bargaining

Bernath and Kovacs (2015) investigated the relationship between the need for mentalizing and negotiation outcomes. The participants (*N* = 150) were assessed on their need to mentalize others and then completed a multi-issue negotiation task. The hypotheses were that the need to mentalize would correlate with joint gains, that higher need for mentalizing would go hand in hand with integrative bargaining strategies, and that the individuals with a higher need to mentalize would be better able to learn the payoff structure of the bargaining task. The results showed a weak correlation between one (of four) subscales of the need for mentalizing and integrative gain ($p > .05$), weak marginally significant correlation ($p = .09$) between one subscale (of four) of the need for mentalizing with integrative strategies, and no association between the need for mentalizing and learning of the payoff structure.

While the subject of the study is 'mentalizing', the authors' understanding of these processes is significantly different from ours. Unlike our understanding of mentalizing as a multifaceted social-cognitive capacity (pp. 22-37), the authors' conceptualization is limited to other-focused theory of mind. The predictor they use is the need to mentalize, which does not capture any kind of ability or performance, but instead a *motivation* to consider the minds of others. More importantly, the theory underpinning the hypotheses in the paper is based on the idea that a higher perceived need to mentalize ought to facilitate integrative negotiation because it promotes prosocial behavior (based on research into couples and parent-children relationships). In other words, the paper lacks a comprehensive theoretical framework about how mentalizing, as a social-cognitive capacity, underpins bargaining. Finally, the effect sizes reported are weak and marginally significant, and the number of predictor variables ought to be corrected for multiple comparisons, likely rendering them insignificant.

Mentalizing in mediation and conflict resolution

In an exploratory paper, Howieson and Priddis (2012) suggested that adopting the 'mentalizing stance' might assist mediators facilitating the 'mediation shift' from the clients' entrenched positions towards negotiating constructive solutions to their disputes. They asserted that the parties who held in mind the perspective of their counterparts may be able to communicate more clearly, better understand their own and the other party's behavior, become calmer, less rigid and open to entertaining alternative future possibilities for resolving their conflict.

In a follow up study, Howieson and Priddis (2015) tested this proposition employing a quantitative and qualitative analysis of mediation transcripts and post-mediation surveys. They found that when mediators adopted a mentalizing stance to facilitate the disputing parties to engage their mentalizing capacities and, in particular, their ability to mentalize the child, the mediation process became more constructive and meaningful. The overriding theory behind this finding was that 'in a family mediation context, the mentalizing capacity is the neurobiological enabler that enables the parties to communicate clearly with one another about their interests', 2015, p. 81).

The authors rely on the conceptualization of mentalizing advanced by Fonagy and collaborators (e.g., J. Allen & Fonagy, 2008) and contend that both self- and other-focused mentalizing may assist in mediation by helping people achieve what they 'want'. However, the study stops short of saying that interests – which is what underpins interest-based mediation – are mental states. Conceptually, the papers limits mentalizing to a communication-enhancing tool. Overall, both the theory and findings are fully consistent with our proposition.

Leary (2008), based on her work with guerilla fighters from the Free Aceh Movement and the Republic of Indonesia under the auspices of the Henri Dunant Centre (Leary, 2004) applied mentalizing as a post-hoc frame to 'understand the relational activity in which mediators engaged in order to sponsor a discussion' between guerilla fighters and the Republic of Indonesia. The key claim the authors make is that mentalizing facilitates the recognition and use of what they call 'critical moments', the moments that drive change in dispute resolution. The authors use psychoanalytic concepts to narrate a story about change under uncertain conditions and postulate that mentalizing is relevant for negotiation because it (1) depends upon being open and openminded and helps avoid assuming with too much certainty that we know what others think and believe, (2) is consistent with the findings that seemingly rational arguments are often influenced by emotional factors, (3) promotes an attitude of accepting that we have something to gain from others and that they can help us reevaluate our own point of view, and (4) can be synergistic when people engage in it mutually. Beyond this, the role of mentalizing nor the theory linking mentalizing and bargaining is not elaborated on any further. The authors' above approach is highly similar to the approach of Howieson and Priddis (2012, 2015) who advocate that mentalizing facilitates the communication and understanding necessary for a 'mediation shift', a concept virtually identical to Leary's 'critical moments'.

Theory of mind in artificial intelligence agents

Finally, artificial intelligence bargaining experiments (de Weerd et al., 2013a; De Weerd, Verbrugge, & Verheij, 2013b; Weerd et al., 2017) tested to what extent various orders of theory of mind benefit computer agents in a formal mixed-motives game. They demonstrated limited effectiveness of the first order theory, but a large effect of the second order theory of mind, as it allowed the computer agents to reason about the way they can communicate their interests. Third and fourth order theory of mind did not provide any additional advantage and was in fact counterproductive.

The results of these experiments are consistent with our model. The papers lend credence to our predictions that mentalizing ought to predict outcomes in negotiation with humans.

Conclusion

The studies presented above investigated the impact of aspects of mentalizing – perspective taking, empathy, theory of mind, emotion recognition and the need for mentalizing – on negotiation processes and outcomes. These empirical findings are consistent with the model we are advancing. To the extent mentalizing – social cognitive inference that models mental states based on observable input – underpins and correlates with the taking of perspective, empathic consonance, emotion recognition and theory of mind, we can expect to see the correlations between mentalizing and outcomes.

However, no study so far considered mentalizing as a social cognitive competency that underpins virtually all decisions in negotiation settings nor provided an overarching model of social cognition in bargaining as a single multifaceted capacity that can be employed comprehensively in different negotiation settings.

Our studies investigate the impact of mentalizing, as assessed by a measure of the global capacity to mentalize rather than any of its specific facets (p. 68-90), combined with the capacity to engage explicit cognition (pp. 103-108), on value creating and value claiming in a multi-issue negotiation with value potential, in a distributive task, and in a dispute resolution task predominantly driven by partisan perceptions.

Outline of the studies

The studies in the following three chapters examine and test the relationship between mentalizing and value creating and claiming in three different negotiation tasks. Study 1 investigates the impact of mentalizing (as assessed by the Reflective Functioning Questionnaire; Fonagy et al., 2016), and the ability to engage explicit cognition (as assessed by the Cognitive Reflection Test; Frederick, 2005) on individual and dyadic outcomes in a multi-issue scorable task. Study 2 extends the research to a purely distributive, zero-sum task. Study 3 investigates the impact of these two capacities on the probability of achieving a successful settlement in a partisan-perception driven dispute.

STUDY 1: REFLECTIVE FUNCTIONING AND COGNITIVE REFLECTION IN A MULTI-ISSUE NEGOTIATION TASK

In the previous chapter we developed the theory that mentalizing fundamentally underpins negotiation processes. In this study we test this. Specifically, we investigate the impact of mentalizing (measured by the Reflective Functioning Questionnaire; Fonagy et al., 2016) and the ability to engage explicit cognition (as assessed by the Cognitive Reflection Test; Frederick, 2005) on individual and joint negotiation outcomes.

Reflective functioning and cognitive reflection in negotiation

In the following three sections we present the two key aspects of cognition in negotiation: the reflective functioning, a measure of mentalizing, and the cognitive reflection, the metacognitive ability to suppress automatic responses and override them with deliberate thinking. We conclude it by showing how training affects each of these two capacities in negotiation settings and outline the predictions for this study.

Reflective functioning

This study tests the proposition that mentalizing underpins the key processes and outcomes in negotiation. Mentalizing, or reflective functioning, is a form of social cognition. It is an imaginative mental activity enabling an individual to conceive of self and others as social agents whose thoughts, feelings, desires, and behaviors are based on underlying intentional mental states. It facilitates the appearance-reality distinction by allowing the understanding of beyond face-value behavior. It enhances communication by allowing the speaker to keep in mind the mental state of the audience and equips an individual for both collaborative and competitive existence with others and promotes individual social survival (Fonagy et al., 2004; Fonagy et al., 2007). For details refer to *Understanding behavior in terms of mental states*, p. 22.

The *Reflective Function Manual* (Fonagy, Target, Steele, & Steele, 1998) states:

> *'[Reflective functioning] or mentalization enables children to "read" other people's minds. By attributing mental states to others, children make people's behaviour meaningful and predictable. As children learn to understand other people's behaviour, they can flexibly activate, from the multiple sets of self-other representations they have organised on the basis of prior experience, the one(s) best suited to respond adaptively to particular interpersonal transactions.'*

As explained in detail in earlier chapters *(Mentalizing in negotiation*, pp. 68-77), mentalizing ought to predict value creating through adaptive (intentional-stance) thinking and interpersonal behavior required to move beyond the generic non-mentalizing assumptions about self and others, and engage in integrative, interest-based negotiation (p. 70). Because the most effective way to mentalize in negotiation is at the level of negotiating unit, dyadic mentalizing should be a strong predictor of joint performance. Finally, we expect that mentalizing at the individual level will predict value claiming, as virtually all value claiming tactics involve 'shaping perceptions' of the other side, and the capacity to effectively infer mental states ought to improve performance (for details see p. 75). A difference in the mentalizing capacity between the negotiators in a dyad ought to be a particularly strong predictor of value claiming, but also an inhibitor of value creating.

Cognitive reflection

Uncertain and ambiguous negotiation situations require that the largely automatic mentalizing (Lieberman, 2007; Luyten et al., 2012) is complemented by controlled inference. The ability to detect that a controlled cognitive check is necessary and to carry it out is a critical aspect of rational negotiation.

The behavioral decision perspective to negotiation views the negotiators' inability to maximize outcomes as the result of biased decision-making. While negotiations normally carry the potential to create value beyond what is immediately obvious (Deutsch, 1973; Pruitt & Rubin, 1986; Walton & McKersie, 1965), negotiators assume the contrary. As noted earlier, the critical bias identified by the decision perspective research is that the value in negotiation is fixed (the 'fixed-pie' or 'zero-sum' assumption), which leads the parties to focus on the competitive distributive aspects of the interaction and to leave value creation out of focus (see e.g., Bazerman et al., 1985; Bazerman & Neale, 1986, 1991; De Dreu, Koole, et al., 2000; Menkel-Meadow, 1983; Neale & Bazerman, 1992b; O'Connor & Adams, 1999; Pinkley et al., 1995; Thompson, 1991). A related bias is that negotiators' interests cannot be compatible (i.e., if I want something, the counterparty will oppose that) and causes lose-lose agreements where the parties both prefer one option, but settle for another (Thompson & Hrebec, 1996). For details about these two biases please refer to *Negotiator cognition and irrationality* (pp. 63-65).

How is the ability to engage explicit cognition relevant for our research? Forty years of the 'heuristics and biases' line of research documented systematic violations of rationality in a wide array of individual thinking tasks (e.g., incorrect probability assessments, faulty hypothesis testing, context dependency, framing; Kahneman, 2011). This research also identified a thinking disposition ('cognitive reflection'; Frederick, 2005) that enables people to detect that the automatic responses may be faulty and that a controlled cognitive check, and potentially override, is necessary. This capacity has been found to provide a measure of protection against bias in a wide array of thinking tasks (for an overview see, e.g., Stanovich, West, & Toplak, 2016).

To generate value by trading (i.e., 'logrolling' to capture 'integrative gain'; Froman & Cohen, 1970) or identifying jointly-preferred options ('compatible gain'; Thompson & Hrebec, 1996) negotiators need to focus on and utilize the correct aspects of not only their own, but also their counterparties' payoffs, which they do not know. The information exchange is fraught with risk and often obfuscated by competitive tactics (Lax & Sebenius, 1986b; Mnookin, 2000; Murnighan et al., 1999), and the imparted information is difficult to verify (Bond, 2008; Bond & DePaulo, 2006; Depaulo et al., 2003; Hartwig & Bond, 2014; Sporer & Schwandt, 2007). The fixed-pie bias in such ambiguous and uncertain ('fuzzy'; De Dreu et al., 2007) situations focuses the negotiators' attention and efforts squarely on the competitive distributive tactics, thus depressing joint value.

We predict that the ability to detect that the automatic responses in negotiation need to be checked with controlled mentalizing will facilitate the conditions necessary to revise these ‘faulty assumptions about the counterparty and the negotiation situation' that are the key culprit for suboptimal outcomes in negotiation (e.g., Thompson, 2005, p. 95). Specifically, insofar the fixed-pie bias (and its close cousin the incompatibility bias) is one of the key barriers to efficiency in negotiation, cognitive reflection might be one of the critical capacities for negotiators. The proposition that we test in this study is that these harmful assumptions are the automatic but erroneous intuitive responses of the automatic processes to the more or less explicit question about the counterparty's preferences and the negotiation situation. This leads to erroneous perceptions about value potential in negotiation and mandates a competitive interaction. These outputs can be, but often are not, detected and corrected by the negotiators' controlled cognitive effort.

In fact, because effective training in negotiation more or less explicitly addresses the zero-sum family of biases by urging negotiators to look beyond the salient features and generate value (Nadler & Thompson, 2003; Patton, 2009; Van Boven & Thompson, 2003), we expect that cognitive reflection will drive the effectiveness of such training. In other words, one of the outcomes of training will be increased cognitive reflection in the participants, and the superior negotiation outcomes in trained groups (vis a vis untrained groups) will be largely due to such improved cognitive reflection in trained negotiators.

Research on cognitive reflection has been so far limited to individual decision-making tasks rather than tasks involving strategic interdependence. No study so far considered cognitive reflection as an independent variable in interdependent, mixed motive tasks. This is the first study that considers cognitive reflection as a predictor of negotiation outcomes.

Relationship between training and the capacities for mentalizing and cognitive reflection

Training improves negotiation performance (Lewicki, 2014; Movius, 2008; Patton, 2009; Thompson, 1991). The effects seem to last (Coleman & Joanne Lim, 2001; Soliman, Stimec, & Antheaume, 2014) and correlate with the intensity of the training (ElShenawy, 2010; Thompson, 1991). Observational and analogical learning are more effective than didactic learning or learning by information revelation (Nadler & Thompson, 2003), and experience-based negotiation training outperforms instruction-based training (Van Boven & Thompson, 2003).

Because the fixed-pie assumption is a critical barrier to efficiency, training needs to generate capacities for correcting this bias. Even the most basic education in interest-based negotiation is effectively a call to appreciate own and other's mental states and to recruit explicit cognition to look beyond the salient features of the negotiation: 'focus on interests, not positions' (Fisher et al., 1991). In the classic anecdote (Fisher et al., 1991, p. 40; originally in Follet, 1925), two quarreling men in the library can achieve the optimal outcome only by engaging both explicit cognition and mentalizing (p. 70). Two things are required to arrive at an optimal solution; first, one must realize that there might be something else going on apart from the salient conflict. In other words, one needs to conceive of the possibility that the men might not be motivated by the open or closed window, but by something else. Second, those motivating mental states need to be made explicit. Both cognitive reflection (the ability to detect that the intuitive 'window open or closed' frame might be incorrect and engage explicit cognition) and the ability to mentalize (infer what the motivators might be) are required to move beyond the initial zero-sum frame and generate an optimal solution.

Training facilitates both. Prescriptive advice for creating value – e.g., to systematically prepare; to take into account the counterparty's perspective and identify value-creating options; to dovetail differences; to add issues to negotiation; and to make simultaneous offers (e.g., Fisher et al., 1991; Lax & Sebenius, 2006; Thompson, 2005) – is largely aimed at facilitating mentalizing and encouraging cognitive reflection, thus creating an environment where the zero-sum and incompatibility assumptions can be revised if necessary, and where value-enhancing opportunities can be found. Also, prescriptive negotiation frames, such as the seven elements framework (Patton, 2005), teach negotiators not only that they ought to mentalize or cognitively reflect, but what specific aspects of the negotiation game they ought to focus their cognitive capacities on. Negotiation is a cognitively demanding and uncertain game (De Dreu et al., 2007) and it is not uncommon for negotiators to focus on irrelevant, random aspects of the game. Even when not focused specifically on changing mindsets (Ade, Schuster, Harinck, & Trötschel, 2018), training has the effect of changing negotiators' understanding of the negotiation game. Van Boven and Thompson (2003) found that the mental models of negotiators who received training reflected greater abstract understanding that the payoff structure might not always be zero sum, and that trading depends on appreciating the perspective of the other side and information exchange.

We expect that untrained participants will be less able to deploy the cognitive mentalizing competencies that are available to them. Therefore, training will enhance the effects of mentalizing on negotiation (as it focuses the participants' social cognition on the relevant and outcome-predicting aspects of the interaction), enhance the reflective capabilities of negotiators (as it instills in the participants that the negotiation situations ought to be thought through using controlled cognition), and mediate the effects of training on negotiation outcomes.

Research questions

This study tests whether two independent predictors, the reflective functioning (the capacity to understand that action is based on underlying mental states) and cognitive reflection (the metacognitive capacity to detect a potential conflict between the automatic response and controlled cognition) predict negotiators' individual gain and dyadic gain. It further tests whether training has an effect on the employment of these two capacities and whether cognitive reflection plays a part in effective training.

Method

Overview

Trained and untrained negotiators completed a multi-issue scorable negotiation task and had their reflective functioning and cognitive reflection assessed by performance and self-reports (Fonagy et al., 2016; Frederick, 2005). We investigated the impact of the capacity for mentalizing and reflective functioning on individual and joint negotiation gains, and analyzed the impact of training on both outcomes and the employment of these two capacities.

Participants and procedure

The participants (*N* = 262) were law students at a large university in the United Kingdom. The untrained group (*n* = 172, 64% female, age 21 - 37) was recruited from the graduate population during the first week of their masters' course (see **Table 3** for a detailed breakdown). The trained group (*n* = 90) was recruited from graduate (*n* = 42, 64% female) and undergraduate students (*n* = 48, 67% female) that participated in our negotiation courses.[3] All cohorts completed the assessment of cognitive reflection. All cohorts except the last one completed the assessment of mentalizing.

Table 3. Summary of studies

Cohort	Predictors	Training	Paradigm	Age range	Mean age	*n*	Dyadic *n*
Graduate	CRT, RFQ	none	Negotiation	20-37	25	172	86
Graduate	CRT, RFQ	trained	Negotiation	21-34	26	42	21
Undergraduate	CRT, RFQ	trained	Negotiation	n/a	n/a	32	16
Undergraduate	CRT	trained	Negotiation	18-24	20	16	8
						262	131

[3] There were no differences in negotiation outcomes, CRT and RFQ scores between graduate and undergraduate groups. Gender also had no impact on the results. This is not discussed further.

Post hoc power analysis with G*power (Faul, Erdfelder, Lang, & Buchner, 2007) showed that this study's sample size had 99% power to detect the overall effects of training, reflective functioning and cognitive reflection on negotiators' individual gain, and 97% power to detect effects of these predictors on dyadic gain in multiple linear regressions with three predictors at alpha p = .05.

Ethics approval

The UCL ethics board provided the required approvals for the study (UCL 8561/002 and amendments).

Untrained group

The untrained group (*n* = 172, 64% female, age 21 - 37) was recruited from the graduate population during the induction week of their graduate degree in law (LLM) in September 2016. These participants did not know each other and had no training in negotiation. The study was conducted in one day in a group setting.

After the participants arrived at the testing room, the procedure was explained to the entire group and any questions were answered. The participants who wished to participate in the study filled out consent forms. They were then instructed to identify and pair up with an individual they did not know. They received individual confidential instructions for the task and had 20 minutes to prepare and 30 minutes to conduct the negotiation in a face-to-face session, after which they submitted the contract sheet outlining the main terms of the agreement if they reached one, or indicating that no such agreement was reached. They also completed the Reflective Functioning Questionnaire (Fonagy et al., 2016) and the Cognitive Reflection Test (Frederick, 2005).

Trained groups

The total trained group (*n* = 90) consisted of both graduate and undergraduate students that received training as part of their education. The first group of 42 participants (64% female) was recruited in November 2016 from the students attending a masters-level fully credited Negotiation module, counting toward their degree as part of the Dispute Resolution Master of Laws (LLM) pathway. The participants were invited to participate during the 10th week of their 12-week module. By then, they had completed 12 negotiation exercises and had 16 contact hours of lectures and tutorials. The topics covered included the basic game-theoretic cooperation-competition model of interdependent bargaining, the Harvard principled negotiation model, detailed instruction on competitive bargaining strategies in distributive (zero-sum) settings, and the three-tensions model (tensions between creating and claiming, empathy and assertiveness, and principals and agents). The participants received no training in logrolling and the task in this study was their first multi-issue scorable negotiation exercise. They completed the negotiation task as a weekly assignment. As usual, they were randomly paired-up at the end of the previous class and were expected to meet with their counterparty and negotiate the case during the week preceding their next class. They were instructed to limit their negotiation to 30 minutes, fill out the contract form and the relevant questionnaires, and return them during the next class.

The second group of trained participants was recruited from the undergraduate population (*n* = 32) during intense professional practice modules in Negotiation. During these modules, the students are immersed in negotiation for 8 hours per day over a period of three days. The topics mirror the ones taught in the fully credited masters module (see the first group, above), except that the emphasis is theoretical insights that are particularly applicable to the practice of negotiation, rather than on a wide range of theory that is part of the academic component of the LLM masters course. The participants received no training in logrolling and the study task was their first multi-issue scorable negotiation. They completed it toward the end of their training.

The third group of trained participants was recruited from the undergraduate population (*n* = 16) during an intense professional practice module like the second group. The only difference is that this group only performed the cognitive reflection test but was not assessed on the reflective functioning. In the present study, this group is considered only in the section that investigates the impact of cognitive reflection alone (pp. 128-132).

In all cohorts, the participants were first informed of the procedure. After a questions and answers session, the participants who wished to participate signed the consent forms against which they received the negotiation exercise and the questionnaires. They prepared in class for 20 minutes and negotiated for further 30. Upon return, they filled out the post-negotiation pack of questionnaires including the Reflective Functioning Questionnaire (Fonagy et al., 2016) and the Cognitive Reflection Test (Frederick, 2005).[4]

Task: New Recruit

The task was a multiple-issue employment negotiation used in prior research in the negotiation field (e.g., Galinsky et al., 2008; Maddux, Mullen, & Galinsky, 2008; Thompson, 1991; Thompson & Hastie, 1990). To reach an agreement, the parties, representing a prospective employer and employee, must find agreement on eight issues. Each issue has five possible options, each worth a different number of points to each party. These point values of the different contract options (the payoff schedule) are outlined in the instructions overleaf. The instructions explicitly state that these payoff schedules must not be shown to the counterparty. The payoff schedule is in **Table 4**.

[4] The differences in cohorts in terms of their level of education (undergraduate or postgraduate) were controlled for and had no impact on the findings. We do not discuss these further.

Table 4. Payoff matrix of the New Recruit negotiation task

Bonus	Recruiter	Candidate	**Job Assignment**	Recruiter	Candidate
10%	0	4000	Division A	0	0
8%	400	3000	Division B	-600	-600
6%	800	2000	Division C	-1200	-1200
4%	1200	1000	Division D	-1800	-1800
2%	1600	0	Division E	-2400	-2400
Vacation Time	Recruiter	Candidate	**Starting Date**	Recruiter	Candidate
25 days	0	1600	01 Jun	0	2400
20 days	1000	1200	15 Jun	600	1800
15 days	2000	800	01 Jul	1200	1200
10 days	3000	400	15 Jul	1800	600
5 days	4000	0	01 Aug	2400	0
Moving Expenses	Recruiter	Candidate	**Insurance**	Recruiter	Candidate
100%	0	3200	Plan A	0	800
90%	200	2400	Plan B	800	600
80%	400	1600	Plan C	1600	400
70%	600	800	Plan D	2400	200
60%	800	0	Plan E	3200	0
Salary	Recruiter	Candidate	**Location**	Recruiter	Candidate
$90,000	-6000	0	New York	0	0
$88,000	-4500	-1500	Boston	300	300
$86,000	-3000	-3000	Chicago	600	600
$84,000	-1500	-4500	Atlanta	900	900
$82,000	0	-6000	San Francisco	1200	1200

Note. Integrative gain is the sum of points negotiators achieve in bonus, vacation, moving expenses, and insurance issues (the maximum is 14,400, the split-down-the-middle compromise is 9,600 points). Compatible gain is the sum of points in job assignment and location issues (the maximum is 2,400 points, the compromise is -600 points). Joint gain is the sum of negotiators' points in integrative and compatible issues, which vary depending on dyadic performance, and distributive issues (salary and starting date) that are a constant -3,600 per dyad (the maximum is 13,200, the compromise is 4,400 points).

The issues in the task fall into three categories. Two (Salary and Starting Date) are distributive (zero-sum): the parties have opposing preferences and equal valuations (e.g., every dollar of salary costs the recruiter exactly as much as it benefits the candidate). Two issues (Job Assignment and Location) are compatible: the parties' preferences and valuations are identical, and there is a jointly preferred outcome where both negotiators benefit equally (e.g., both parties prefer the job to be in San Francisco, rather than in Chicago, Boston, Atlanta or New York). The last four issues (Bonus, Vacation time, Moving Expenses, and Insurance) are integrative: the parties have opposing preferences, but different valuations (e.g., an extra dollar of bonus benefits the candidate more than it costs the recruiter), which allows maximizing value by trading ('logrolling'; Tajima & Fraser, 2001).

Apart from the confidential payoff schedules, the instructions are identical for both parties. They consist of a page-long description of the task with a table displaying the minimum and maximum possible individual payoffs which differ based on the role-specific payoff matrix. The instructions state each party's goal is to maximize its own individual points. There are therefore no explicit instructions that the goal is to 'win' more points than the counterparty. Because both in distributive and integrative issues the payoffs are in conflict (e.g. the recruiter prefers the bonus low, the candidate high), the parties have an indirect incentive to keep their counterparty's payoffs low.

Predictors

Reflective functioning (mentalizing)

Assessing the capacity to mentalize using self-reports (as opposed to the prohibitively expensive Adult Attachment Interview-based assessment; Steele & Steele, 2008) is not without challenges. The main issue is that the ability that the self-report instrument measures is the exact same ability needed to fill out the questionnaire. Individuals must use mentalizing to answer questions about mentalizing, and it's implausible to expect someone to use mentalizing correctly to say that they are poor at it (Fonagy et al., 2016, pp. 28 - 29). Because the large part of reflective functioning occurs outside of conscious awareness and control, the participants may have little access to such processes and will be consequently unable to judge their own capacity, resulting in bias regarding the participants own ability.

The Reflective Functioning Questionnaire (Fonagy et al., 2016) was designed to tackle these challenges and discern genuine mentalizing – which appreciates the opaqueness of mental states – from hypomentalizing (the inability to infer own or others' mental states with sufficient breadth and depth), and hypermentalizing (where the individual's inference and mental models of self and others are excessively detailed, unmoored from reality and lack meaning (see *Phenomenology and neurobiology of poor mentalizing*, p. 32).

The questionnaire comprises eight items that are answered on a 7-point scale (1 = strongly disagree, 4 = neither agree nor disagree, 7 = strongly agree). The questions relate to certainty and uncertainty regarding one's own mental states (e.g., "I always know what I feel") and mental states of others (e.g., 'People's thoughts are a mystery to me'). When scored, these eight items translate into two scales: the certainty (RFQ-c) and the uncertainty scale (RFQ-u). The certainty subscale (RFQ-c) measures the confidence individuals have in understanding their own and others' mental states by indicating agreement or disagreement with statements reflecting such confidence (e.g., "I don't always know why I do what I do"). The items are rescored so that low levels of agreement reflect hypermentalizing and high levels reflect genuine mentalizing: using a 7-point Likert scale, the items are rescored to 3, 2, 1, 0, 0, 0, 0. The uncertainty subscale (RFQ-u) asks the participants to rate statements such as "Sometimes I do things without really knowing why" and the responses are recoded to 0, 0, 0, 0, 1, 2, 3. The very high scores demonstrate a general lack of appreciation of mental states, and low scores reflect the understanding that mental states are opaque, which is a mark of good mentalizing.

Since 2016, a number of studies used the RFQ for research purposes (e.g., Badoud et al., 2015; Ciccarelli, Nigro, D'Olimpio, Griffiths, & Cosenza, 2021; Huang et al., 2020; Li, Carracher, & Bird, 2020; Malcorps et al., 2021; Salaminios et al., 2020). The scale has been translated to multiple languages (e.g., Badoud et al., 2015; Y. Lee, Meins, & Larkin, 2020; Morandotti et al., 2018). It has been used also as a basis to develop further reflective functioning measures (De Roo, Wong, Rempel, Fraser, & parenting, 2019; Luyten, Mayes, Nijssens, & Fonagy, 2017). The most important findings of the these studies (for our purposes) are the correlational patterns of the certainty (RFQ-c) and uncertainty (RFQ-u) scales of the RFQ. Namely, high scores in the certainty subscales are often positively associated with mental health, suggesting that the subscale captures adaptive rather than maladaptive characteristics; similarly, the uncertainty subscale correlates with various indices of psychopathology, suggesting high scores indicate maladaptive rather than adaptive capacities. The original face validity of the scales makes sense in the context of a clinical sample.

RFQ-c is more likely to detect hypermentalizing rather than adaptive levels of social cognition if the proportion of clinical participants in the sample is significant rather than small or nonexistent. Because in our studies our participants presumably suffered no marked mental health issues, we took the view that high scores on the certainty scale of the RFQ would indicate relatively high levels adaptive, genuine mentalizing, and high scores of the uncertainty scale would indicate relatively weaker mentalizing.

Cognitive reflection

The second (social) cognitive ability we tested as a predictor of negotiation outcomes is the capacity to supervise and if need override one's own automatic responses. Contemporary dual-process theories (for an overview see Stanovich, 2011) see decision-making in terms of a power-expense tradeoff between automatic and controlled cognition, where the default outputs of the former system can be intervened on by the latter (Evans & Stanovich, 2013; Kahneman, 2011; Stanovich, 2011). The heuristic-based automatic cognition rapidly and effortlessly processes a large amount of information and its first approximations tend to be sufficiently accurate most of the time, particularly in 'benign' environments (Kahneman & Klein, 2009). However, they tend to be predictably and systematically off-mark in complex tasks (Kahneman, 2011) and need to be overridden by serial controlled computation. Because controlled processes involve high computational cost, people experience them as aversive and tend to default to automatic processes (the 'cognitive miser' phenomenon; Frederick, 2005; Kahneman, 2011; Stanovich, 2011; Stanovich et al., 2016; Toplak et al., 2014).

The degree of miserliness in information processing is different among people. The Cognitive Reflection Test (the 'CRT'; Frederick, 2005) is the quintessential measure of the capacity to detect conflicting responses of automatic and controlled systems.

The first item in the CRT reads:

> *'A bat and a ball cost $1.10 in total. The bat costs a dollar more than the ball. How much does the ball cost?'*

The intuitive answer (the result of automatic processes, also termed reflective processes or System 1) is 10 cents. The correct answer – only apparent if an intentional ‘check’ is performed (controlled processes, also termed reflective processes or System 2) – is, however, 5 cents. This item is difficult not because it requires any complicated calculation, but because it demands detecting that the automatic answer (10 cents) that ‘springs “impulsively” to mind’ (Frederick, 2005, p. 27) needs to be checked and corrected by controlled thinking (to arrive at the correct answer of 5 cents). The CRT is thus a performance test of the capacity to suspend and check the automatic response.

The CRT predicts performance in a wide array of independent thinking tasks better than assessments of cognitive ability, thinking dispositions, and executive functioning (Cokely & Kelley, 2009; Oechssler, Roider, & Schmitz, 2009; Toplak, West, & Stanovich, 2011). However, research so far investigated the impact of cognitive reflection on performance in individual thinking tasks (for an overview of thinking problems see West, 2011) that do not involve strategic interaction and mixed motives. The following section discusses a potential impact of cognitive reflection in mixed-motive negotiation.

We found the CRT particularly appropriate for this study for two reasons. First, it measures the ability to suspend and override immediate and attractive (and erroneous) solutions generated by each item (e.g., 10 cents in the ‘Ball and Bat’ item), which mirrors the immediate and attractive fixed-pie assumption in negotiation settings. Second, the CRT measures performance rather than asks for self-reports. Instead of providing opinions about their own readiness to engage controlled cognition, the test measures performance on tasks that include a ‘trap’ to ensnare the cognitive miser, thus avoiding an array of social desirability biases and problems with the reliability of reports of implicit processes (e.g., Nisbett & Wilson, 1977).

List of independent variables

The predictors were:

1. Reflective functioning of the focal negotiator (individual RFQ-c) based on the scale of mentalizing of the Reflective Functioning Questionnaire (Fonagy et al., 2016).

2. The advantage of focal negotiator against their dyadic counterparty, calculated by deducting from each negotiator's RFQ-c score the RFQ-c score of their respective counterparty (for example, if in a dyad the recruiter's RFQ-c was 2 and the Candidate's score was 3, the Recruiter's RFQ-c difference score would be -1, denoting that the Recruiter's score was 1 point lower than the score of their counterparty).
3. Dyadic mentalizing (dyadic RFQ-c), calculated by averaging the RFQ-c score of the negotiators in each dyad.
4. Individual cognitive reflection (individual CRT score), based on the Cognitive Reflection Test (Frederick, 2005).
5. Dyadic cognitive reflection, calculated as a mean of the CRT scores of the negotiators in the dyad.
6. Training, denoted by a categorical predictor indicating whether participants received any training before the exercise or not (0 = no training received, 1 = training received).

Outcome (dependent) measures

The outcome variables we used to assess the impact of the predictors listed above were all aspects of negotiation outcomes. They fall into two categories: dyadic and individual gain. First, variables measuring *dyadic gain* (value creation) were integrative gain, correctly settled compatible issues and joint gain. *Integrative gain* was measured by summing up the points negotiators realized by trading integrative issues. The optimal dyadic integrative gain was 14,400 points. A compromise split 'down the middle' outcome yielded 9,600 points. We coded the variable measuring the dyad's ability to correctly settle compatible issues on a 0 – 2 interval scale (0 = no compatible issues settled correctly, 1 = one of two issues settled correctly, 2 = both issues settled correctly). This measures the capacity of the dyad to correctly settle compatible issues more accurately than looking at the parties' combined dyadic score in compatible issues. Joint gain was measured by adding the points from integrative and compatible issues.[5] Optimal joint gain was 13,200 points. Splitting issues down the middle resulted in 4,400 points.

The variables measuring *individual gain* (effectiveness of the focal negotiator and the focal negotiator's ability to claim the value generated in the interaction) were the focal negotiator's gain as (1) an absolute number and (2) as a percentage of the joint gain of that dyad. The maximum possible gain for an individual negotiator that is theoretically possible is 13,200. However, in practice, that would entail the counterparty to accept the absolute minimum gain of -8,400, which is unlikely (although it is not uncommon that a small minority of participants accept deals where their individual value is negative). If the parties split the options in issues 'down the middle', the negotiator's gain is 4,800 points. To measure the negotiators' success in claiming value, we also calculated each negotiator's gain as percentage of joint gain in each respective dyad. For example, a negotiator's gain of 3,000 may represent a small fraction of a large joint value (e.g., 25% of a joint gain of 12,000), denoting a negotiator who was poor at claiming value in a dyad that did well in creating joint gain; the same gain of 3,000 may on the other hand represent a large fraction of a small joint gain (e.g., 75% of a joint gain of 4,000), denoting a negotiator who successfully claimed most of the joint value in a dyad that performed poorly at value creation.

Statistical analysis

We used Stata to analyze the data. Integrative and joint gain in dyads who reached no agreement were treated as zeroes in nonparametric analysis. For parametric analysis we created adjusted variables where the no-deal zeroes were replaced with the minimum scores as per common practice in negotiation research (e.g., De Dreu, Weingart, & Kwon, 2000). When assumptions for parametric models were violated, we used scaled multivariate Box-Cox transformations (mboxcox and mbctrans

[5] Because distributive issues are zero-sum, they have a constant effect on dyadic gain. In New Recruit, the net effect of distributive issues is 3,600, which is what we deducted from the sum of the compatible and integrative gain to arrive at the net dyadic gain.

commands; Lindsey & Sheather, 2010) to correct. This involved adding constants to some variables to avoid non-positive values or rounding to integer for transformed categorical variables before conducting the transformations. For the data that still violated normality, we followed nonparametric procedures, mainly quantile regression (*qreg* command) or robust regression (*robreg* command) using the M or MM-estimator (Jann, 2021). Hierarchical (nested) models with dyad as the random effect were used to adjust for dyadic performance. In multilevel models, R-squared was estimated either by running an OLS regression (if the LR test showed there was no difference between OLS and mixed regression), or by estimating Snijders-Bosker R-squared using the MLT module in Stata (Möhring & Schmidt-Catran, 2013). For mediation analysis, we used Stata's SEM and PROCESS macro for SPSS (Hayes, 2020) and estimated the confidence intervals by bootstrapping 5000 samples.

Predictions

The predictions were that in a multi-issue negotiation exercise with compatible, integrative and distributive issues:

1. Training will increase all aspects of dyadic gain (value creating).
2. The parties' reflective functioning scores will predict joint gains (value creating).
3. The negotiators' reflective functioning scores will predict their competitive success and individual gains (value claiming).
4. Cognitive reflection will predict dyadic and individual gain (value creating).
5. Cognitive reflection will mediate the impact of training on dyadic outcomes.

Results

Six dyads (5%) failed to reach an agreement. The mean joint gain (including no-deals as zeroes) was 9,073 (*SD* = 3,006) and the mean integrative gain was 10,946 (SD = 2,975). Negotiating dyads on average discovered 1.4 (*SD* = .75) of the two compatible issues in the payoff matrix and realized the mean compatible gain of 1,551 (*SD* = 1,175).

Training

As predicted, training increased all aspects of dyadic gain; one-way MANOVA showed a significant effect on both adjusted compatible and integrative gain; $F(2, 128) = 14.25$, $p < .001$, $\Lambda = .82$. The contrast analysis is in **Table 5**.

Table 5. Contrast analysis of joint, integrative and compatible gain between trained and untrained samples

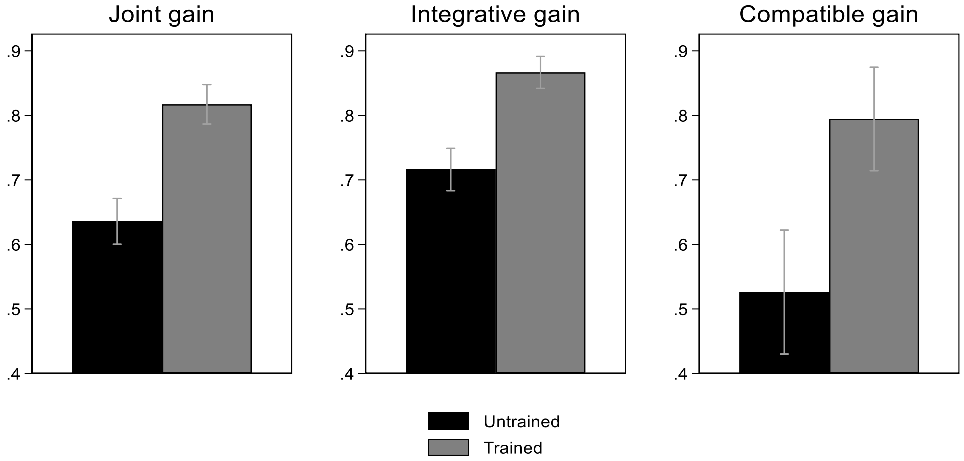
		Joint gain		Integrative gain		Compatible gain	
	N	*M*	*SD*	*M*	*SD*	*M*	*SD*
Untrained	86	8,560	2,698	10,897	3,165	1,263	1,228
Trained	45	10,787	1,954	12,480	1,729	1,907	934
t		5.41		5.15		2.97	
p		<.001		<.001		.002	
d		.901 [0.58, 1.22]		.979 [0.66, 1.30]		.472 [0.16, 0.78]	

Note. The test statistic is Welch-adjusted t-test (single-tailed). 95% CI are in square brackets. Results remain significant if adjusted for multiple comparisons.

Trained negotiators outperformed their untrained peers by 29% in (unadjusted) joint gain, resulting from a 21% improvement in (unadjusted) integrative gain and a 33% increase in compatible gain (**Figure 13**).[6]

[6] If the data pertaining to the cohort that had not completed the RFQ is removed from the dataset, the results are virtually identical: trained negotiators outperformed their untrained peers by 27% in

Figure 13. Joint gain, integrative gain and compatible gain as percentage of optimal outcome in untrained and trained groups (unadjusted)



Reflective functioning

Value creating at the individual level

At the individual level, we fitted a series of hierarchical (mixed-effects) regressions with RFQ-c score and training as predictors, and dyad as the random effect. The impact of RFQ-c on the negotiator's gain when controlled for training was significant; *b* = 389, 95% CI [34, 744], z = 1.80, *p* = .036.

The interaction between training and RFQ-c was significant, and the model with the best fit was the one including only the interaction term and no main effects; *b* = 674, 95% CI [327, 1022], *z* = 3.19, *p* = .001. Details are in **Table 6**.

(unadjusted) joint gain, resulting from a 20% improvement in (unadjusted) integrative gain and a 28% increase in compatible gain. All tests remain significant. See Appendix, p. 288.

Table 6. Multiple mixed-effects regression models of the impact of training and individual RFQ-c scores on the individual gain of negotiators

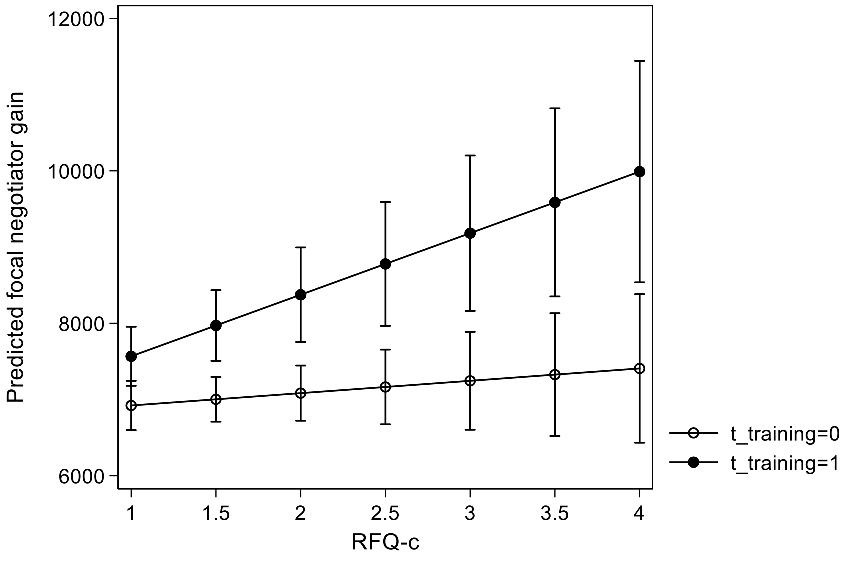
	1	2	3	4
RFQ-c	389.23*	-172.36		
	(215.83)	(284.10)		
Training	929.93**		329.64	
	(334.51)		(424.72)	
Interaction term		411.59*	304.46*	674.32**
		(172.19)	(167.98)	(212.13)
Constant	6390.98***	6223.49***	6392.73***	6985.70***
	(384.61)	(436.57)	(382.27)	(176.04)
Wald	9.44	9.50	9.20	10.19
Model *p*	.009	.025	.009	.001

*** p<.001, ** p<.01, * p<.05

Note. Standard errors are in parentheses. The models were fit using a scaled multivariate Box-Cox power transformation.

Training was a moderator of the impact of RFQ-c on the negotiator's individual gain. The impact was not significant in untrained samples (*p* = .455) but the RFQ-c was predictive in trained samples; *b* = 808, 95% CI [262, 1122], *z* = 2.66, *p* = .008. See **Figure 14** for the interaction graph.

Figure 14. Interaction between RFQ-c and training and its effect on individual gain with .95 CI



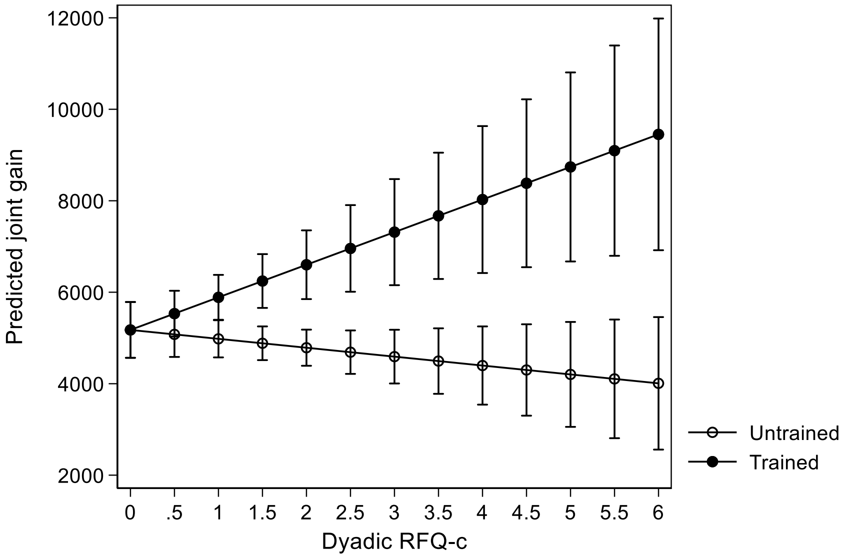
Note. The variable has been transformed using the (scaled) Box-Cox transformation.

Value creating at the dyadic level

Joint gain

At the dyadic level, the impact of RFQ-c was conditional on training. The interaction term was significant indicating that in trained (but not untrained) negotiating pairs dyadic reflective functioning drove joint gains. The model explained 10% of variance in joint gain; $F(2, 188) = 6.38$, $p = .002$, $R^2 = .098$. Refer to **Figure 15** for the graph of the interaction.

Figure 15. Interaction between RFQ-c and training and its effect on joint gain with 95% CI

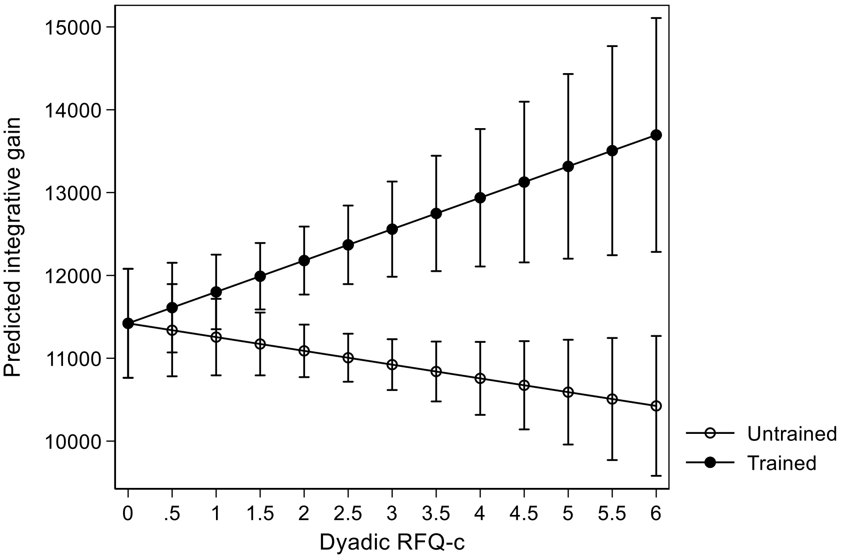


Note. The variable has been transformed using the (scaled) Box-Cox transformation.

Integrative gain

We found the same conditional effect of RFQ-c we observed on joint gain also on integrative gain, except stronger. The interaction between RFQ-c and training was significant again suggesting that trained negotiating dyads used their reflective functioning to generate integrative gain, whereas untrained dyads did not. The model explained 13% of variance in integrative gain; $F(2, 188) = 9.10$, $p < .001$, $R^2 = .133$.[7] Refer to **Figure 16** for the graph of the interaction.

Figure 16. Interaction between RFQ-c and training and its effect on integrative gain with .95 CI



Compatible gain

The reflective functioning, conditional on training, also influenced the odds of the dyads reaching optimal settlements in compatible issues; $\chi^2(1) = 7.30$, $p = .026$, pseudo $R^2 = .044$. Robust regressions of the compatible gain on the interaction term were not significant.

Value claiming

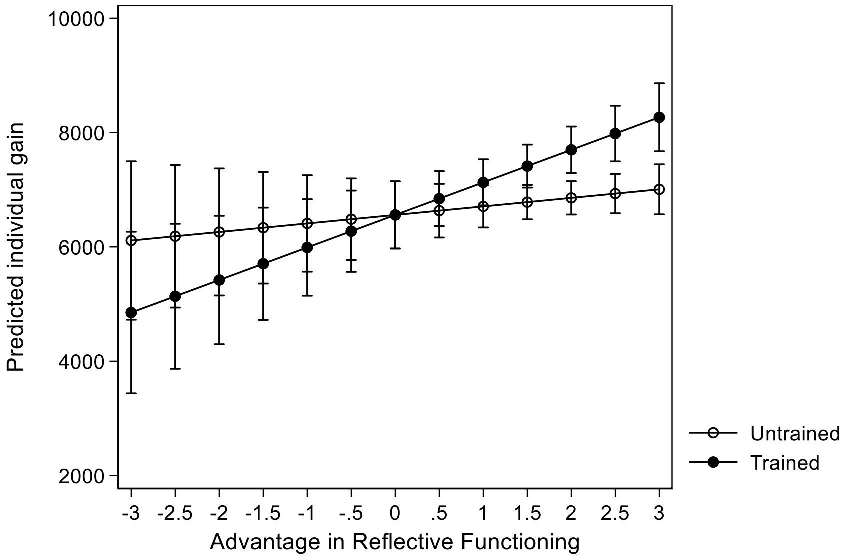
RFQc advantage assists individual gain

The reflective functioning advantage negotiators held vis-a-vis their counterparties (measured as the difference between their RFQ-c scores) was predictive of the

[7] *Post hoc* analysis showed two datapoints with excessive leverage and residuals. If we omit those, the R^2 of the model increases to 20%.

negotiators' individual gain, but that was conditional on training, suggesting that the trained individuals were able to use their mentalizing to their advantage; *Wald*(1) = 12.06, p = .002, R^2 = .050. The interaction plot is in **Figure 17**. Training was a moderator of the impact of the advantage in reflective functioning on negotiators' gain.

Figure 17. Interaction between negotiators' advantage in RFQ-c and training and its effect on individual gain



Note. The variables were subject to a scaled multivariate Box-Cox power transformation.

We also investigated the impact of the individual negotiator's reflective advantage on the share in joint gain they managed to claim. Assumptions for regressing the full sample were violated to the extent that we were not able to find a transformation that fixed that. Instead, we used quantile regression and the interaction term was bordering significance in the full sample (p = .06). In the trained sample, we were able to establish normality by removing an extreme outlier and multilevel regression showed the advantage in reflective functioning to be a significant predictor of negotiator's share in joint gain; b = .08, 95% CI [.012, .15], z = 1.95, p = .026, R^2 = .092.

RFQc advantage inhibits dyadic gain

At the same time, the mentalizing advantage negotiators held over their counterparties had an *inhibiting* effect on joint and integrative gain (**Table 7**). The mere presence of the difference between the RFQ-c scores between the negotiating parties was sufficient to depress integrative gain, which translated to losses in joint gain. No such effect was observed for compatible gain.

Dyadic reflective functioning and dyadic difference in reflective functioning were not correlated ($r = .13$, $p = .147$, $r_s = .13$, $p = .155$), confirming our assumption that they assess different aspects of dyadic mentalizing, the former denoting dyadic reflective capacity, and the latter the differences in individual reflective capacities of the negotiators.

Table 7. Impact of dyadic RFQc and the dyadic difference between negotiators' RFQc scores on adjusted joint and integrative gain

	Joint Gain	Integrative Gain
Difference in RFQc^a	-1015.20*	
	(430.21)	
RFQc * training^a	1260.12**	
	(424.99)	
Difference in RFQc		-776.97**
		(280.76)
RFQc * training		575.32***
		(129.26)
Constant	5761.79***	11588.68***
	(407.34)	(273.50)
R^2	.082	.177
F	5.27	12.65
Model *p*	.006	<.001

*** p<.001, ** p<.01, * p<.05.

Note. ^{a} denotes a variable transformed using a scaled multivariate Box-Cox power transformation.

The models in **Table 7** highlight that, conditional on training, both value creating and value claiming were predicted by reflective functioning: the RFQ-c influenced value creating and the advantage in RFQ-c influenced value claiming. Trained negotiators were able to use their mentalizing for both value creating and value claiming.

Cognitive reflection[8]

Dyadic and individual CRT scores

We then investigated to what extent cognitive reflection assists dyadic and individual negotiation outcomes. Dyadic CRT was predictive of joint gain and its components, the integrative and compatible gain, when controlled for training (**Table 8**). There was no interaction between the terms.

Table 8. Multiple regression models of the impact of dyadic CRT and training on joint, integrative and compatible gain

Dyadic N = 131

Joint gain	*b*	SE	*t*	*p*	[95% CI]	
Dyadic CRT	399.75	102.94	3.88	<.001	229.19	570.32
Training	1595.02	380.69	4.19	<.001	964.36	2225.83
Constant	5658.84	715.93	7.90	<.001	4472.65	6845.04
Pseudo R^2	.196					
Wald(2)	37.29					
p	<.001					

Integrative gain	*b*	SE	*t*	*p*	[95% CI]	
Dyadic CRT	275.40	73.51	3.75	<.001	153.60	397.20
Training	1769.26	386.50	4.58	<.001	1128.89	2409.64
Constant	7328.12	271.32	27.01	<.001	6878.58	7777.66
R^2	.265					
F	23.05					
p	<.001					

Compatible gain	*b*	SE	*t*	*p*	[95% CI]	
Dyadic CRT	81.08	48.57	1.67	.049	0.61	161.56
Training	303.78	153.04	1.98	.025	50.20	557.35
Constant	1052.66	302.90	3.48	.001	550.80	1554.53
Pseudo R^2	.200					
Wald(2)	9.16					
p	.010					

Note. Variables are adjusted for no-deals. Joint gain and compatible gain were regressed using robust regression with the M estimator (Jann, 2021). Integrative gain variable was transformed using a scaled multivariate Box-Cox power transformation.

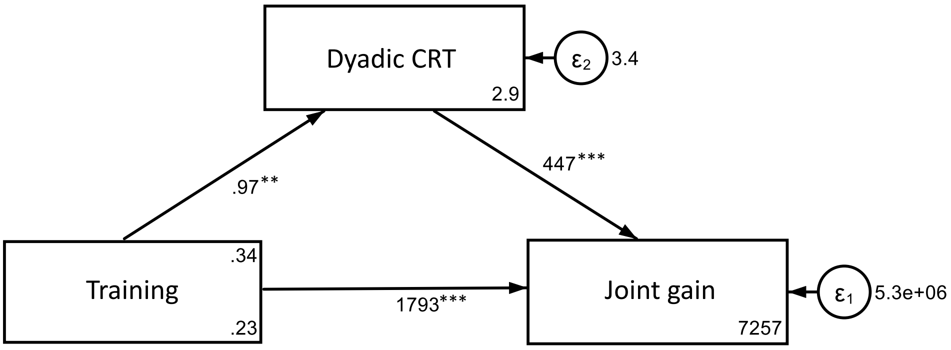
[8] This part of the study used the sample including the third trained cohort (*N* = 262). Excluding this cohort enhances the results.

A mixed-effects (hierarchical) regression with individual CRT scores and training (control variable) as predictors, and the dyad as the random effect was significant; *Wald*(2) = 36.23, *p* < .001. Both individual CRT scores (*b* = 313.54, 95% CI [114.54, 512.53], *p* = .005) and training (*b* = 1044.71, 95% CI [702.59, 1386.84], *p* < .001) were predictive. There was no interaction between the terms (*p* = .67).

Mediation analysis

As expected, the relationship between training and joint gain was partially mediated by dyadic CRT; *F*(1, 129) = 24.00, p < .001, R^2 = .157.[9]

Figure 18. Regression coefficients for the relationship between training and joint gain mediated by dyadic CRT



Training predicted the CRT scores; *b* = .97, 95% CI [0.42, 1.52], *z* = 2.88, *p* = .002. The indirect effect was significant and explained 19.5% of the total effect; *b* = 433.45, 95% CI [117, 830]. The direct effect of training on outcomes remained significant; *b* = 1792.75, 95% CI [1072, 2514], *z* = 4.09, *p* < .001. This suggests that training works partly through increased cognitive reflection in the participants.

[9] We used non-transformed variables in this part of analysis; we address this in the discussion (p. 141).

Combined impact of reflective functioning and cognitive reflection

Finally, we fit a series of models investigating the joint impact of mentalizing and cognitive reflection on negotiation outcomes.

Individual level

Cognitive reflection and reflective functioning were not correlated (r = .068, p = .29; r_s = .080, p = .21), confirming our assumption that they measure two different capacities.

At the individual level, negotiator's individual gain was a function of the dyadic cognitive reflection (marginally significant, p = .07) and either the reflective functioning (conditional on training) or the advantage in reflective functioning (conditional on training, see **Table 9** for details). The reason for the similarity of these two models is that the RFQ-c and the RFQ-c advantage are relatively highly correlated (r_s = .64, p < .001). The model featuring the interaction between training and the reflective advantage of individual negotiator provided a marginally better fit (R^2 = .058, p = .004) and suggested that both cognitive reflection and the ability to mentalistically outsmart your counterparty influence the individual gain.

Table 9. Models of the focal negotiator's individual gain predicted by the CRT, RFQ-c and the advantage in RFQ-c

	1	2
CRT	185.64^a	184.85^a
	(126.72)	(126.75)
RFQ-c * training	585.23**	
	(205.23)	
RFQ-c advantage * training		414.46**
		(135.89)
Constant	6296.11***	6277.87***
	(258.51)	(262.76)
Wald	12.57	13.45
Model *p*	.002	.001

*** p<.001, ** p<.01, * p<.05, ^{a} p < .10

Note. The variables have been adjusted using a scaled multivariate Box-Cox power transformation.

Dyadic level

On the dyadic level, joint gain was positively associated with the dyadic cognitive reflection and the dyadic reflective functioning, conditional on training (**Table 10**). This shows that the ability to engage explicit cognition, as well as mentalizing in the trained population, assist with value creating. The presence of differences in dyadic RFQ-c had a marginally significant (p = .07) negative effect on joint gain, suggesting that individual mentalizing must have been used for competitive purposes, which depressed dyadic value.

In integrative gain, the effects were stronger. Cognitive reflection and dyadic RFQ-c (conditional on training) were significant positive predictors of integrative gain, whereas the difference in dyadic RFQ-c scores was a negative predictor. The larger effect sizes suggest that logrolling is particularly adversely affected by value claiming.

Finally, compatible gain was assisted by the dyadic reflective functioning (conditional on training) and marginally by the dyadic cognitive reflection. The interaction of the RFQ-c advantage and training was not predictive. Details of these regressions are in **Table 10**.

Table 10. Models of dyadic gain predicted by the CRT, RFQ-c and the dyadic difference in RFQ-c

	Joint gain	Integrative gain	Compatible gain
Dyadic RFQ-c * training	630.89**	166.32*	97.60*
	(224.50)	99.01	(41.31)
Dyadic RFQ-c difference	-523.11^{a}	-660.14*	
	(355.19)	(299.10)	
Dyadic CRT	853.38***	692.14***	151.16^{a}
	(190.16)	(152.29)	(106.07)
Constant	4017.208***	10170.99	1061.33**
	(486.195)	(7530.64)	(307.96)
(pseudo) R^2	.227	.203	.205
F	9.75		
Wald		41.59	9.31
Model *p*	<.001	<.001	.010

*** p<.001, ** p<.01, * p<.05, ^{a} p < .10

Note. The variables in the regression of joint gain have been subjected to a scaled multivariate Box-Cox power transformation. Integrative gain and compatible gain have been regressed using a robust regression with the MM estimator at .95 efficiency (Jann, 2021).

Discussion

The findings of this study support our theoretical proposition that mentalizing, conditional on training, influences both value creating and value claiming in negotiation. We further found that the capacity to engage in explicit (social) cognition to supervise automatic (biased) responses correlates with negotiation outcomes. Finally, we found that training 'focuses' the reflective functioning of the participants and that cognitive reflection partly mediates the effects of training on negotiation outcomes.

In the following section we discuss these results. The discussion largely follows our reports in the Results section (pp. 121-132). We first discuss the impact of mentalizing and how it is moderated by training and continue by elaborating the effect of cognitive reflection on dyadic and individual outcomes. We then discuss the overall impact of mentalizing and cognitive reflection on negotiation results, make suggestions for negotiation training, and conclude with limitations and a short summary.

Reflective functioning

The findings of this study show that mentalizing, as measured by the Reflective Functioning Questionnaire, influences individual and dyadic gain in negotiation, subject to training. Moreover, mentalizing, again conditional on training, affected both value claiming and value creating aspects of negotiation. This suggests that mentalizing does indeed underpin, or at least *ought to* underpin, bargaining.

Reflective functioning influenced the negotiator's individual gain in the full sample when controlled for training, however the superior model was the one where the reflective functioning was conditional on training. In other words, the positive association between mentalizing and gains in the full sample was driven predominantly by the strength of the association in the trained sample. Negotiators who had received training were able to employ their capacity for reflective functioning to maximize individual gain far better than their untrained counterparts.

The situation was identical with dyadic gain. Both components of joint gain – the integrative gain, maximized by logrolling, and the compatible gain, maximized by correctly settling compatible issues – were driven by reflective functioning, conditional on training. Value creating involves meeting both parties' interests and preferences and to craft optimized deals, the negotiators must understand these motivating mental states. Mentalizing does that and consequently facilitates correct action choices that leads to value creating (see p. 70).

We also found evidence that negotiators use reflective functioning for value claiming. The advantage in reflective functioning negotiators held over their counterparties was associated with individual gain, again conditional on training. In other words, the better the trained negotiator's reflective functioning compared to their counterparty's, the higher their gain. The advantage was also marginally predictive of the negotiators' share in the joint gain in the full sample, and significantly associated with the share in the joint gain in the trained sample. This is not surprising. Claiming value by 'shaping perceptions about what is possible' (see p. 50 and p. 75) is an epitome of an action that requires mindreading (in fact, for most novice negotiators, negotiation *is* such social-cognitive competition). The effective use of competitive maneuvers such as anchoring, lockups, ultimatums and converging concessions depends on the ability of the user to convince the counterparty that the claiming negotiator's reservation value is much closer to the counterparty's bottom line than it really is. This presupposes a clear idea about specific aspect of the counterparty's mind, specifically on an explicit (inferred) estimate of the counterparty's reservation value, and getting an idea about the other's mind *is* mentalizing. Given that negotiators use mentalizing for the purposes of such perception-shaping competition, it is not surprising that an advantage in this capacity correlates with the success in value claiming and ultimately in negotiators' individual gain.

Perhaps the most interesting finding was that, while in the trained group dyadic mentalizing increased joint gain, the mere presence of the difference in reflective functioning between the dyadic parties depressed it. This effect was not conditional on training; it plagued trained and untrained negotiators alike. Our interpretation is as follows. The difference in reflective functioning allowed, and likely 'invited', competitive claiming of the negotiator with the better capacity for mentalizing. Negotiators claim value by social-cognitively 'outsmarting' one's opponent, and competitive tactics employed to such ends focus the whole interaction on distribution. This leaves value on the table. Negotiation literature is clear on this point: *value claiming inhibits value creating* (see *Tension between creating and claiming value*, p. 56). Distributive tactics distort the picture of preferences and take a toll on communication and relationship, frustrating generation of optimizing options (Lax & Sebenius, 1986b), and push the interaction into a positional concession-making, unmoored from interests (Fisher et al., 1991). If mentalizing is focused on competitive moves (e.g., 'What am I willing to accept on each issue?', 'What is the counterparty willing to accept?', 'What does the counterparty think I am willing to accept', 'how can I convince them that am willing to only accept this much?'), the parties can focus relatively less, or not at all, on the aspects of the negotiation process that are required for value creation (e.g., each other's relative valuations required for logrolling). In our study, the difference in reflective functioning meant not only that the negotiator with the relatively higher mentalizing managed to claim more joint value, but also that this joint value was *lower* than in dyads where the RFQ-c scores were more even. The mere presence of the difference in the social cognitive capacity entailed lower joint gain.

The mechanism underlying this is likely that the party with the advantage in mentalizing capacity was more effective at 'shaping' the perceptions of their social-cognitively weaker counterparty. This resulted in a straightforward distributive compromise, rather than requiring the dyads to work harder and find more value in a superior outcome. There is support for this in the literature. In a classic paper on strategic choice in negotiation, Pruitt (1983b) suggested that some level of contending and resistance to yielding was a necessary precursor to successful problem-solving and value optimizing (pp. 170-171; see also Carnevale & Pruitt, 1992). Premature yielding results in suboptimal joint value because if one negotiator easily gives in to the other's claims, there is no reason to seek out integrative options. In our study, such yielding was caused by superior claiming of negotiators who had an advantage in mentalizing, and resulted in both higher individual value for such parties (because of effective claiming), and lower dyadic value (because of inhibited value creating). Effective does seem to come at a price.

We did not find evidence that the difference in reflective capacities influences compatible gain. This, again, is not surprising. In compatible issues, value claiming is pointless because the preferences of negotiators are identical (e.g., both parties want the job to be in San Francisco, see **Table 4**, p. 113). However, value claiming negotiators may use compatible issues as leverage for concessions on other issues. For example, a claiming negotiator would misrepresent that their compatible preference is not what it is, but something else (e.g., not San Francisco, but New York), only to 'give in' against another concession (e.g., agree to San Francisco for a concession on Salary). This is consistent with prior studies. For example, Loschelder et al. (2014) showed what they called the 'folly of revealing compatible preferences' by demonstrating that making the first offer may backfire when the offeror reveals information about preferences that an astute recipient can leverage to their advantage.

Why is the relationship between mentalizing (reflective functioning) and both value creating and value claiming conditional on training? In other words, why does reflective functioning assist trained negotiators, but not their untrained peers? A possible explanation is as follows. Negotiation situations are highly ambiguous and uncertain ('fuzzy'; De Dreu et al., 2007). Negotiators make their bargaining choices without knowing their counterparty's interests and reservation values, and consequently what kind of, if any, deals are possible. Furthermore, communication, required to reduce the asymmetry of the relevant information between the parties, is heavily obfuscated. Because negotiators' mental states are opaque, there is no reliable way to determine whether the counterparty's intentions are collaborative or competitive (it is notoriously difficult to tell truth from lies; Bond, 2008; Bond & DePaulo, 2006; Hartwig & Bond, 2014), and consequently whether the information they impart is trustworthy. With so many moving parts, people who lack a formal framework for negotiation tend to focus on the aspects of the negotiation process which are uncorrelated with good outcomes, or are counterproductive (e.g., are they 'winning', is the other side hurting more, are they squeezing the last available penny from the counterparty; Fisher et al., 1991; Patton, 2005). Trained negotiators, on the other hand, should be able to focus on the aspects of negotiation that are relevant and predictive of outcomes, and thus employ their social cognitive capacities for such purposes. This is particularly true in complex negotiations such as a scorable multi-issue tasks. To effectively create value, a negotiator must understand very specific aspects of the counterparty's mind. With integrative gain, where the effects were strongest, the negotiator needs to understand their own and the counterpart's relative preferences and craft the trades that are mutually beneficial. Training will focus the mentalizing on the correct aspects of the interaction. To claim value, a negotiator must figure out which issues are valuable to the counterparty to then mislead them into thinking those issues are valuable to the focal negotiator, too, with a view to trade them for large concessions on the issues that are particularly valuable to the focal negotiator (a form of bait and switch tactic). A solid understanding of the negotiation process seems to be essential for both value creating an value claiming.

The finding of this study that mentalizing influences value creating and claiming is consistent with research showing that these aspects of the negotiation process depend on perspective taking (Galinsky et al., 2008; Gilin et al., 2013) and emotion recognition (Elfenbein et al., 2007), and that perspective-taking reduces the effectiveness of anchoring attempts (Galinsky & Mussweiler, 2001). No study so far, however, used the Reflective Functioning Questionnaire (see *Reflective functioning* above, p. 114), an instrument designed to measure aspects of respondent's life that are affected by (good or poor) mentalizing, such as understanding of others' minds ('people's thoughts are a mystery to me'), agency as rooted in mental states (e.g., 'I don't always know why I do what I do'), and understanding of oneself in terms of mental states (e.g., 'Strong feelings often cloud my thinking'). Also, no study so far tested the advantage in mentalizing as a predictor of outcomes. Finally, no study so far found any evidence for the inhibiting effect of such an advantage on dyadic gains.

Cognitive reflection

Cognitive reflection was a significant predictor of dyadic and individual gain in a multi-issue scorable negotiation task. To the extent that the implicit fixed-pie assumption is the major barrier to efficient outcomes, suppressing and overriding this assumption is a *sine qua non* for value creation, and the ability to detect that such suppression and override are necessary – cognitive reflection – is a key trait of an effective negotiator. To some degree, this insight is not surprising. The capacity to take a mental step back and think outside of the zero sum-frame to generate value has long been considered a key trait of an effective negotiator (e.g., the window in the library example; Fisher et al., 1991; Follet, 1925). Several studies have also indicated that the CRT may measure a cognitive trait that includes more characteristics than originally suggested by Frederick (2005). Instead of assessing only the relatively narrow capacity to detect potentially asynchronous outputs of automatic and controlled systems, the CRT may capture the more general disposition to suppress impulsiveness and conduct an elaborative domain-specific heuristics search in situations where normative models are unavailable (Campitelli & Labollita, 2010; Cokely & Kelley, 2009). Such open-mindedness would facilitate adaptive action in yet unencountered contexts, such as novel negotiation tasks.

The fixed pie bias is a particularly potent barrier to maximizing integrative gain, which is where we found the strongest effect of cognitive reflection. To logroll, a negotiator must realize there are differences between their own payoffs, which they know, and the counterparty's, which they do not. If the pie is assumed to be fixed, there is no need to pay attention to payoffs and explore whether additional value could be created by trading on differences as no differences are thought to exist (or if they do, they are deemed irrelevant).

It is worth noting that detecting the need to examine the automatic zero-sum assumption is essential, but only does half of the job. Unlike in heuristics and biases tasks, a negotiator cannot work out the solution alone, but needs to interact with the counterparty and exchange information, which is subject to the information dilemma and verification problems (e.g., Depaulo et al., 2003; Lax & Sebenius, 1986b; Murnighan et al., 1999). These barriers are the reason that the 27% effect size, while substantial, is smaller than it perhaps ought to be.

Identifying compatible issues is more random than optimizing integrative gain. The compatible option is the best outcome for both parties (choosing any other option is a 'lose-lose' agreement; Thompson & Hrebec, 1996). While a reflective detection and override of the incompatibility bias will likely lead to optimization, it is not necessary as the parties can stumble upon the correct solution by chance when one party one-sidedly discloses their general preferences on the issue (Loschelder et al., 2014). This is likely the reason for the relatively weaker effect of cognitive reflection on compatible gain that we found in our study.

Our prediction that training is effective to the extent that it brings cognitive reflection to the fore was partly supported by the results. While training did increase the CRT scores, the mediation was partial: the indirect effect was in the region of 20% of the total effect of training on outcomes. We interpret these results as follows. While suspending the output of automatic processes is procedural, it is difficult to separate process and knowledge considerations in decision-making tasks; the mindware plays a critical part (Stanovich, 2011; Stanovich et al., 2016). A person who has been trained that negotiation situations may appear zero-sum but often carry hidden value potential (Van Boven & Thompson, 2003) is more likely to detect that their immediate fixed-pie perception needs a controlled cognitive check than a person who has not received such training. This would explain the mediation effect. At the same time, stimulus discriminations, and decision-making rules and principles that have been practiced to automaticity can be part of the implicit cognition (Kahneman & Klein, 2009). In other words, a trained negotiator's automatic investigation of interests and an eager student's blind following of the instruction to 'make multiple offers' may be uncorrelated with their general tendency to resist miserly processing and engage cognitive reflection. While such heuristic tricks of the trade are grounded in the recognition of how incorrect the fixed-pie bias is, their application does not require cognitive reflection. This raises interesting possibilities for future research. For example, can cognitive reflection be trained with lasting results? Which components of negotiation training increase it? Would negotiators who use cognitive reflection be more effective than the ones using heuristic tricks of the trade in novel situations (e.g., settling a legal dispute, negotiating a border, diffusing a hostage situation, or agreeing a ceasefire in a military conflict)?

This study extends cognitive reflection research to situations involving mixed motives and strategic interdependence. These results are novel and have considerable practical importance because of the pervasiveness of negotiation in human affairs, the vast amount of value at stake, and our poor record of value optimization, wasted resources, incurred social costs and increased conflict (Pruitt & Rubin, 1986).

Combined effect of mentalizing and cognitive reflection

Considered together, the reflective functioning (the capacity to mentalize), conditional on training, and cognitive reflection (the ability to detect potentially invalid automatic responses and correct using controlled cognition) jointly predicted the negotiators' individual and dyadic gain.

In individual gain, the effect of cognitive reflection was marginally significant and predicted individual gain jointly with either the negotiator's reflective functioning or the negotiator's advantage in reflective functioning over their counterparty. The reason for equal and exclusive effect sizes of these two predictors is likely a combination of the relatively high correlation between them and the different contributions of these two capacities to the joint gain. The reflective functioning alone contributes to dyadic reflection, which we know significantly increases dyadic and consequently individual gain. The advantage in reflective functioning over the counterparty, on the other hand, facilitates superior value claiming, thus boosting individual gain at the expense of the other side. It is an open question whether the predictors drove the individual gain via successful value claiming, value creating or a combination of both. Further studies are needed to investigate this.

The dyadic joint gain and its component integrative gain were positively associated with cognitive reflection and mentalizing (conditional on training). At the same time, they were negatively predicted by the difference between negotiators' capacity for mentalizing. This is perfectly in line with our predictions. Cognitive reflection facilitates the revision of the harmful zero-sum assumption and increases joint gain. Mentalizing (conditional on training) facilitates the understanding of the relevant mental states, which also benefits dyadic gains. Finally, the difference in mentalizing capacity seems to invite value claiming, which has an inhibiting effect on joint gain. Put simply, the joint gain that negotiators manage to generate in a negotiation will be driven by the negotiators' capacity to keep their instinctive reactions in check and by their quality of mentalizing (if they are trained), and reduced if one negotiator has a stronger capacity to mentalize (because that will trigger claiming, which will in turn depress joint value).

Training

While negotiation training was not the focus of our study, our findings highlight its effect and importance. Training alone improved mean joint gain by 29%, resulting from a 21% increase in integrative gain and a 33% increase in compatible gain. This is consistent with previous research (Movius, 2008; Nadler & Thompson, 2003; Patton, 2009; Soliman et al., 2014), but extends it by showing that it facilitates both integrative logrolling and the identification of compatible issues. More importantly, training improved cognitive reflection and recruited mentalizing capacities, both contributing to improved dyadic gains. These results are encouraging and suggests that – given the prevalence and importance of negotiation in human affairs – training in negotiation ought to be part of a wide array of educational curricula.

That mentalizing only has an effect when training has already occurred is an interesting finding. There is something going on with untrained people negotiators that stops them from being able to take advantage of their capacity to think about mental states in themselves and in others in the context of negotiation. It seems that they simply switch off that social part of their mind and adopt a competitive cognitive approach that Tomasello (2014) would call individual intentionality, rather than the relatively collaborative joint intentionality mode (see also pp. 42-46, above). Training, on the other hand, seems to create a space for thinking socially. We speculate that this may be because negotiation training is in itself an intensely social experience and experiencing others learning about negotiating makes the participants more aware of other minds around them and the participants' own minds in relation to them. That is, training might have a non-specific priming function for mentalizing. Further studies are necessary on this point.

Limitations

This study is subject to several limitations. One limitation is our use of a multi-issue scorable task as a proxy for negotiation. While this type of task possesses some features necessary for research (e.g., it quantifies various types of gain and removes the necessity of the parties to engage in their own inherently subjective and noisy valuations of issues) and is a popular tool in negotiation research (De Dreu et al., 2007), there are limits to the generalizability of the findings. First, the task is artificial and requires the participants to imagine themselves being in the roles introduced in the case. Second, the values of issues are arbitrarily assigned in a way that is unmoored from any real interests the participants may hold (e.g., participants may, or may not, care more about five days of holiday than they do about an extra 2% bonus). Because of this, the mental states (that negotiators need to mentalize) are not something that is inherently important and affectively relevant to the individual participants. In other words, what they try to infer from each other's behavior is not their emotionally infused, subjectively important interests in a bonus, holidays and job assignment, but rather what the instructions say the relevant points are. It may be that such cold cognitive knowledge is not something that is easy to mentalize. At the same time, our participants – all students of law, a relatively competitive profession – might have cared relatively more about 'winning' and their reputation of being a successful negotiator amongst their classmates, which might have motivated them sufficiently for research purposes. However, in spite of such shortcomings, multi-issue cases like the New Recruit are commonly used in negotiation research and the consensus in the academic community seems to be that while the magnitude of the effects may not directly generalize, we can expect the effects themselves to manifest in real-life negotiation (De Dreu & Carnevale, 2005a).

Second, although the findings lend general support to the proposition that cognitive reflection improves outcomes because it enables higher-level cognitive override of the implicit fixed-pie bias, the study correlated outcomes with CRT scores rather than with any record of the hypothesized cognitive processes. We did not administer questionnaires about implicit processes during the negotiation to avoid providing hints to participants that could interfere with the experiment, and have not conducted a post-negotiation survey because participants' self-reports of such implicit processes tend to be unreliable (Nisbett & Wilson, 1977). Future studies might consider tackling this challenge using qualitative methods to tease out the processes that enable dyads with higher CRT scores to capture higher dyadic gains.

Third, the improved results in the trained group may be partly due to an epistemic understanding of negotiation as carrying integrative potential that the negotiators shared at the dyadic level (i.e., both negotiators knew they attended the same training advocating negotiation as commonly containing hidden value). This may have resulted in a tacit value-claiming ceasefire that allowed negotiators to explore options boosting integrative and compatible gain. To the extent that such epistemic collusion is a significant factor, the demonstrated benefits of training may not fully generalize to situations outside of the joint learning environment. At the same time, this also serves as a reminder of the benefits of a widespread education in negotiation and conflict resolution. This is an area for further study.

We used original, non-transformed variables in linear regressions of the structural equation model of the mediation effect of CRT on training. Stata and SPSS do not feature robust regressions in the SEM and mediation analyses. Because power transformations did not fully remove the abnormality of residuals, we decided to use the original non-transformed variables. We believe this does not significantly distort the model because our sample size was relatively large (23 observations per parameter), and linear regression models without normally distributed errors tend to be valid, provided the sample size is sufficient – and sufficient in this context means at least 10 observations per parameter (Schmidt & Finan, 2018) or even less (Austin & Steyerberg, 2015).

The next set of limitations relates to our assessments of the cognitive capacities. The Reflective Functioning Questionnaire (Fonagy et al., 2016) has been developed fairly recently and its properties are still under investigation. The questionnaire is primarily designed to detect severe malfunctioning or absences of mentalizing that point to serious personality disorders rather than to provide a fine-tuned measure of the capacity for social inference in a non-clinical population, which is what we require to adequately model the relationship between mentalizing and negotiation outcomes. As the RFQ or other measures of mentalizing develop, further studies will need to retest these predictions. The Cognitive Reflection Test (Frederick, 2005) has the opposite problem. The CRT is rather popular as a research instrument and its popularity might have the unfortunate effect of reducing the test's validity as its items become widely known (Haigh, 2016; Thomson & Oppenheimer, 2016). However, the performance on the CRT seems to be stable over time and robust to multiple exposures. Stagnaro, Pennycook, and Rand (2018) identified 3,302 unique participants who had completed the CRT two or more times and found a strong correlation between their first and last CRT scores (*r* = .81). Bialek and Pennycook (2018) found that multiple exposures do not invalidate the CRT. Meyer, Zhou, and Shane (2018) examined over 14,000 Mechanical Turk participants who took the test up to 25 times and found that prior exposure failed to improve scores; the participants' increase in scores was a mere 0.024, and even that was chiefly driven by the minority who spent time reflecting on the questions. Finally, the later scores retain the predictive validity of earlier ones, as the initial success and subsequent improvement measure the same ability.

Related to that, it is possible to object that rather than assessing the tendency to avoid miserly cognition, the Cognitive Reflection Test (CRT) simply measures participants' numeric ability. Because the New Recruit task involves a fair amount of calculation (of how different options affect the points negotiators are asked to maximize), simple numeric literacy might be the source of correlation between the CRT and the results (i.e., the dyads who can calculate gains better can realize better gains). However, a series of studies (Toplak et al., 2011, 2014) demonstrated that while the CRT has a significant correlation with the general cognitive ability, regression analysis showed that the test was a unique predictor of resilience to various judgmental biases and explained variance not accounted for by intelligence and other individual differences. The authors suggest that the reason is that neither intelligence tests nor assessments of executive functioning test 'the tendency toward miserly processing in the way that the CRT does' and go on to argue that 'the CRT is a particularly potent measure of the tendency toward miserly processing because it is a performance measure rather than a self-report measure' (Toplak et al., 2011, p. 1275).

In addition, the CRT was designed to be a performance-based test of general reasoning (e.g., the ball and the bat problem) and not a test of social cognitive performance. However, because what the CRT measures is the penchant for suspending the immediate, effortless (easy) intuitive judgment in order to engage effortful explicit cognition, it is possible to treat the CRT as a general test of cognitive alertness (the negative of cognitive misery). Irrespective of that, ideally, our studies measuring the impact of the ability to shift from implicit to explicit mentalizing would employ a social-cognition-based CRT-type of test. A development of such an instrument would be a worthwhile task.

Finally, the study provides evidence for the role of mentalizing and cognitive reflection in multi-issue tasks with hidden value potential. However, negotiation comes in many forms. To replicate and extend the findings of this study, the effects of our predictors ought to be tested in different negotiation tasks, in particular in single-issue distributive (zero-sum) situations without value potential (that involve predominantly value claiming), and in partisan-perception driven dispute-resolution tasks (that feature predominantly value creating). We conduct such tests in Study 2 and Study 3.

Conclusion

Overall, this study provides support for our theory that mentalizing underpins bargaining. The outcomes at the individual and dyadic levels were a function of mentalizing and cognitive reflection. It seems that the capacities to imagine people in terms of mental states and to resist impulsive intuitions and instead engage in an effortful deliberation improve performance in negotiation. Taking a step back and thinking carefully about the situation and the other’s mind increases gain. Haste makes waste, and so does mind-blindness.

STUDY 2: REFLECTIVE FUNCTIONING AND COGNITIVE REFLECTION IN DISTRIBUTIVE NEGOTIATION

The present study investigates whether the capacity to mentalize (assessed by the Reflective Functioning Questionnaire; Fonagy et al., 2016), particularly a negotiator's advantage over their counterparty in the mentalizing capacity, and cognitive reflection (as measured by the Cognitive Reflection Test; Frederick, 2005), affect individual outcomes in a single-issue, zero-sum task negotiated by untrained participants.

Study 1 provided support to the idea that in complex multi-issue negotiation settings, trained parties use their mentalizing to claim value. The present study attempts to extend these findings to value claiming of untrained negotiators in zero-sum tasks. In other words, we investigate whether participants without training could make use of their mentalizing for competitive purposes in a simpler negotiation paradigm where the only outstanding issue is price and no hidden value exists. Because distributive bargaining and sequential concession-making is what novice negotiators expect when they approach any negotiation (O'Connor & Adams, 1999), and such 'positional bargaining' is something people are uniformly familiar with (Fisher et al., 1991), we expected that no training would be necessary for the effects of mentalizing on individual gain to appear. In other words, we expected the untrained participants in a single-issue distributive task to be behave like the trained negotiators in the multi-issue task in Study 1.

This study also tests whether cognitive reflection – the metacognitive capacity to detect potentially faulty automatic responses and override them with controlled cognition – facilitates individual outcomes in a distributive negotiation. Study 1 suggested that cognitive reflection assists individual gains of both trained and untrained negotiators, and provided strong support for the hypothesis that it assists with value crating. While cognitive reflection might facilitate dyadic gains, presumably by allowing negotiators to revise the faulty fixed-pie and incompatibility biases, it is plausible that the ability to second guess one's own intuitive responses and proceed algorithmically also assists in value claiming. The present study tests this proposition.

Method

Overview

Untrained negotiators completed a single-issue, distributive (zero-sum) negotiation task and had their reflective functioning and cognitive reflection assessed by performance and self-reports (Fonagy et al., 2016; Frederick, 2005). We investigated the impact of the capacity for mentalizing and reflective functioning on individual negotiation gains.

Participants and procedure

The participants

The participants (*N* = 82) were students of law at a large university in the United Kingdom. They were recruited between November 2017 and February 2018 from the population pursuing undergraduate and graduate degrees in law during the first day of an intensive training in negotiation. The participants received no performance-linked rewards or compensation for their participation.

The participants were recruited during two separate intensive negotiation courses in November 2017 and February 2018. These participants had no prior training in negotiation. They conducted the negotiation during the first day of their course after being given a lecture on the interdependent nature of bargaining but before any instruction on specific value-claiming strategies.

The participants were first informed of the procedure. After a questions and answers session, the participants who wished to participate signed consent forms against which they received the negotiation exercise and the questionnaires. They prepared in class for 10 minutes and negotiated for further 15 minutes. Upon return, they filled out post-negotiation questionnaires, including the RFQ and the CRT.

We based our power analysis on the effect of the difference in reflective functioning on the share of joint gain in trained samples in Study 1. This required a sample size of 80 for 80% power to detect effects of a single predictor in a linear model at alpha p = .05. Because some of the participants have not correctly completed the questionnaires, our sample size was five people short and would be sufficient for a 77.5% *a priori* power. The power analysis was made using G*power (Faul et al., 2007).

Ethics approval

The UCL ethics board provided the required approvals for the study (UCL 8561/002 and amendments).

The task: Navy Contract

Navy Contract involves a distributive (zero-sum) negotiation over a price of components (navigation chips) required by a buyer to perform its obligations under a contract with its customer. Each negotiator is provided with a party-specific payoff schedule (**Table 11**) outlining the impact of potentially agreed prices on their profit margins and on their company generally. Some of the prices are accompanied with precedents: prior transactions the seller has concluded with other customers, outlining how many units were sold, at which price, and to which customer. This information is identical for both parties and ensures that they have ample data to derive estimates of each other’s alternatives and reservation values, and decide their distributive strategies in an informed manner. The information about profit margins and financial consequences of prices for each party is confidential.[10]

Table 11. Payoff schedule of the Navy Contract task

Moon Microsystems

Price p/unit	Profit Margin	Notes / details of specific contracts
2,200	9.1%	Minimum acceptable margin
2,300	13.0%	
2,400	16.7%	
2,500	20.0%	Apple (40,000 units)
2,600	23.1%	Ministry of Defence (10,000 units)
2,700	25.9%	UC hospitals (5,000 units)
2,800	28.6%	
2,900	31.0%	
3,000	33.3%	
3,100	35.5%	
3,200	37.5%	Abrams (15,000 units)
3,300	39.4%	
3,400	41.2%	Argyll (10,000 units)
3,500	42.9%	
3,600	44.4%	
3,700	45.9%	Emmerson Co (3,000 units)
3,800	47.4%	Sumner Inc (5,000 units)
3,900	48.7%	
4,000	50.0%	
4,100	51.2%	
4,200	52.4%	
4,300	53.5%	
4,400	54.5%	
4,500	55.6%	
4,600	56.5%	
4,700	57.4%	
4,800	58.3%	
4,900	59.2%	
5,000	60.0%	

Stealth Inc

Price p/unit	Profit Margin	Notes / details of specific contracts
2,200	9%	
2,300	13%	
2,400	17%	
2,500	20%	Apple (40,000 units)
2,600	23%	Ministry of Defence (10,000 units)
2,700	26%	UC hospitals (5,000 units)
2,800	29%	
2,900	31%	
3,000	33%	
3,100	35%	
3,200	38%	Abrams (15,000 units)
3,300	39%	
3,400	41%	Argyll (10,000 units)
3,500	43%	**Tolerable loss**
3,600	44%	
3,700	46%	
3,800	47%	Sumner Inc (5,000 units)
3,900	49%	
4,000	50%	
4,100	51%	
4,200	52%	**Serious cash flow problems**
4,300	53%	**10% chance of bankruptcy**
4,400	55%	
4,500	56%	
4,600	57%	**50% chance of bankruptcy**
4,700	57%	
4,800	58%	
4,900	59%	
5,000	60%	**99% chance of bankruptcy**

The seller's reservation value is somewhere between 2,200 (the 'minimum acceptable margin') and 2,400 (just below the Apple contract, driven by the instructions stressing the precedential effect). The buyer's reservation value is, similarly, somewhere between and 4,500 (just short of 50% chance of bankruptcy) and 4,100 (slightly less

[10] The full exercise is available from the author upon request.

than ‘serious cash flow problems’). The situation is zero-sum as a dollar more for one side is a dollar less for the other.

Predictors

The independent variables were the differences between the dyadic parties’ RFQ-c scores as assessed by the Reflective Functioning Questionnaire (Fonagy et al., 2016); and the cognitive reflection score as per the Cognitive Reflection Test (Frederick, 2005).

Outcome (dependent) measures

The dependent variable was the negotiator’s individual gain, calculated as the financial benefit of the negotiating party based on the agreed price and their reservation value. The reservation values we took were 2,200 for the seller and 4,500 for the buyer.[11] For example, if the negotiators struck a deal at a price of 3,000, the seller’s individual gain was 800 and the buyer’s 1,500.

Predictions

We expected that negotiators’ advantage in reflective functioning would predict individual gain. We also made a tentative prediction that cognitive reflection (CRT) would assist negotiators’ individual gain.

[11] An alternative specification of negotiators’ gain is based on the actual results of the negotiation, using the minimum (2,450) and maximum (4,100) prices achieved by negotiators in the exercise. Using this approach makes the reported results more conservative.

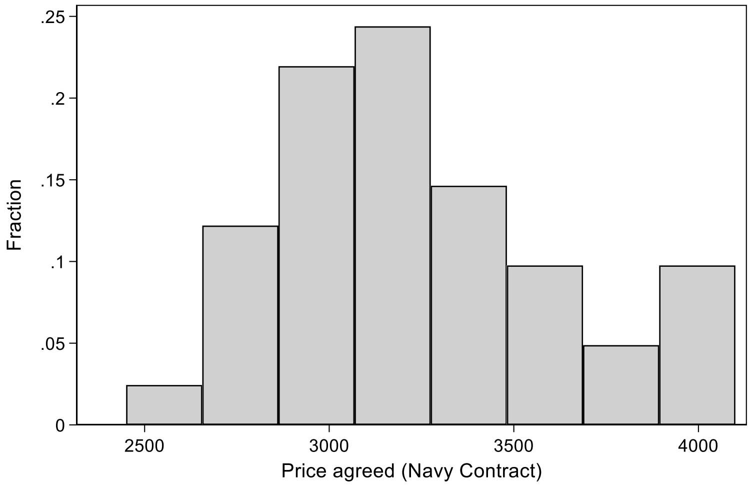
Statistical analysis

The analysis was performed with Stata. Mixed effects (hierarchical) regression with the dyad as the random effect was used when modelling the impact of mentalizing and cognitive reflection on individual gain. Overall, the same steps were followed as in Study 1.

Results

All participants reached an agreement. The mean price was 3224 (*SD* = 385). See **Figure 19** for a histogram of reached prices.

Figure 19. Price agreed in the Navy Contract negotiation



Hierarchical (mixed-effect) regression of the individual negotiator's gain on the difference between negotiators' RFQ-c scores,[12] allowing the slopes to vary, was significant; *b* = 241, 95% CI [10, 473], *z* = 1.71, *p* = .043. Cognitive reflection on its own was not predictive (*p* = .84). However, a model with both cognitive reflection and the difference in reflective functioning was significant; *Wald*(2) = 5.34, *p* = .035.

[12] Standard errors were calculated using a robust variance estimator to adjust for heteroscedasticity.

Cognitive reflection predicted individual outcomes; *b* = 66, 95% CI [9, 123], *z* = 1.90, *p* = .042. The RFQ-c difference remained significant; *b* = 236, 95% CI [11, 461], *z* = 1.73, *p* = .029.

Discussion

In a single-issue, purely distributive negotiation, the advantage in the RFQ-c and cognitive reflection were predictive of the focal negotiator's gain.

While the findings of Study 1 suggested that trained, but not untrained, negotiators can use their reflective functioning for claiming value in a complex, multi-issue negotiation, the present study investigated whether untrained negotiators can effectively employ their reflective functioning in a less complex, one-issue, zero-sum negotiation without any hidden value potential.

Our interpretation of the results of Study 1 was that training is important in complex negotiations as it provides a basic framework for preparing and conducting a negotiation, and an understanding of negotiation as a process. We then made a tentative prediction that in simple, zero-sum negotiations, where frameworks and understanding negotiation normally required for dealing with complex situations were not strictly necessary, negotiators might be able to use mentalizing effectively to claim value even in the absence of training. This prediction turned out to be correct.

The reason that no training was necessary for the effects of mentalizing on gains to appear seems to be that we are by default prepared for competitive negotiation. Distributive bargaining is what novice negotiators expect when they approach any 'negotiation' (O'Connor & Adams, 1999). They assume the situation is zero-sum and without any value potential and that impasse is a likely outcome. They also expect that the process will be one of sequential offers, what Fisher and colleagues call 'positional bargaining' (Fisher et al., 1991) where the seller starts high, the buyer low, and then they exchange concessions, attempting to convince the other side to accept the final price as close as possible to their aspiration value. These expectations are so strong that one can conduct distributive bargaining using hands and fingers if the counterparty does not speak the same language, and in some cultures is entrenched and expected (Fisher et al., 1991, pp. 72-73). While real-life negotiations are not commonly a single-issue, zero-sum affairs (Fisher et al., 1991; Pruitt & Lewis, 1975; Raiffa, 1982), when they actually are – such as in this case – naïve negotiators are relatively well prepared and know what to expect. This is the reason behind the finding of the present study: the negotiators were able to focus their reflective functioning on the correct aspects of the negotiation.

Cognitive reflection was associated with the negotiators' individual gain. As it is the capacity to engage explicit cognition (Evans & Stanovich, 2013; Kahneman, 2011; Lieberman, 2007; Lieberman, Gaunt, Gilbert, & Trope, 2002) in situations that do not apparently require it (Frederick, 2005; Toplak et al., 2011), the role of cognitive reflection does not seem to be constrained to the dyadic level in negotiation situations that offer non-obvious opportunities to create value (Fisher et al., 1991; Pruitt, 1983b; Raiffa, 1982, 2002) that are stunted by the implicit assumptions negotiators commonly bring to the table, mainly the zero-sum and incompatibility assumptions (Bazerman & Neale, 1991, 1992; Brett & Thompson, 2016; Thompson, 1990a, 1991; Thompson & Hastie, 1990). In such situations, the ability to engage explicit thinking assists in revising these value-inhibiting assumptions and assists the negotiators at the dyadic level (i.e., at the level of joint gain).

However, the findings of this study show that cognitive reflection seems to also assist value claiming. This is not entirely surprising. Several studies have demonstrated that the CRT may assess a metacognitive trait that is wider than the relatively narrow capacity to detect and override faulty automatic responses as originally suggested by Frederick (2005). Instead, the CRT is likely to capture a general disposition to conduct an elaborative domain-specific heuristics rather than rely on impulsive (automatic) responses, particularly in novel situations (Campitelli & Labollita, 2010; Cokely & Kelley, 2009). This trait, a combination of open-mindedness, non-impulsiveness and a propensity to think algorithmically, is likely to facilitate adaptive action in a wide array of tasks, which would include competitive negotiation.

The limitations of this study are similar to the limitations of Study 1. The first one is the use of an artificial negotiation task. One-page written instructions about negotiating on behalf of a corporation deviates in important ways from negotiation briefs and incentives present in real life, so the generalizability of the found effects may be somewhat constrained. However, the ecological validity of the Navy negotiation task is likely relatively higher than the one of the New Recruit used in Study 1; while it does not possess the detailed information and realistic incentive structures (e.g., promotion or bonus depending on the outcome of the negotiation), the negotiation brief is not fundamentally different from those present in distributive situations in real life (i.e., the key information regarding the context, the issue and the reservation values are sufficient for generating an ecologically valid distributive situation). In addition, there seems to be a consensus in the negotiation research community that observed effects themselves are likely to replicate in real life settings, even if the magnitude of the estimates turn out to be incorrect (De Dreu & Carnevale, 2005).

Second, while consistent with Study 1, the effects sizes found in this study are relatively weak. In addition to being slightly underpowered, the likely reason is that that the effective employment of reflective functioning depends on the level of training of the participants (as demonstrated by Study 1 for multi-issue tasks). Our participants in this study were completely untrained. We believe that if the participants received specific training in distributive tactics (e.g., converging concessions, opponent reservation value perspective taking, anchoring, non-offer offers) the effect size of reflective functioning would increase. A new study is needed to test this hypothesis.

Overall, this study indicates that in a sample of untrained negotiators, reflective functioning and cognitive reflection are predictive of individual (competitive) success in single-issue, distributive negotiation.

STUDY 3: REFLECTIVE FUNCTIONING IN PARTISAN-PERCEPTION DISPUTES

Negotiation is involved in many aspects of human interaction. One classification in this field separates deal-making and dispute resolution (Mnookin, 2000): some negotiations are predominantly about reaching a deal (where the alternative is status quo), and others are mainly about resolving a dispute (where what normally follows is a risky and uncertain formal dispute resolution process like litigation or arbitration). While there is certainly an overlap between these two – dispute resolution often involves some sort of a deal (e.g., clauses in the settlement) and deal making often includes conflict management – they are sufficiently different that they warrant separate investigation. This is particularly important because in disputes, the value of shared and different (logrolling) interests of the parties can exceed the value of the dispute, but that remains hidden to the disputants, who often focus solely on the value of the dispute (Patton, 2005).

The present study investigates whether the capacity to mentalize (assessed by the Reflective Functioning Questionnaire; Fonagy et al., 2016) influences the resolution of disputes that are driven predominantly by differences in the parties' perception of the conflict situation. In the introductory part of the thesis we discussed at length how naïve realism – the inability to appreciate differences between own mental states and mental states of others stemming from biased perception of one's own objectivity – stems from poor mentalizing (pp. 84-87). Conversely, good mentalizing is marked by the appreciation that others' mental states are observationally opaque and that while we can approximate what others feel, think, believe, desire, and so on, we cannot know their mental states with certainty. Genuine mentalizing comes with measured modesty and curbed confidence in one's own social-cognitive representations, stemming from the awareness that our understanding of others' minds is based on inference, an approximation, rather than shared knowledge. Reflective functioning should therefore act as an antidote to excessive naïve realism. The present study tests this proposition.

Method

Overview

Trained participants conducted a negotiation of a dispute that was predominantly driven by the parties' erroneous partisan perceptions. They were assessed on reflective functioning. We investigated to what extent the participants' capacity for mentalizing assists with the successful settlement of the dispute.

Participants and procedure

The participants (*N* = 204) were students of law at a large university in the United Kingdom. They were recruited from the population pursuing graduate and undergraduate degrees in law. The participants received no performance-linked rewards or compensation for participating.

Participants were recruited on three occasions. The first group (*n* = 34) was recruited during an intensive negotiation workshop in October 2016 and the second group (*n* = 52) during an identical workshop in February 2017. The participants conducted the negotiation during the last day of training. Before that, they had negotiated and debriefed multiple negotiation cases and had received training and instructions in strategic interdependent nature of bargaining, interest-based principled negotiation (Harvard model), distributive tactics, and the tension between creating and claiming value. They completed the negotiation task used in this study prior to (and educationally as part of) the presentation of the tension between empathy and assertiveness (Mnookin, 2000). On the penultimate day of the negotiation module, the participants were informed of the procedure. After any questions were answered, the participants who wished to participate signed consent forms, against which they received the negotiation task (the students who did not wish to participate in the research also did the task but did not receive the questionnaires). The participants then prepared overnight and negotiated the case in the middle of the third day of the module. They had 45 minutes to complete the negotiation.

The third group (*N* = 118) was recruited in November 2019 from the population pursuing a graduate (masters) degree in law. The participants were recruited during the first semester course in a regular masters level module *Negotiation for Lawyers*. Before negotiating the task, the participants received training and instructions in the strategic interdependent nature of bargaining, interest-based principled model of negotiation (Harvard model), the basics of the decision-analytic negotiation model, distributive tactics, and parts of the three-tension model (Mnookin, 2000). As the other participants, they completed the negotiation task used in this study prior to (and educationally as part of) the presentation of the tension between empathy and assertiveness. The participants received the negotiation instructions as part of the coursework and were asked to pair up with an individual they did not know. They completed the negotiation as a weekly assignment.

Based on the results of regressions of joint gain on reflective functioning and the dyadic difference in reflective functioning in Study 1, G*power (Faul et al., 2007) suggested a full sample of 196 (dyadic *n* = 98) for 90% power to detect an effect in a linear regression with two predictors at alpha *p* = .05. We recruited 204 participants (our cutoff policy was to recruit by cohort until the required sample size was reached). However, nine participants have not completed the RFQ, so our effective sample size was 195 (dyadic *n* = 95), translating to 89% *a priori* power. If we base the required sample size calculation on regressions of integrative gain in Study 1, the required sample size was significantly smaller; we would need 150 participants (dyadic *n* = 75) for 99% power for linear regression with two predictors at alpha = .05.

Ethics approval

The UCL ethics board provided the required approvals for the study (UCL 8561/002 and amendments).

The task: Hijacked Performance

Hijacked Performance is a negotiation case involving a dispute between two students over the discharge of a debt. The case features a student in financial difficulties (defendant) who is forced to borrow money for tuition from a wealthy fellow student (the claimant). After working hard to repay the loan, the students are robbed at the moment when the defendant hands over the money to the claimant in an attempt to repay the loan.[13] The students do not know each other very well and are plagued by stereotypes; the defendant perceives the claimant as a spoiled and overly rich person and the claimant suspects the defendant is a habitual drug user. Ultimately, the claimant feels taken advantage of and is anxious that the defendant perhaps even plotted the whole robbery with his 'drug buddies'. The defendant, on the other hand, cannot afford to, and feels it is unfair to be asked to, again repay a loan he has already discharged, particularly when, as he feels, the robbery was entirely the fault of the claimant.

The negotiation is between the representatives of the disputing students and the instructions of the task are written in a way that encourages partisan perceptions. Each representative is only exposed to the version of events and facts as perceived by their client. In designing the task, we were careful that the instructions do not differ in facts, but only in the clients' perception of these facts (for examples please see **Table 12**).[14]

[13] The first version of the case grew out of a joke: Two guys are walking down the road. A robber jumps them: "Your money or your life!" As the two men are begrudgingly pulling their wallets out of their pockets, one guy says to the other: "Here's the twenty bucks I owe you."

[14] Full text of the exercise is available upon request from the author.

Table 12. Examples of partisan perceptions in Hijacked Performance task

Confidential information for the representative of the defendant (Alvin Wright)	**Confidential information for the representative of the claimant (Reginald Beaucy-Gardiner)**
Feeling terrible about the affair, Alvin worked double shifts over the next two months to be able to repay the money, completely neglecting his studies and social life. He did not contact Reginald during that time, as he puts it "half because I was embarrassed and half because I worked 16 hours, 7 days a week."	After that, Reg did not hear from Alvin for over two months. He tried calling, but no one picked up. He sent emails which went unanswered. He tried contacting a few joint acquaintances, but no one had seen Alvin anywhere. He also looked up Alvin's lecture schedule and waited after a couple of lectures, to no avail. Alvin literally disappeared.
The conversation was awkward at best. Reginald was distant and requested Alvin to wire the money to his bank account. Alvin had to insist on handing him the cash, as he was still more than £600 in overdraft and would not be able to repay the full £3,000 if it had to go through his bank. Eventually Reginald agreed with an annoyed "Fine, whatever," and hung up.	The conversation was awkward at best. With a rude attitude, Alvin asked where to deliver the £3,000. By that time Reg was uncomfortable as much as annoyed. He politely asked that the money be wired to his bank account, however Alvin aggressively insisted on handing it over in cash. Eventually Reg agreed with a resigned "Fine, whatever," and hung up.
[no information as Alvin is not aware of the effect his appearance, due to his heavy work schedule and lack of sleep, has on Reginald]	As [Reg] was passing the underground station he bumped into Alvin, who just got out of the tube. Alvin looked terrible. He had bloodshot eyes, his clothes were shabbier than ever, and Reg got uncomfortable. Before he had his suspicions, but now he was sure this whole business was related to drugs.
The silence was painfully uncomfortable for Alvin. His mumbling attempts at conversation were met with cold silences.	Alvin kept mumbling something unintelligible to himself. Reg's mind was racing anxiously.
[no information as Alvin had not noticed the dozen 'dangerous looking teenagers']	As they entered a small dark alley, a shortcut between the main road and the pub, Reg noticed – to his increasing alarm – a group of about a dozen dangerous looking teenagers standing on the side of the alley. Completely scared by now and desperately trying not to show it, Reg plodded along. Acutely aware of the unfriendly stares of the gang, Reg tried very hard not to make eye contact. They passed by the gang and nothing happened.
At some point, in the middle of a narrow alley between the main road and the pub, Alvin had had enough: "It was so humiliating. I just wanted to get it over with, you know, so I pulled the envelope with the three grand out of the bag. Reginald was not even looking at me. 'Look mate,' I said, 'here is your three grand… Just take this, so we're done and I can f*** off,' and I pushed the envelope into his hands. Reg took it and thumbed through the stack of bills, checking if everything was there. Then, with a start, he looked up, and his eyes went wide."	Then suddenly Reg felt a shove and heard Alvin's voice: "Here's your three grand, take it, so we're done and… f*** off!" He looked at Alvin, who was pushing an envelope into his hand. Stunned, he automatically took it. It contained a fat stack of £20 notes. All this – the shove, the money in a dark alley – was so unexpected that the whole situation got that eerie surreal feeling, Reg tells you. Dumbfounded, he just stood there, thumbing the bills. The stack was about an inch thick. "Must have been three thousand in there, I reckon," Reg says. After a couple of seconds, he remembered where he was. With a start he looked up.

There is a large amount of anecdotal evidence available about the behavior and outcomes in Hijacked Performance because the task forms part of the regular curriculum in the *Negotiation for Lawyers* course for graduate and undergraduate students of law. Hijacked Performance is used to demonstrate the difference between problem-solving and pleading approaches in dispute resolution (Moffitt, 2005) and highlight the benefits of empathy and perspective-taking (Mnookin, 2000) in negotiation settings generally.

Unlike other negotiation tasks used in this research, Hijacked Performance is a case where the main value can be created through a successful resolution of a dispute. The task does involve minor aspects of value claiming, because the parties must agree the amount and terms of repayment, if there is to be any. However, the value claiming aspect is dwarfed by the challenge of agreeing or not agreeing a settlement at all, obstructed by the partisan perceptions of both parties. The key aspect of the task is to agree – or not – a settlement that generates value by avoiding litigation. Careful analysis of the disputants' interests shows that litigation is a lose-lose proposition and can only be justified by an unwarranted and overconfident belief in the strength of one's own case, underpinned by naïve realism (see pp. 84-87). Settlement is only possible if the parties effectively inquire about and share information about each of their client's minds.

Anecdotally, the deal-rate in Hijacked Performance is around 60%. Negotiators who manage to resolve the dispute and settle tend to engage in a thorough exploration of their clients' perspectives. This results in a realization that there is a multi-layered version of the 'truth' that was differently experienced by the two clients. Multi-perspectival understanding of the situation underpins an exploration of interests and crafting of an agreement that avoids litigation. No-deals tend to be the result of the process in which the parties focus on argumentation (pleading) and legal entitlements, normally underpinned by the perception of the problem as a binary win-lose situation (as a court would see it). In the debrief, we discuss the differences between the problem-solving and pleading approaches in terms of the information the parties focus on, whether they consider contributions of each of the parties (or remain focused on fault), their intentions and emotions, and whether they pay attention to the future relationship or remain focused on what had happened in the past (for a comprehensive analysis of the differences between pleading and problem-solving, see Moffitt, 2005).

If the parties remain in the pleading mode, they commonly perceive Hijacked Performance as a single-issue distributive negotiation. The perceived task is then to distribute the pain of the stolen money. The negotiators who frame the issue this way often reach no agreement, because both clients feel very strongly that they should not be required to bear that cost. The claimant feels his generosity has been taken advantage of by a malicious hustler. The defendant cannot afford to pay the debt again and feels it would be unfair to require him to do so (again). Both parties feel that the robbery was the other one's fault. On the other hand, a problem-solving approach – or, more precisely, an approach based on mentalizing – tends to make possible a careful exploration of the parties' mental states, both their perceptions of the facts and events that gave rise to the dispute and their motivating interests. This refocuses the negotiation on interests and away from claiming value in a distributive contest of assigning blame. Such an approach often has an effect of the negotiators realizing their clients were both victims, which dispels the conflict and the perception that the other side is unreasonable and at fault. This insight then facilitates solutions where the parties' needs are met in different ways.

Because litigation is a poor outcome for both parties, a good outcome in Hijacked Performance is any deal that avoids it: it avoids litigation that is highly uncertain (a 'toss-up' according to one of the expert barristers we consulted when writing a case), allows the defendant to continue his studies, and ensures the claimant does not feel taken advantage of and keep the unfortunate event secret. A myriad of deals achieving this outcome are possible. The most common arrangement is a deferred payment of a certain proportion of the debt in question, usually around 50%, that the defendant needs to transfer to the claimant in such a way that his studies are not jeopardized, normally over a period of years or months, or becoming due once the defendant completes his degree or secures employment.

Predictors

The independent variables were the RFQ-c and the RFQ-u scale of mentalizing of the Reflective Functioning Questionnaire (Fonagy et al., 2016) for each role, the combined reflective functioning score calculated by deducting the RFQ-u from the RFQ-c (for each participant), and the absolute value of the difference between the RFQ-c scores of the dyadic parties (refer to Study 1, p. 114, for details).

Dependent measures

The outcome variable was a binary variable whether a deal that avoided litigation (successful settlement) was achieved or not (0 = no deal, 1 = deal agreed). Any features of payment, such as deferred disbursements, a loan with or without an interest rate, and such, were disregarded.

Statistical analysis

The analysis was performed with Stata. The impact of dyadic and individual reflective functioning was assessed at the dyadic level. We fit a logistic regression model with dyadic RFQ-c, the difference between negotiators' RFQ-c scores and the interaction term as predictors, and reached agreement as the outcome variable. We also explored individual (role-specific) reflective functioning, where we fit a series of logistic regression models using individual RFQ-c, RFQ-u and the difference between them as predictors and reached agreement as the outcome variable. There were 4 outliers in the RFQ-u scores, which were capped at mean ± 2.5 SD.[15]

[15] Keeping them unchanged does not alter the analysis in any meaningful way.

Predictions

In a case driven by partisan perceptions, we expected good mentalizing to facilitate an understanding of the other side's perceptions and provide some protection against the fundamental attribution error and naïve realism. Because a mark of good mentalizing is an appreciation that mental states in oneself and others are opaque, and must be inferred rather than observed, some uncertainty and curiosity about thoughts and perspectives of the counterparty should facilitate deals. Therefore, we expected that mentalizing (as measured by the RFQ) would be positively associated with the odds of resolving the dispute and reaching an agreement.

Furthermore, in Study 1 we found that the dyadic difference in reflective functioning exerts a significant negative impact on value creating, presumably by facilitating an environment where value claiming takes the central position, thus depressing dyadic gains. We therefore expected that the (absolute value of the) difference in the RFQ-c at the dyadic level would exert a negative impact on the odds of successfully resolving the dispute.

Further, while we expected both parties' (and possibly joint) reflective functioning to contribute to dispute resolution, the mental capacities of the representatives of the claimant ought to have a relatively stronger effect. The reasons are the following. First, the case is asymmetric when it comes to the parties' partisan perceptions. While the defendant's view of the events differs from the claimant's, his perceptions are mainly about himself and his situation, rather than about the claimant. In other words, the defendant's perceptions are not the key driver of the dispute. The claimant's perceptions, however, are largely about the defendant himself: the defendant seems to be a person who actively hustled him and possibly even committed a criminal offence. Because of that, the claimant's understanding of the defendant's mental states would be key for the resolution of this dispute. Second, Hijacked Performance is a settlement negotiation, where the key driver of the underlying litigation is the claimant, who is the ultimate decision-maker when it comes to abandoning his legal claim. The claimant – rather than the defendant – needs to be convinced to let go of litigation and settle; the defendant would be happy with the status quo.

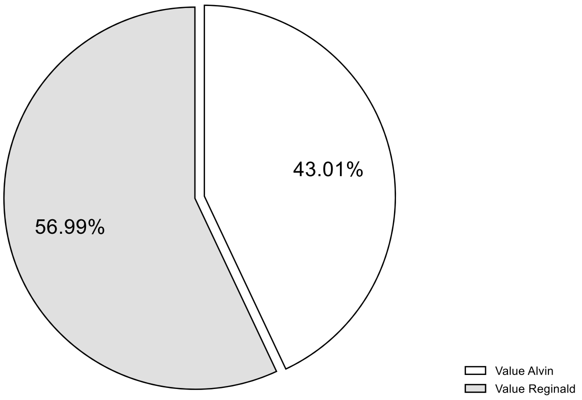
In exploring the role specific impact of reflective functioning, we used the uncertainty subscale RFQ-u in addition to the certainty subscale RFQ-c. The reason was that we expected the uncertainty regarding the mental states to negatively predict the ability of negotiators to revise their naïve realism driven perceptions and consequently to reduce the odds of successful settlement.

Results

Deals and value distribution

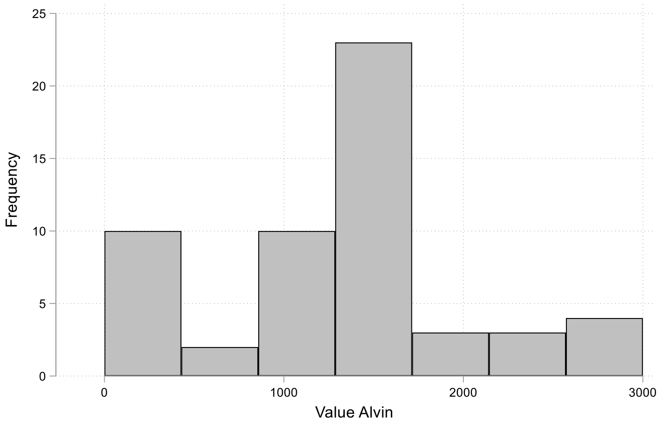
Sixty percent (60%) of the negotiating dyads reached an agreement. In the concluded deals, representatives of the claimant claimed more (*M* = 1700, *SD* = 810) than the representativeness of the defendant (*M* = 1300, *SD* = 810). This difference was significant; *t*(108)= 2.59, *p* = .006. See **Figure 20** for the pie chart.

Figure 20. Percentage of value claimed by representativeness of claimant and defendant



The distribution varied quite a bit, but the most common result was a 50-50 split which was also the median (*med* = 1,500). See **Figure 21** for the histogram.

Figure 21. Histogram of value claimed by representatives of the defendant in concluded deals in Hijacked Performance negotiation



Reflective functioning

Dyadic reflective functioning

Logistic regression of the odds of successful agreement on dyadic reflective functioning was predictive, but the effect size was small; *b* = .41, 95% CI [.029, .791], *z* = 1.77, *p* = .036, pseudo R^2 = .029.

In the series of models considering both the dyadic RFQ-c and the RFQ-c difference between the dyadic parties, the best fit was the model featuring the dyadic RFQ-c difference and the interaction term. The dyadic RFQ-c difference was negatively associated with the odds of a successful deal; *b* = -1.68, 95% CI [-3.11, -.25], *z* = -1.94, *p* = .027. The interaction term was significant, too; *b* = .58, 95% CI [.06, 1.09], *z* = 1.82, *p* = .034, 95% CI [.06, 1.09].

Individual (role-specific) reflective functioning

Spearman correlations of the RFQ scores and the incidence of successful agreements indicated that the RFQ-c of the representatives of the claimant (r_s =.27 , p = .007), but not of the defendant (p = .94), were significantly associated with the deal odds. As predicted, the RFQ-u scores of the defendant did not correlate with reaching of a settlement agreement (p = .78), and for the claimant the relationship was marginally significant (p =.09).

We fit a series of logistic regression models on the individual RFQ-c and RFQ-u scores of the claimant. They showed a positive impact of the RFQ-c and a negative impact of the RFQ-u scores of the claimant. See **Table 13** for the summary.

Table 13. Logistic regression models for the impact of RFQ-c and RFQ-u of the parties on the probability of achieving a mutually acceptable settlement

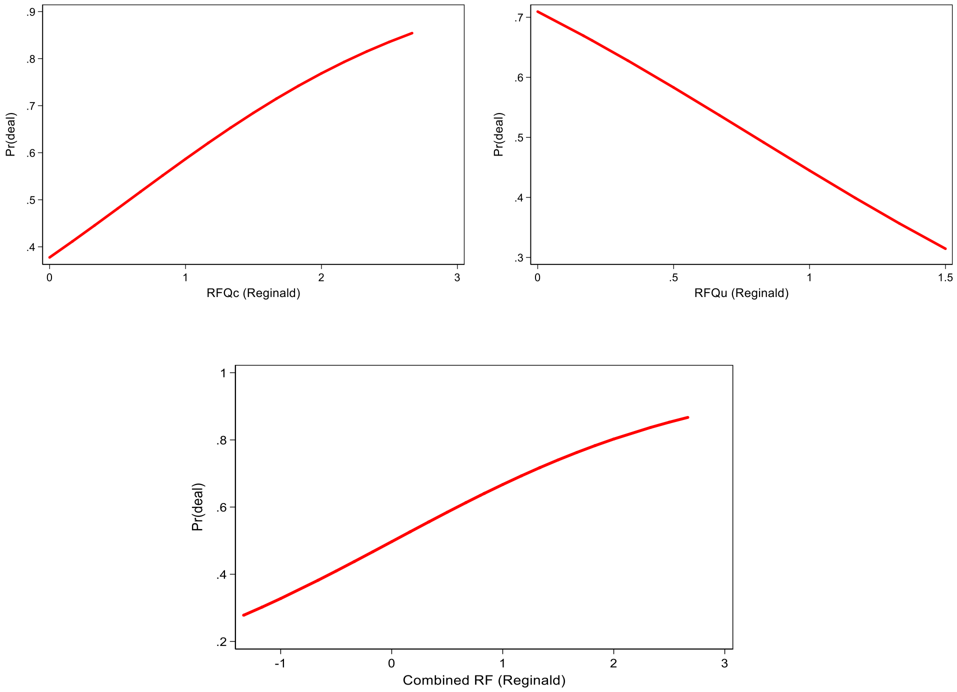
	1	2	3
RFQ-c Claimant	.85*		
	(.34)		
RFQ-u Claimant		-1.11*	
		(.52)	
Combined RF score			.71**
			(.25)
Constant	-.5	.89**	-.012
	(.41)	(.31)	(.27)
R^2	.052	.037	.073

*** p<.001, ** p<.01, * p<.05, ^{a} p < .10

Note. Standard errors are in parentheses.

See **Figure 22** for the predicted impact of the claimant's reflective functioning on the probability of agreement.

Figure 22. Predicted impact of the claimant's RFQ-c, RFQ-u and combined RF score on the probability of settlement



Discussion

The results support our hypothesis that reflective functioning (as measured by the RFQ; Fonagy et al., 2016) assists in dispute resolution. Overall, the dyadic RFQ-c score was associated with the odds of reaching a successful settlement. We also observed an inhibiting effect of the difference in the capacity for reflective functioning between the dyadic parties on the value creating aspect of the negotiation, the same effect that we observed in Study 1. When adjusted for roles, the reflective functioning of the representatives of the defendant were nonpredictive, but the reflective functioning of the claimant significantly affected the odds of successfully settling the dispute. Both certainty and uncertainty subscales of the RFQ of the claimant predicted successfully concluded settlements.

We predicted that the reflective functioning capacities of claimants would exert larger effect on the odds of good outcome than the same capacities in defendants. This role-dependency of effects is not uncommon in negotiation research. Unlike game theory tasks, such as the prisoner's dilemma or social dilemmas, that feature symmetric information and rules (e.g., R. Axelrod & Dion, 1988; Batson & Ahmad, 2001; Batson & Moran, 1999; Dawes, McTavish, & Shaklee, 1977; Liberman, Samuels, & Ross, 2004; Rilling et al., 2002; Weber, Kopelman, & Messick, 2004; Zhong, Loewenstein, & Murnighan, 2007), ecologically valid negotiation tasks are normally role-asymmetric. Both goals (interests) and information are different for the participating parties. The participants playing different roles are therefore not looking at an identical task, but rather at different ones. As the result, it is not uncommon that the effects of predictors is role-specific (e.g., Galinsky et al., 2008, for an example of role-asymmetric effect of perspective-taking).

There are two factors that may individually or jointly explain this role-specific effect of mentalizing on the frequency of settlements in this case and perhaps even on the resolution of partisan-perception driven disputes generally. *First*, for a settlement to take place, legal proceedings must be discontinued, and in that, the claimants' decision making is more relevant than the defendants'. The defendants may suffer a variety of biases that make them prone to refuse settlement and behave litigiously in commercial disputes (e.g., loss aversion that causes risk-seeking attitude generally; also, bias in specific instances of high- and low-probability cases, see the 'fourfold pattern', the 'certainty' and 'possibility effects', and the Prospect Theory in general; Kahneman, 2011; Tversky & Kahneman, 1992). However, the structure of the legal process in litigation demands that for a lawsuit to stop (failing a court decision), the claimant must discontinue their legal claim (which may be particularly difficult as dropping a claim may be experienced as a 'loss' due to the 'endowment effect' and loss aversion; Kahneman, 2011; Kahneman, Knetsch, & Thaler, 1990). Because of this crucial dependency of the settlement on the action of the claimants, it is reasonable to expect that the claimant's reflective functioning may exert a disproportionate influence on the odds of settlement: the more a claimant can understand herself and – particularly in partisan-perception driven disputes – the mental states of the other side, the likelier the settlement. *Second*, in the present negotiation task, the claimant's perceptions of the events and the counterparty are relatively more biased than the defendant's when it comes to the issues crucial for this dispute. While the defendant's views about the claimant and what transpired on the night of the robbery do differ significantly from the claimant's (and were colored by the defendant's own bias), they do not constitute the core of the dispute. For example, the defendant has no perceptions that stand in the way of understanding the claimant's interests (e.g., the claimant's feelings of being taken advantage of). By contrast, the claimant's (erroneous) perceptions of the defendant underpin his belief that he was being taken advantage of by the defendant, and possibly criminally assaulted. These perceptions had to be revised for a settlement to take place. This is likely why the claimant's representatives' reflective functioning mattered relatively more than dyadic reflective functioning and the reflective functioning of the defendant. The study supports the idea that mentalizing assists in the resolution of affect-driven disputes involving partisan perception of the warring parties (this is also consistent with research showing that mentalizing facilitates successful mediation; Howieson & Priddis, 2012; Howieson & Priddis, 2015).

The certainty aspect of the mentalizing capacity assessed by the RFQ-c positively predicted settlement odds, and the uncertainty aspect predicted them negatively. This makes sense; the certainty aspect of the RFQ assesses an individual's confidence that one's own, and others', agency is underpinned by mental states. The individuals scoring higher on the certainty scale are likely better able to navigate their social worlds by inferring mental states and considering mental causation. In the present case, such individuals were better able to facilitate communication and understand their counterparty's mental states, which eventually led to successful settlements. Conversely, high scores on the uncertainty scale reduced the odds of a successful settlement. The uncertainty score reflects the individuals' distrust or incapacity to sufficiently consider mental causation. The higher such uncertainty reigns, the less likely such individuals are able to resolve the challenges of biased (partisan) perceptions.

Our results also provide tentative support for the claim that reflective functioning may reduce the social enmity effects of naïve realism. The capacity to correctly infer mental states of others is associated with an appreciation that mental states are opaque and that the perceptions of another's motivations and beliefs are inferences, rather than what the person is 'really like', and that these inferences are more or less correct, rather than exactly equivalent to reality. Such awareness may provide the necessary trigger of explicit mentalizing to enrich or revise the automatic social cognitive inference (Van Overwalle & Evandekerckhove, 2013). High-quality mentalizing may therefore facilitate a sense of intellectual modesty and curiosity that counteract overconfidence and safeguard against the three tenets of naïve realism (see also *Psychic equivalence*, p. 81).

The predictive value of reflective functioning in the context of resolution of disputes is encouraging because in such situations the disputing parties are likely biased *against* easy revisions of their beliefs about their opponents. The parties are reluctant to understand each other's perspectives for fear of appearing concessionary (hence the famous pro-tip: "understanding is not agreeing"; Fisher et al., 1991, p. 21). Also, settlement negotiation is fraught with dangers of unwitting and harmful disclosure of information that may be damaging in the future, which causes the parties to be less than forthcoming with information and questions. Finally, there is a perceived risk of appearing weak, so the prospective litigants tend to huff and puff, exaggerate the willingness to walk away and inflate the attractiveness of litigation more than usual. All this is to say that the finding that, in spite of all that, reflective functioning does make a difference is encouraging and supports claims of Howieson and Priddis (2012, 2015) that mentalizing facilitates mediation shifts (see p. 100). Perhaps reflective functioning training ought to form part of any dispute resolution education.

This brings us to the limitations of the present study. Like in our other studies, the first limitation is our use of an artificial task as a proxy for real life negotiation. It requires the negotiators to imagine the roles they are asked to play and their real-life interests are not linked with the interests of the fictional parties they represent. However, our participants were students and they were motivated to do well. Also, the Hijacked Performance task is more ecologically valid than the multi-issue scorable tasks used in our other studies because it provides the negotiators a relatively comprehensive background explaining the interests and background of their clients, very much like these things happen in real life. Nevertheless, direct generalizability might be limited. Second, our participants were students of law and might have been more litigious than people from other professions. It is unclear whether that enhanced or depressed the effects we found. Third, the study has a limited number of outcome variables. In addition to the binary variable of the existence of a successful settlement, future research should include various other measures such as post-negotiation questionnaires assessing the employed process and subjective value attached to the process and the outcome (e.g., Curhan et al., 2005; Curhan et al., 2010). This will provide a more comprehensive picture of the effect of social-cognitive predictors on the negotiation as a whole.

Overall, this study found that reflective functioning, particularly of the protagonist in the dispute (the claimant), contributed to the out-of-court resolution of disputes by presumably enabling the revision of erroneous perceptions and attenuating the adverse effects of naïve realism in negotiation of disputes driven by partisan perceptions.

CONCLUSIONS AND REMARKS

We have tested the theory that mentalizing must underpin negotiation because it facilitates negotiators' understanding of the mental states that are critical for negotiation: their motivators and their beliefs about optimal bargaining strategies. Mentalized content is the critical input into negotiators' strategic choice and behavior. In the theoretical chapter (pp. 68-78) we stated the general model of mentalized-focused bargaining that negotiating parties use social cognitive inference (mentalizing) to infer observationally opaque mental states of their counterparties based on observable behavior, and then use these inferred mental states to choose their own actions in light of the counterparties predicted behavior.

In the three studies that followed (Study 1, 2 and 3) we tested the two key hypotheses that emanate from the theory that mentalizing underpins bargaining: that mentalizing predicts value creating and value claiming in negotiation. Below we summarize the key findings of these studies and discuss their implications. We start by discussing the findings of the three studies related to mentalizing and cognitive reflection, continue by commenting on the joint impact of these two cognitive capacities on the negotiation processes and outcomes. In the final part we discuss general implications of our findings for negotiator cognition and outline the research question for the second part of the thesis.

Reflective functioning

In Study 1, the capacity for reflective functioning (measured by the RFQc scale of the Reflective Functioning Questionnaire; Fonagy et al., 2016), conditional on training, was a significant predictor of the negotiators' individual and dyadic gain in a multi-issue scorable task. Negotiators who had training were able to use their reflective functioning more effectively than their untrained peers.

Trained negotiators used mentalizing for value creating. The reflective functioning scores, conditional on training, were predictive of the overall joint gain, integrative gain, and odds of dyads reaching optimal settlements in compatible issues. The negotiators also used mentalizing for value claiming, to two effects. First, the advantage in the reflective functioning capacity (trained) negotiators held over their counterparties was predictive of the individual competitive gain of negotiators, presumably because such negotiators social-cognitively outperformed their less clever peers and managed to claim a higher proportion of dyadic gain. Second, such difference in mentalizing capacities was negatively associated with dyadic joint gain in both trained and untrained samples. In other words, the larger the difference in reflective functioning between the dyadic negotiators, the smaller the dyadic joint gain. Negotiators seized any opportunity presented by their reflective advantage opportunity to claim value, rather than create it. We will return to that later.

In Study 2, untrained negotiators engaged in a single-issue, purely distributive negotiation. Consistent with Study 1, the advantage negotiators held in the RFQ-c over their counterparties predicted the focal negotiator's gain. The reason no training was necessary for the mentalizing to be effective is because naïve negotiators assume zero-sum payoffs, sequential concession-making, and a conflict of wills whenever they approach any 'negotiation' (Fisher et al., 1991; O'Connor & Adams, 1999). They were thus prepared to engage in the purely distributive task we used in the study.

In Study 3, we tested whether reflective functioning assists with the resolution of partisan-perception disputes, i.e., disputes where the core of the conflict lies in mismatched (partisan) perceptions. Because one of the key drivers of such disputes is naïve realism, and naïve realism is rooted in the inability to appreciate the perspective of the other side (which is the definition of poor mentalizing), we expected reflective functioning to improve the odds of successfully resolving the dispute. The results provided some support for our expectations. Dyadic reflective functioning was indeed positively associated with the odds of settlement. In addition, and consistent with the findings of Study 1, the dyadic difference in reflective functioning inhibited the odds of successful dispute resolution. Finally, both the certainty and uncertainty scales of the reflective functioning of the protagonist of the dispute (the claimant) were significantly associated with the odds of reaching a settlement agreement, presumably by facilitating the revision of erroneous perceptions and attenuating the adverse effects of naïve realism in dyadic negotiation of disputes driven by partisan perceptions.

In summary, reflective functioning (as measured by the RFQ; Fonagy et al., 2016) predicted successful value creating and claiming (conditional on training) of participants in multi-issue tasks (Study 1), successful value claiming of untrained participants in single-issue distributive task (Study 2) and the odds of reaching a settlement agreement in a partisan-perception dispute (Study 3). This provides support for the theory that mentalizing underpins the key processes in negotiation.

Cognitive reflection

In the present studies we investigated the impact of the cognitive reflection – the metacognitive ability to suspend and potentially override automatic responses with controlled cognitive effort (as measured by the Cognitive Reflection Test; Frederick, 2005) – on dyadic and individual gains of negotiators engaged in a multi-issue task (Study 1) and in a single-issue, distributive bargaining task (Study 2). Cognitive reflection was employed as a proxy for the ability of negotiators to engage explicit reasoning and social cognitive inference to complement their implicit mentalizing.

In Study 1, cognitive reflection was a significant predictor of dyadic and individual gain in both trained and untrained participants negotiating a multi-issue scorable task. Because the implicit zero-sum bias is the critical barrier to efficient outcomes, overriding it is essential for value creation, and cognitive reflection seems to facilitate it. Substantial evidence shows that the fixed-pie bias is a particularly harmful barrier to value creation and affects particularly gains that need to be realized by logrolling. This is where we found the strongest effects of cognitive reflection. Because the fixed-pie bias is essentially an automatic social-cognitive response (an erroneous perception of the counterparty's mental states), cognitive reflection facilitates the suspension and examination of this bias by explicit reasoning about the preferences and interests of the other side (controlled mentalizing). This improves value creating.

Both individual and dyadic gains were significantly affected by the individual and dyadic cognitive reflection abilities. This is not surprising: being able to mentally step back and reason about (and around) the zero-sum frame has long been deemed an important capacity of negotiators. The capacity to suppress impulsiveness and conduct an elaborative search for what could work in novel situations, purportedly measured by cognitive reflection, facilitates open mindedness and is likely to enhance performance in novel negotiation tasks. Furthermore, the study showed that cognitive reflection partially mediates the effects of training of negotiators, suggesting that a part of the effectiveness of training is indeed due to the enhanced ability of the negotiators to engage in cognitive reflection.

In Study 2, cognitive reflection assisted in value claiming in a single-issue distributive task. In pure distributive tasks, the fixed-pie assumption does no harm because there is no hidden value in the task (i.e., the assumption is correct because the payoffs are zero-sum). Our interpretation is that cognitive reflection allowed the negotiators to plan and execute their competitive distributive tactics and strategies in a controlled cognition-mediated way rather than rely on intuitive impulses. Cognitive reflection seems to have merit beyond the revision of bias.

Mentalizing and cognitive reflection

Cognitive reflection, reflective functioning and the difference in reflective functioning between the dyadic negotiators were highly predictive and explained almost a third of variance in outcomes. This is not surprising. The CRT and the RFQ measure different aspects of cognitive functioning. Cognitive reflection denotes a metacognitive capacity to detect situations where erroneous automatic responses might be faulty and to check and potentially override them with controlled thinking. In negotiation, that likely translates to detecting and revising the fixed-pie and incompatibility assumptions related to the task payoffs. Higher cognitive reflection means higher odds of getting a solid understanding of the task payoffs, and correspondingly a higher chance of exploiting the differences and similarities. Reflective functioning, on the other hand, is the capacity to understand that the behavior is driven by observationally opaque mental states and act accordingly. This attunement to mental states can be employed to both create joint value, as well as to maximize individual gain at the expense of value creating. The former is reflected in the positive association between dyadic gains and dyadic mentalizing, and the latter in the negative association between dyadic gains and the difference in the reflective functioning capacity. In short, dyadic gains seem to be driven by the ability to engage controlled cognition if necessary (cognitive reflection), and how the capacity for mentalizing is used for value creating (reflective functioning) and value claiming (difference in reflective functioning).

General discussion

In the studies summarized above we have investigated the effects of individual differences in mentalizing and cognitive reflection on various types of negotiation outcomes in different negotiation paradigms. While this kind of trait-based research does suggest that the dispositions toward reflective functioning and cognitive reflection correlate with outcomes, it does not tell us what these dispositions bring to the strategic interaction to have such effect. In other words, we know that trained negotiators who are better at mentalizing and have a propensity to reflect on their intuitive reactions do better than their peers in whom these capacities are less pronounced. What do the negotiators with superior mentalizing and cognitive reflection do differently from their less gifted peers when they negotiate a task? In what different ways do they approach the interaction? Also, we have seen that training improves performance by increasing cognitive reflection (which is a mediator of the effect of training on outcomes, see p. 129), and by focusing the negotiators' mentalizing (training is a moderator of the effect of mentalizing on outcomes, see p. 122-128). What do trained negotiators do with these capacities so that the outcomes improve? What strategies do they employ? In what ways are their mental models about negotiation different from their untrained, less effective peers?

Negotiation research shows that negotiators suffer 'faulty assumptions' about 'the counterparty and the negotiation situation' (Thompson, 2005, p. 95), meaning the fixed pie and the incompatibility biases (Bazerman et al., 1985; Bazerman & Neale, 1986; De Dreu, Koole, et al., 2000; Pinkley et al., 1995). In addition to assuming their interests are in conflict or incompatible with the interests of the other side, untrained negotiators expect that the bargaining process involves conflict, impasse and competition (O'Connor & Adams, 1999).

Our studies provide support to the idea that both cognitive reflection and mentalizing – joint models explained almost a third of variance in joint gain – assist the revision of these biases and help negotiators optimize joint gain (and also depress it should they use mentalizing for competitive purposes). This suggests that training, which partially works through increasing cognitive reflection and focusing the mentalizing capacities of negotiators, does not enhance gains by honing negotiators' value claiming skills, but by facilitating cooperative behaviors that lead to the maximization of joint gain. This is not surprising considering that the evolutionary pressures that selected people with superior social cognition were predominantly collaborative (see *Mentalizing as an evolutionary adaptation*, pp. 40-47). It is also consistent with the major message stressed by virtually all mainstream negotiation prescriptive literature that negotiators ought to focus on maximizing joint interests (Fisher et al., 1991), creating value by moving north-east on the pareto graph (Lax & Sebenius, 2006) and expanding the pie (Thompson, 2005), and with research that shows that trained negotiators' mental models resemble the models of collaborative negotiators (Nadler & Thompson, 2003; Thompson, 1990a; Van Boven & Thompson, 2003). In other words, with mentalizing, negotiators shift the negotiation slightly away from competition and towards the collaborative Vygotskian stance, away from individual intentionality and slightly towards the joint intentionality mode of constructive negotiation.

This raises a new question. Why do trained negotiators use their capacities for value creating, but untrained do not? What is it about 'negotiation' that makes untrained negotiators assume a fixed pie and a zero-sum contest? We believe that a part of the reason is what negotiators construe the interdependent, mixed-motive task to be. People make their action choices based on what they believe the task is. In other words, they behave in ways they believe they ought to in order to do well. And because people decide how to act based on mental representations of a reality they construe rather than on any kind of direct access to objective reality (Griffin & Ross, 1991; Lieberman, 2005; Ross & Nisbett, 1991), the way negotiators *construe* the interdependent mixed-motive task determines their mindsets and consequently their strategies and outcomes.

In the second part of the thesis, we develop the theory that once an interaction is construed as ‘negotiation’, the parties perceive the task to be predominantly a competition. The strategies that they need to follow therefore focus on the distributive, zero-sum elements, and leave out of focus the strategies necessary to create value. The fixed-pie and incompatibility biases, and the expectations of non-simultaneous offers, impasse and competition are logical consequences of such competitive construals. We test this theory in a series of three studies.

Part II:

Task construal, strategic choice and negotiation outcomes

In the first part of the thesis, we advanced a theory that mentalizing – understanding one's own and others' behavior in terms of mental states such as beliefs, desires, feelings and thoughts – must underpin negotiation because negotiation is about satisfying interests, and interests are motivating mental states. Mentalizing is an essential component of bargaining because it provides the critical inputs to negotiators' decision-making, namely the models (representations) of the interests of both parties and of the relevant strategic beliefs of the counterparty. In addition to facilitating a clear sense of one's own interests, mentalizing generates a mental model of the counterparty's mind that allows predicting their behavior in light of their (inferred) motivating mental states and beliefs about optimal strategic action. In other words, negotiators mentalize to infer (hidden) counterparty's mental states from their (observable) behavior, and to decide their own action by inferring the likely reaction of the counterparty (based on their modelled motivational and epistemic mental states).

This model suggests that individual differences in mentalizing should predict both value creating and value claiming aspect of negotiation. We tested these hypotheses in three studies and found support for the theory. In Study 1 (p. 103), mentalizing facilitated both value creating and value claiming of trained negotiators in a multi-issue negotiation task. In Study 2 (p. 145), mentalizing correlated with successful value claiming in a purely distributive negotiation, and in Study 3 (p. 154) mentalizing assisted the resolution of partisan perception disputes.

One perhaps worrying finding of our studies is that people without specialized training perform rather poorly in negotiation. In other words, they do not use their arsenal of social-cognitive capacities as well as they could and should. This is consistent with the evidence that shows we are rather ineffective negotiators, predictably and dramatically failing to generate value in situations where such value is available (Deutsch, 1973; Pruitt & Rubin, 1986; Raiffa, 1982, 2002; Walton & McKersie, 1965). We discussed how we reach no deals where good deals are possible, and the deals we do close are not great (see *Ineffective mentalizing and negotiation processes*, p. 78).

This inefficiency is a serious problem because it incurs significant personal and social costs, wastes resources and productivity, and increases conflict (Pruitt & Rubin, 1986). It is also paradoxical considering the sophisticated, primarily collaboration-oriented cognitive capacity of the human brain (Sánchez-Amaro et al., 2019; Tomasello et al., 2012), our default intuitive predilection for collaboration (Levine et al., 2018; Rand et al., 2012), and the common sense notion that negotiators presumably want to maximize gain (even if for selfish reasons: a slice of a big pie is bigger than a slice of a small pie).

Approaching the problem of the pervasive suboptimality of negotiation results from the perspective of individual differences will always account for only a small proportion of the variance given that the challenges for rational negotiation are universal and may therefore be definition poorly captured in individual difference designs. What universal modifiers of the human cognitive approach to tasks can be powerful enough to dampen the advantages associated with cognitive perspective taking, empathy, and mentalizing? In other words, given the human ability for cooperative social thinking and action (see pp. 42-47), why are negotiation outcomes so poor?

In this part of the thesis, we focus on the impact of the negotiators' situational sensemaking on their ability to bargain effectively. The theoretical proposal we advance and test experimentally is that perceiving (construing) an interdependent task with value potential as a 'negotiation' triggers the understanding that the task is a zero-sum distribution, focusing on competitive tactics, and leaving out of focus the strategies that lead to value creating. Put simply, the competitive mindset and the consequent poor outcomes are relatively independent of the objective task parameters; people are more likely to approach an identical task competitively if they feel they are 'negotiating' it. In the next section we outline our approach to this question.

PRIOR EXPECTATIONS, SENSEMAKING AND STRATEGIC CHOICE

Research investigating suboptimal outcomes in negotiation highlighted the role of expectations negotiators bring to the table. People tend to assume their interests are in conflict or incompatible with the interests of the other side (the 'fixed-pie' and incompatibility biases; Bazerman et al., 1985; Bazerman & Neale, 1992; Pinkley et al., 1995; Thompson, 1991; Thompson & Hastie, 1990; Thompson & Hrebec, 1996), and expect sequential issue settlement, impasses, and competitive behavior (O'Connor & Adams, 1999). These expectations affect both information exchange and information processing (Pinkley et al., 1995), essential for successful outcomes (Pruitt & Lewis, 1975; Thompson, 1991).

Other theoretical accounts highlighted the importance of the parties’ perception of the negotiation task, suggesting that it affects their approach and behavior including the zero-sum assumptions discussed above ('mental models'; Bazerman, Curhan, Moore, & Valley, 2000; 'problem-solving' vs 'adversarial' orientations; Menkel-Meadow, 1983). In addition, the level of information exchange and cooperation in negotiation appears to be a function of trust (Butler, 1999; Gunia, Brett, Nandkeolyar, & Kamdar, 2011; Kimmel, Pruitt, Magenau, Konar-Goldband, & Carnevale, 1980; Kong, Dirks, & Ferrin, 2014).

Cooperative and competitive choices in mixed-motive (Schelling, 1980) negotiations have been compared to social dilemmas, explaining suboptimal outcomes and inefficient information exchange (the 'negotiator's dilemma', Lax & Sebenius, 1986b; the tension between creating and claiming value, Mnookin, 2000; the 'information dilemma', Murnighan et al., 1999).

Finally, a series of studies by De Dreu and colleagues looked at the quality of agreements as a function of exchanging and processing information, which in turn depend on the social and epistemic motivation of negotiators (the 'motivated information processing model of negotiation'; De Dreu, Beersma, Stroebe, & Euwema, 2006; De Dreu, Koole, et al., 2000; De Dreu, Weingart, et al., 2000). Their studies show that social motivation affects the type of information in focus, and the epistemic motivation determines the quality and depth of information processing.

The common unarticulated theme of the above research is that the way negotiators think and behave in interdependent negotiation settings, and the outcomes they achieve, depend on how they *make sense* of situations the literature calls 'negotiations'. The authors cited above have not explicitly framed their research as being about sensemaking of the negotiation game. However, their findings, jointly with insights from theory on perception from social psychology and computational neuroscience, are helpful in developing the predictions and theory of this part of the thesis.

In the following chapters we outline and experimentally test the theoretical proposition that perceiving an interdependent situation as 'negotiation', other things being equal, triggers a distribution-focused competitive mindset that causes the parties to focus their cognitive capacities predominantly on the distributive aspects of the interaction. This results in competitive strategies and suboptimal outcomes. A consequence of this task-construal is also a low level of epistemic trust, which distorts information exchange, and is relatively independent from construing the situation as a competitive interaction (i.e., it is possible to trust someone in a competitive setting such as a zero-sum 'price' negotiation as much as it is to mistrust another in a cooperative scenario such as the stag hunt, see pp. 44-46). Cognition focused on competitive distribution lacks the processes required to generate solutions to problems presented by situations with hidden value potential (e.g., multi-issue differential payoff matrices, deal optimization through creative options) and leaves out of focus the specific information required to facilitate such solutions.

Given the pervasiveness of negotiation in human affairs and the value at stake, the suboptimal outcomes in negotiation are an issue of considerable practical importance. Understanding how and why suboptimality occurs may provide insight into how to tackle this through process-interventions and training. Negotiators' competitive interactions yielding poor outcomes also raise important theoretical questions. The ability of humans to cooperate underpins our dominance of the planet and our scientific and cultural achievements. Evolutionary accounts suggest that the human default attitude is collaborative rather than competitive (e.g., Moll & Tomasello, 2007;

Tomasello et al., 2012). What is it about a ‘negotiation’ that changes our default attitude?

In the following sections, we conduct a brief review of the negotiation literature on the topic of mental models in negotiation. We then outline the social psychological approach to prior expectations before moving on to explore the role of prior expectations under the active inference framework, which treats the human brain as a hypothesis-testing machine. Because hypotheses that have priority in terms of being tested (by matching those with the incoming sensorium) in uncertain situations are largely a function of priors, this theory is particularly effective in explaining why prior expectations exert such an imperative dominance in negotiation. We continue by discussing to what extent these findings can inform negotiation theory, in particular why priors in negotiation – such as the zero-sum and incompatibility assumptions – carry such disproportionate weight and remain robust and resistant to correction. We then briefly review the existing relevant literature. This is followed by an overview of a series of studies we conducted to investigate whether and to what extent the strategies negotiators employ and the key components of dyadic value in negotiated agreements can be affected by experimentally manipulating the perception of the task (by labelling it ‘negotiation’ versus ‘problem-solving’ or ‘deal-design’).

In the discussion, we integrate the findings of these studies. We suggest that how negotiators make sense of (construe) the interdependent mixed-motive situation profoundly affects and determines their mindsets, and that construing a task as ‘negotiation’ results in competitive strategies and suboptimal outcomes. There is something about being in ‘negotiation’ that makes us competitive. We discuss the implications in the general discussion at the end of the thesis.

Sense-making and mental models in negotiation

> *“An important emerging feature of research on negotiation is the study of how players define and create the negotiation game—both psychologically and structurally. Interdependence theory explored the ways in which social actors transform the given matrix of*

> *outcomes into an effective matrix by their own personal interpretations, relationship-specific motives, and social norms.*
>
> *"[We suggest] that how parties understand the game is a critical determinant of how they play the game. To give rational advice, we need to understand the actual preferences and mental models of negotiators, rather than simply inferring that they accept the utility structure that an experimentalist provides. Understanding how negotiators differentially define the game may be key to better understanding why parties do not reach agreements when we think they should."*
>
> Bazerman et al. (2000, pp. 286 - 287)

In the field of game theory, Nalebuff and Brandenburger (1996) suggested that it was more important to focus on understanding the way the players define the game they are playing than on the moves they make. There is causality there: the way the players make sense of what the task requires determines what they do in the task to do well.

In the text below, we review how researchers in the field of negotiation historically considered and investigated the way negotiators make sense of the negotiation 'game', i.e., how they mentally model the interdependent, mixed-motive interaction with hidden value potential. This is important to set up a backdrop for our contention that the way negotiators construe a negotiation situation determines their expectations, which in turn bias them towards competition.

A terminology caveat is in order, though. Most of the research below did not explicitly focus on, or label their variables, situational sense-making, construals or mental models. However, the mechanisms that were under investigation by various approaches were indeed cognitive structures the negotiators make use of to make sense of the ambiguous and uncertain task that is negotiation. After reviewing the research, we turn to the key biases that determine negotiation behavior and inhibit creation of dyadic value – the incompatibility and zero-sum biases – before considering the psychological aspects of expectations and, further on, the active inference framework.

Early research on mental models in bargaining

Early research in social psychology of negotiation focused mainly on individual differences and structural variables relevant to negotiation settings (for a review see Bazerman et al., 2000; Rubin & Brown, 2013). The individual differences approach, covering predominantly demographics and personality traits, yielded little explanation of negotiation behavior, strategies and outcomes (Ross & Nisbett, 1991; Thompson, 1990b; Thompson & Deharpport, 1998). More importantly, any effects of the individual differences were sensitive to changes of the situation: small changes in how the background, task and objectives were perceived dwarfed most effects of individual differences (Ross & Nisbett, 1991; Thompson, 1990b; Thompson & Deharpport, 1998). Structural variables, such as the impact of incentives and payoffs (S. Axelrod & May, 1968), whether there are deadlines or not (Pruitt & Drews, 1969) and the number of players (Marwell & Schmitt, 1972), were found to be more robust predictors. It is worth stressing that this older, social-psychological investigation into the impact of the structure on bargaining behaviors and outcomes has not focused on the impact of construal of that structure or task, which rendered the results relatively useless and consistent with the naïve, intuitive 'folk' understanding (Bazerman et al., 2000, p. 281).

Behavioral decision research on mental models in bargaining

In the last two decades of the previous century negotiation scholars focused on 'behavioral decision research', consisting of both prescriptive and descriptive theory. The prescriptive approach focuses on normative tools, such as the analytical frameworks and behavioral strategies that lead to optimization rather than 'satisficing', i.e., accepting the first satisfactory and sufficient solution (Simon, 1955). Descriptive research, on the other hand, focuses on investigating and amassing data on how people behave in real life situations (that is, how they 'satisfice' rather than optimize). The wisdom of the combined approach is that the suboptimal behavior oftentimes occurs without explicit awareness, and the knowledge of which situations make negotiators particularly vulnerable to bias in order to recognize them, is necessary for people to be able to suspend intuitive judgment and activate normative behaviors advised by the prescriptive theory (e.g., descriptive theory will show that people's choices are affected by framing, so negotiators are prescribed to conduct a comprehensive expected value based analysis before making their choice).

The behavioral decision research is largely based on the findings of Nobel laureate Kahneman and Tversky (who in turn built upon the early research of another Nobel laureate, Herbert Simon). It highlights the bounded rationality of negotiators who, due to cognitive constraints, rely on simplifying heuristics rather than conduct full rational analysis. Because of that, they satisfice rather than optimize. The founders of this approach include Howard Raiffa (e.g., his seminal work; 1982), who grounded his mixed descriptive-prescriptive approach in mathematics and game theory, Bazerman and Neale, who used the frameworks of cognitive psychology to identify value optimizing behaviors in the face of specific instances of negotiation 'irrationality' (Bazerman & Neale, 1991, 1992; Neale & Bazerman, 1992a), and Thompson, who investigated the impact of unconscious assumptions about the structure of negotiation and negotiators (Thompson, 1990b; Thompson & Deharpport, 1994; Thompson & Hastie, 1990; Thompson & Hrebec, 1996; Van Boven & Thompson, 2003). This research found, among other things, that negotiators are more concessionary under a positive frame, are badly affected by anchoring and availability biases, tend to be overconfident and overoptimistic about their outcomes, like to escalate conflict beyond what rational analysis recommends, are bad at perspective taking or rather fall prey to focusing failures, and reactively devalue concessions made by counterparties (for a detailed review of this literature see Bazerman et al., 2000). Finally, and critically, negotiators tend to assume negotiations are zero-sum affairs and that their preferences are incompatible with the preferences of the other side, which we have reviewed in detail earlier in the thesis (*Negotiator cognition and irrationality*, pp. 63-65).

Social psychological research on mental models in bargaining

Although the decision perspective research had a valuable effect on the research and practice of negotiation, it has been criticized for neglecting the psychological aspects of the interaction (Greenhalgh & Chapman 1995). In the next couple of decades, social psychological research made its return to the negotiation field and used the social aspects of the interaction as specific research topics. The key topics are summarized here (for a detailed review of this literature see Bazerman et al., 2000). First, the research focused on *social relationships in negotiation.* The perspectives considered were how judgment and preferences are affected by social settings (e.g., Clark & Chrisman, 1994), how social relationships within the negotiating pairs affect the processes and results (e.g., Valley, Neale, & Mannix, 1995) and the effect of the relationships on the constituencies and other factors (e.g., Baker, 1984). The second key topic was *egocentrism in negotiation*. The key aspect these studies highlighted was that the negotiators' opinions about what is fair are far from objective. Instead, negotiators are biased by self-interest (Babcock & Loewenstein, 1997; Diekmann et al., 1997). The third aspect was the topic of *motivated illusions in negotiation*, highlighting how negotiators view themselves significantly more positively than reality can warrant (Taylor & Brown, 1988). In negotiation, for example, they feel they will get outcomes far better than possible given the ordinary and normal distribution of the outcomes (Kramer, Newton, & Pommerenke, 1993). The fourth topic was *emotion and negotiation*, where the researchers investigated the impact of specific emotional states on negotiation. For example, good moods increase cooperation (Forgas, 1998) and improve outcomes (Carnevale & Isen, 1986). Anger, on the other hand, was found to be counterproductive (Allred, Mallozzi, Matsui, & Raia, 1997).

Greenwald and Banaji's (1995) term 'implicit social cognition' and their priming approach (allegedly activating pre-conscious concepts) also had a considerable effect on negotiation research. It was employed to investigate independent and interdependent orientations, power and gender and found that relatively fine primes significantly affected the negotiators' priorities and behaviors (Howard, Gardner, & Thompson, 2007) and power (Magee, Galinsky, & Gruenfeld, 2007). Two important studies (already mentioned in the literature review, see pp. 93-97) investigated the impact of empathy and perspective-taking on dyadic and individual value in negotiation (Galinsky et al., 2008; Gilin et al., 2013). In terms of training of negotiators, Nadler and Thompson (2003) looked at which training is most effective in improving negotiation skills and found that didactic training was heavily outperformed by analogical training. Looking at the effects of mental models on outcomes Van Boven and Thompson (2003) found that negotiators who reached optimal settlements had mental models that reflected greater understanding of the negotiation's payoff structure, and of the logrolling processes than their peers who failed to optimize, and that experience-based training outperformed instruction-based training.

Further, a line of research considered the impact of the social orientation of negotiators. The crucial difference between proself and prosocial oriented negotiators pertains to whose outcomes they predominantly care about: proself-oriented bargainers care mainly about their own and prosocial-oriented endorse joint outcomes. Social motives affected the strategy the negotiators pursued; as expected, proself negotiators employed mainly distributive strategies and reached worse outcomes than their prosocial-oriented peers, who employed integrative behaviors (Beersma & De Dreu, 2002; De Dreu, Giebels, & Van de Vliert, 1998; De Dreu, Koole, et al., 2000; De Dreu, Weingart, et al., 2000; Giebels, De Dreu, & Van De Vliert, 2000). In groups, prosocial parties paid more attention to the other negotiators' behavior and interests than proself-oriented negotiators (Weingart, Brett, Olekalns, & Smith, 2007).

Finally, DeDreu and colleagues developed a two-factor model of the negotiators' ability to optimize outcomes based on their social and epistemic motivation. Optimal outcomes are more likely in negotiators who have a prosocial orientation and high epistemic motivation, that is, the negotiators who care about joint rather than solely individual outcomes, and are intrinsically motivated to understand the negotiation situation, the payoff matrix, and the counterparty (De Dreu et al., 2007; De Dreu et al., 2006; Ten Velden, Beersma, & De Dreu, 2010).

Summary

In summary, negotiation research considered the impact of mental models on negotiation processes and outcomes, although this was rarely done under the name 'mental models'. The key findings that are relevant for the purposes of this thesis are the competitive expectations the negotiators bring to the table and the zero-sum (fixed-pie) and incompatibility biases – that the negotiators assume the preferences of the parties are in conflict – that inhibit the value-creating aspects of the bargaining process. The key proposition of the present thesis is that the way the parties construe the interdependent, mixed motive situation (that we normally call 'negotiation') exerts decisive influence on their strategies and outcomes.

In the following sections we will review how prior expectations influence perception, first, briefly from the perspective of social psychology, and then, more in depth, using the active inference (free energy minimization) framework.

Perception and prior expectations in psychology

There are good reasons for us to assume that expectations exert critical influence on social behavior. Below we review the key aspects of research on the constructive nature of perception and the impact of expectations in psychology, providing the theoretical background for our claim that construing an interaction as 'negotiation' triggers a set of expectations that inhibit value-creating and promotes competitive value claiming.

Constructive nature of perception

People decide how to act based on mental representations of a reality they construct using prior knowledge and situational input, rather than on any kind of direct access to objective reality (Griffin & Ross, 1991; Lieberman, 2005; Ross & Nisbett, 1991). This is true for all perception, non-social and social. Evidence for the active nature of perception is particularly clear in the context of ambiguous social settings. Research shows that perception is constructive and affected by goals, needs and expectations (Bruner, 1957), for example temporarily (Förster & Liberman, 2007; Higgins et al., 1977) as well as chronically accessible constructs (Bargh et al., 1986; Bargh & Thein, 1985; Higgins, 1996), exposure to material objects (Kay & Ross, 2003; Kay et al., 2004), scripts, schemas and knowledge structures (Ross & Nisbett, 1991), among others.

Although social perception is actively (albeit implicitly) construed, we do not feel it to be self- generated, but instead experience it *as reality*, i.e., as if we had a direct, unmitigated access to whatever is out-there (thus ignoring the fact that the brain is skull-bound with limited available sensorium that needs to be cognitively processed for any kind of perceptual experience to take place). We experience an aggressive person, not a person we decided was aggressive (Hastorf & Cantril, 1954). This contributes to the stability and obdurateness of perception, because a review by controlled cognition (Evans & Stanovich, 2013) does not feel necessary. Maladaptive consequences, such as the fundamental attribution error, have been explored under the heading of *naïve realism* in the context of social conflict and misunderstanding (Griffin & Ross, 1991; Robinson et al., 1995; Ross, 1995; Ward et al., 1997), and studied in the earlier parts of this thesis (pp. 84-87).

Expectancies in social psychology

If an organism is to survive, it must, at some level at least, understand and be able to predict its environment. Prior expectations, or expectancies – 'beliefs about a future state of affairs, subjective estimates of the likelihood of future events ranging from merely possible to virtually certain' (Roese & Sherman, 2007, p. 91) – inform and constrain behavioral choices so that they lead to the desired outcomes. An evolutionary actor with a defined system for anticipating opportunity and danger is in a vastly superior position compared with a competitor who does not possess such a mechanism.

How do expectancies work? Roese & Sherman offer driving as an example (2007, p. 92). Chauffeuring a car involves continuous processing of various differential streams of exteroceptive and proprioceptive stimuli (e.g., the road, signposts, surrounding traffic, the speed of one's own vehicle, the position of the foot on the pedal and the arms on the steering wheel, and so on). The perception and action in driving consists largely of continuously comparing these stimuli to their desired counterparts (i.e., current speed versus desired speed) and attempting to correct any deviations (absence of such correction, such as falling asleep behind the wheel, could lead to disastrous consequences). The idea is that this quick feedback loop, labelled a TOTE unit ('test-operate-test-exit'; Miller, Galanter, & Pribram, 1960), applies not only in quick here-and-now situations like the driving described above (i.e., immediate situations involving proximate goals), but also in situations that occur over longer time periods and involve distant goals. Furthermore, the process applies more or less continuously, a phenomenon known as process fluency (e.g., Benjamin & Bjork, 1996; Johnston, Hawley, & Review, 1994; Whittlesea & Williams, 2001). In this, the key function of *expectancies* is to establish a 'set of broadly generic roadmaps for ongoing behavior' (Roese & Sherman, 2007). In other words, the expectancies provide the 'desired counterparts' in the example above, resulting in a set of behavioral policies that, on an expected value basis, lead to action that has the best odds of attaining the preferred goals.

Expectancy is a function of accessibility and can be primed. Accessibility, in the context of expectancy, refers to the probability with which the expectancy is to be applied to the present situation and consequently to affect the subsequent judgment. Expectations may be more or less accessible due to their vividness or recent activation (Bazerman, 2013; Kahneman, 2011). For example, in a classic priming study (Higgins et al., 1977), Donald's intended actions (such as wanting to cross the Atlantic in a sailboat) were perceived as adventurous or reckless, depending on the recently activated (primed) expectancy.

Expectancies are more or less implicit. This is of particular importance for the present research. Expectancies can be fully explicit or fully implicit. However, the majority are likely to be placed somewhere on a continuum between the two extremes. In most cases, an expectation becomes fully explicit (and reportable) when a social agent is asked to reflect on it explicitly. The implication is that expectancies can, and very often do, guide behavior without the agent having any idea this is so.

There is research on how expectancies can affect reality. For example, a belief in future success facilitates future success (e.g., Lewin, Dembo, Festinger, & Sears, 1944; Oettingen & Mayer, 2002; Vroom, 1964), presumably through increased performance (e.g., Locke, Shaw, Saari, & Latham, 1981), confidence (e.g., Feather, 1966) and perseverance (e.g., Battle, 1965). This is largely attributable to the increase of a positive affect that motivates positive action (e.g., Bandura & Locke, 2003; Erez & Isen, 2002) and making elaborate plans that make implementation easy and effective (e.g., Brandstätter, Lengfelder, & Gollwitzer, 2001). Expectancies also assist in avoiding misfortune: an agent anticipating a future negative state is able to act correctively today to avoid it (regret avoidance). People routinely consider future effects of decisions and actions and act in ways to minimize future regret (e.g., Zeelenberg, 1999).

Expectations can also be self-fulfilling. They may improve performance, as outlined above, or work counterproductively (e.g., a 'self-fulfilling prophecy'). Research showed the impact of stereotypes on perception and action in terms of gender (e.g., Eagly, Wood, & Diekman, 2000) and racial stereotypes (e.g., Word, Zanna, & Cooper, 1974). Crucially, the negative effects take place implicitly and without awareness of the actor (e.g., Chen & Bargh, 1997) and sometimes of the target (e.g., Vorauer & Miller, 1997).

This section summarized the key research on expectancies in social psychology. In the next section we turn to how the impact of prior expectations (expectancies) on perception and action is conceptualized in the free energy minimization (active inference) framework.

Free energy minimization (active inference) framework

The free energy minimization (active inference) framework is a relatively recent theory in mathematical neuroscience by Karl Friston and his collaborators (e.g., Friston, 2005, 2010; Friston & Frith, 2014; Friston, Harrison, & Penny, 2003; Friston et al., 2013; Friston et al., 2014; Hohwy, Roepstorff, & Friston, 2008; Moutoussis, Fearon, et al., 2014; Moutoussis, Trujillo-Barreto, et al., 2014). It holds considerable promise and is used increasingly to support theoretical predictions in experimental research as well as advance the existing theory of perception and action (Hohwy, 2014).

Before outlining the key aspects of the free energy principle, it is important to note that the theory is somewhat controversial in the literature. It's epistemic status seems to be unclear; it has been called ‘an imperative, a tautology, a stipulative definition, paradigm, law of the life sciences, law of nature, an a priori first principle, a unifying explanation, and a simple postulate or axiom’(Colombo & Wright, 2018), and a ‘normative, axiomatic and self-evidently true natural law’ (M. Allen & Friston, 2018). The main criticism is that it seems to be unfalsifiable, which puts it beyond scientific inquiry (Hohwy, 2020). Also, it is difficult to understand and impossible to use by even expert audiences (Freed, 2010). Finally, Biehl, Pollock, and Kanai (2021) questioned certain technical aspects of the free energy mathematics and highlighted the differences between older and newer formulations of the principle, but also stated that the indicated problems do not conclusively negate the value of the general ideas behind the free energy principle. While appreciating these criticisms, we do not consider them sufficiently persuasive to dissuade us from using the free energy principle (active inference) theory as our chosen framework for explaining the problem we set for ourselves, namely explaining the persistent failure of negotiators to reach optimal outcomes.

Brain as a hypothesis-testing machine

At the core of the active inference framework is the theory that the brain is a hypothesis-testing machine that makes perceptual and active inference about the hidden world. It does so by minimizing prediction error, which is done by comparing the predictions generated by prior expectations and prior inference to the (sampled) sensory input that is available. It is a powerful and attractive theoretical account that holds promise to parsimoniously explain both perception and action and is supported by many theoretical arguments as well as some empirical studies.

The free energy principle (active-inference) turns the traditional notion of perception on its head. The brain’s role in perception (and action, and everything in-between) is not to make bottom-up sense of what the incoming stream of sensory data means, but to actively sample the sensorium to see whether the incoming material conforms to what it expects to find. It is prior expectations, weighted by how likely they are to explain the sensorium if they are true, also called the ‘hypotheses’ about hidden states, that direct what kind of sensorium the brain will preferentially pay attention to. Should the sensory data fit the expectations, we will perceive what the hypothesis states; if not, the brain processes will continue the query (also using action, hence the 'active' inference framework) or discard the falsified hypothesis and substitute it for another one. The key aspects of the free energy principle date back to late 18th century and Hermann von Helmholtz's idea of unconscious inference; Helmholtz reportedly theorized, in reaction to Kant, that our perception is based in unconscious inference from the answers the world delivers when our brain queries it (Hohwy, 2014, p. 5).

Perceptual and active inference

Perception is an inference problem. The sensorium, that is, the data we can detect with our sensors, is the caused by the world out there. It is important to stress that this world is inaccessible directly. Our brain is locked in a skull and can, via exteroceptive senses, sample not the world, but the sensorium the world produces. The challenge of perception is to infer the sensorium's causes: what must the world be like so that it produces this kind of sensorium? This is a *problem* because the same cause can produce very different sensory data (e.g., a rattlesnake can hiss and rattle), and the same sensory data can be explained by many different putative causes (rattling can be a rattlesnake or a toy rattle). Two further constraints are necessary to make such selection possible: first, the *likelihood*, which is the relationship between the sensorium and its likely causes (how likely is it that the hypothesis, if true, explains the detected sensory signal, e.g., how well does a rattlesnake explain the rattling), and second, *prior expectations*, independent, *a priori* plausibility assessments of each hypothesis (e.g., how likely are we to encounter a rattlesnake where we are). Perceptual inference thus works according to Bayesian conditional probability considering both priors and likelihood to arrive at the hypothesis with the highest posterior probability, which is what we then perceive as 'reality'.[16]

[16] The inferential processes involving priors and likelihoods lend themselves to Bayesian conditional probability calculation. Perception involves maximizing the hypothesis' *posterior* probability by minimizing the prediction error arising from the mismatch between predicted and actual sensory feedback. Posterior probability in a simplified version of the Bayes theorem equals the product

Action (active inference) assists perception by exerting an influence on the external (hidden) world so the brain can compare the resulting sensory input to the one predicted by the current hypothesis (the one that has the highest posterior probability). For example, looking at a large canvas we spy something that looks like an eye, which makes the brain infer that we are most likely looking at the picture of a face. Based on that hypothesis, the prediction is that if we move our gaze (action) in a saccadic motion down from the 'eye' we should encounter a nose and then a mouth (Friston, 2012); if we encounter these further facial features, we perceive ('see') a face; if not, the brain supplants the primary hypothesis in favor of a different one (perhaps it is a picture of an eye only). Active inference is particularly useful in situations where the initial sensorium is ambiguous or uncertain so that perceptual inference alone cannot determine a clear favorite hypothesis or test causality. These mechanisms work on various levels of cortical hierarchy and use prediction error and expected precisions as the key parameters (refer to the Appendix for a more detailed description of these processes, p. 290).

Cognitive penetrability of prior expectations

As perception depends on priors, do we not, at least to some extent, perceive what we expect to perceive? In other words, does cognition penetrate perception? This turns out to be relevant in situations where the sensorium is ambiguous in the sense that perceptual processes cannot discriminate between a number of competing hypotheses (if the situation is clear, the sensorium constrains cognition because outlandish hypotheses return a large prediction error and are discarded). If the situation is uncertain,[17] disproportionate weight is given to prior expectations and cognition thus penetrates perception.

of the priors and the likelihood: $P(hi|e) = P(e|hi) * P(hi)$, where *P(hi|e)* stands for the posterior probability of the hypothesis being true given the sensory evidence provided, *P(e|hi)* for the likelihood that the evidence would be produced given a certain hypothesis, and *P(hi)* the prior probability of the hypothesis itself.

[17] More precisely, cognitive penetrability of higher-level priors depends on how precise or imprecise the higher hierarchical cortical units estimate the lower-levels' prediction error signal to be. If the expected precisions are high, the sensorium (via the prediction error of lower units) will effectively

Say that during an evening stroll in the Slovenian alps we encounter a large animal. The sensory feedback is too vague to be trusted and a reliable inference cannot be made. If the animal is on our path and we *must* make an inference (i.e., see what it is), the perceptual processes will assign priors increased importance. In other words, we rely on prior knowledge (e.g., the animal is most likely a cow) as the alternative would be to take a wild guess. This would then dictate further action, such as engaging active inference by looking closer, a very different reaction than if your priors were a hypothesis of a bear, in which case we would probably slowly and quietly tiptoe away. Therefore, in situations of high ambiguity or uncertainty, prior beliefs can cognitively penetrate perceptual experience.

Active inference, mentalizing and negotiation

Expectations always influence perception and action. This is critical for this thesis because it highlights that perception starts with a cognitive representation of what the reality is expected to be, and that these priors determine the specific aspects of sensorium that will be considered first in the process of perceptual and active inference.

The prior expectations are particularly important when situations are ambiguous; in such cases, they are given extra weight and can penetrate perception. This is important for two reasons. First, mentalizing is an epitome of a difficult inference problem. Mental states are observationally opaque and need to be inferred from the observable sensorium, which is not an easy feat as evidenced by everyday social misunderstandings, even between people who know each other well. Second, negotiation situations are particularly ambiguous and uncertain (De Dreu et al., 2007), and mentalizing in negotiation is particularly difficult because of the tension between creating and claiming value (see *Creating and claiming value: Negotiator's dilemma*, p. 61), the incentives that negotiators have to make their deception undistinguishable from honesty, and how difficult it is to tell truth from lies (Bond & DePaulo, 2006). In

constrain the higher units' prior hypotheses (as in the example of 'supervision by the world' above). If the prediction error signal is not to be trusted, priors will be granted disproportionate weight.

such situations, we can expect that priors will have greater sway, penetrate perception and dictate what subjective reality is. And we know that the expectations in negotiation are competitive.

We now turn to how our understanding of the nature of perception and characteristics of negotiation situations can inform our understanding of competitive cognitions and interactions in situations perceived as ‘negotiations’.

Competitive cognition in ‘negotiation’

Beliefs and expectations, in the form of priors, are likely to determine (cognitively penetrate) perceptual content predominantly in situations where the sensory input is ambiguous or uncertain (e.g., where the prediction error signal cannot be trusted). Below, we put forward a theoretical account of how such cognitive penetrability of priors influences cognition in negotiation and biases it towards competition. We start by describing the inherently high levels of uncertainty and ambiguity in negotiation settings, which is exacerbated by the difficulty in discerning (mentalizing) negotiators’ cooperative and competitive intentions. We then explain how this uncertainty and ambiguity results in negotiators overweighing their prior expectations linked with negotiation, and that these expectations are competitive. We conclude with our theoretical proposition that construing a situation as ‘negotiation’ triggers such competitive mindsets and results in higher levels contending and low levels of problem-solving, which ultimately translates to suboptimal negotiation outcomes.

This section is an attempt at integration of two fields of research, the free energy framework and the aspects of negotiation research that demonstrate suboptimal outcomes. We are asking whether the free energy model can explain the robust findings from the negotiation literature. Below, we attempt to explicate our theoretical proposition and propose empirical designs that may validate it.

Uncertainty and ambiguity in negotiation

Negotiation is a complicated, ambiguous and uncertain 'fuzzy' affair (De Dreu et al., 2007). Not only is the situation itself riddled with information asymmetry, uncertainty and ambiguity, the critical inferences that negotiators must make relate to the *mental states* of their counterparts, and mental states are a paragon of the observationally opaque. More importantly, because of the mixed motives of the parties – negotiators experience simultaneous motivations to both compete and collaborate – and given the possibility of deception, coupled with our inability to detect lies, it is particularly difficult for negotiators to infer the mental states of their counterparties with any level of confidence. We review these two features of negotiation below.

Inherent uncertainty and ambiguity of negotiation situations

Negotiators do not know their counterparty's interests (motivating mental states). With analysis and preparation they may get a sense what these may be, but they cannot know for sure, and what they do know is far more vague than what is required for optimal decisions (i.e., they are unlikely to know specific preferences or valuations such as the reservation value, which are key for a successful creating and claiming of value). In fact, the negotiators' own interests require a certain dose of introspection and self-focused mentalizing as most interests are not obvious and require conscious reflection to explicate them (see p. 70). The parties are therefore entering the negotiation half blind: while they (hopefully) have a reasonably solid understanding of their own interests, their counterparties' mental states are an informed guess at best. The implication is that the negotiators cannot know the overall payoff structure of the negotiation task and therefore do not have the information necessary for generating compatible and integrative gain (see p. 51). Consequently, they have no idea what the hidden value potential is or, in fact, whether there is any value to be created at all. Finally, both the focal negotiator's and the counterparty's alternatives to the negotiated agreement – needed to determine the bargaining range and decide whether there is a deal to be made at all – are often uncertain (e.g., an opportunity cost of investing in an uncertain venture, a job interview, a course of action subject to a number of contingencies), and unknown to the other party.

This information asymmetry can be mitigated by communication, but this communication is besieged by the possibility of deception and the negotiator's dilemma. From the perspective of an individual negotiator, the counterparty's communicational intent is highly opaque; are they cooperatively telling the truth, or competitively 'shaping perceptions'? We address this in the following section.

Exacerbated opacity of negotiators' intentions

Competition looks like cooperation

Individual outcomes in negotiation are interdependent and in all aspects of the interaction negotiators base their behaviors on the perceived and expected behaviors of their counterparties. To make these choices, it is imperative that they can differentiate between collaborative and competitive intent in their counterparties; the complex strategic decision making in mixed-motive scenarios depends heavily on predicting the moves of the other. Mentalizing the counterparty's intent is particularly difficult in such settings because virtually all value-claiming strategies *look like* value creating. In other words, the value claiming negotiator will be motivated to make themself appear to be a value creating negotiator – or else the deception will not work. Lax and Sebenius, for example, in their largely prescriptive treatise *3D Negotiation* (2006) suggest that to effectively claim, a negotiator must shape the counterparty's perceptions of what is possible (see p. 50 and 75): you will get the best value if the counterparty believes you *really* will not pay more than a certain price. Value claiming only works if the counterparty does not think you are claiming, but that you are genuinely informing them of your true interests, reservation values, alternatives, and such. For example, a statement 'I really cannot offer you more than a thousand for this item' may be motivated by an honest, value-creating intention to inform the counterparty of a realistic constraint and thus facilitate a deal within the bargaining range, or by a deceitful (the 'shaping perceptions' type) intention to mislead about the reservation value and thus claim a large portion of the available value. Value creating and claiming are, from the perspective of the recipient, virtually indistinguishable, and this is caused by a particularly bad case of the opacity of the negotiators' key mental states in negotiation.

To catch a liar

There is no easy way of catching a liar in negotiation. One of the key findings from research on deception is that the behavioral tell-tales are weak and scarce; there are no systematic signs that accompany deception (Depaulo et al., 2003; Sporer & Schwandt, 2007). Meta analyses show that human lie detection is only slightly better than chance alone (Bond, 2008; Bond & DePaulo, 2006; Hartwig & Bond, 2014).

In a meta study, Bond and DePaulo (2006) analyzed 206 studies comprising 24,483 people judging deception in real time with no special machinery (e.g., polygraph) or training. On average, only 54% of lie-truth judgments were correct (classifying 47% of lies as lies and 61% of truths as truths). The main criticism of this research was that because the participants were college students lacking motivation to deceive and any emotional investment in the result, the effects cannot possibly generalize. In other words, motivation and strong emotion may be moderators of lie detection. If liars are invested in their lie, that is if they sufficiently care about the content and success of their lying, cues to deception will leak through (the 'leakage hypothesis'; Porter & Brinke, 2010). If the stakes are high enough, the deception will become more detectable (Frank & Ekman, 1997). In response, Hartwig and Bond (2014) conducted a meta study of detectability of lies from multiple cues in 144 samples consisting of 9,380 judges of truths and lies who opined on 26,866 truthful or deceptive messages and found that neither motivation nor emotion mattered, thus lending generalizability to the prior studies.

In sum, it is hard for negotiators to detect deception. This makes negotiation all the more ambiguous and uncertain. Having no other choice, negotiators rely on priors. We explore this in the next section.

Ineffective perceptual and active inference cause overweighing of priors

How can negotiators, given these constraints, make critical negotiation decisions at all? The ambiguity and uncertainty of the key aspects of the negotiation situation, coupled with the intentional obscuring of motives in the case of value-claiming and by our inability to detect deception, make perceptual and active inference of the counterparty's mental states exceedingly difficult.

The statement from the example above 'I really cannot offer you more than a thousand for this item' *can* be explained by two hypotheses: first, the counterparty is making either an *honest attempt at value-creating* within the constraints of the bargaining range, motivated by a cooperative intention, and second, the counterparty is mounting a *deceitful attempt at value claiming*, motivated by a competitive intention. Both the cooperative-intent and the competitive-intent hypotheses explain the sensorium equally well, that is, they both have high likelihoods. However, because of the exacerbated opacity of the negotiators' mental states, perceptual inference alone cannot determine which hypothesis is correct.

Active inference (involving action) cannot not distinguish between the two competing hypotheses either. Asking questions and taking action attempting to clarify the counterparty's motive are subject to more or less the same constraints as perceptual inference based on the sensory input alone. A deceitful counterparty will do their best to look and sound truthful, answer questions, be helpful and understanding, and in general do all else as if the information they were imparting were truthful. And so will the honest one.

In ambiguous situations where the prediction error signal is unreliable, the brain relies heavily on prior expectations – the alternative is blind guessing (see p. 195 above). For the reasons above, this applies heavily in negotiation (at least in the communication transactions that involve critical information). So once given decisive weight, negotiation-activated priors will influence perception. And there is plenty of evidence demonstrating that prior expectations in negotiation favor competitive hypotheses.

Expectations (priors) in negotiation are competitive

The expectations and assumptions people bring to the negotiating table – that is, their priors – are predominantly competitive. The most salient assumptions in negotiation relate to the most uncertain and ambiguous aspects of the bargaining interaction: the preferences and intentions of the other side, the payoff task, and the negotiation process. People expect the counterparty and the interaction to be competitive and likely to end in an impasse (e.g., O'Connor & Adams, 1999), and that the task is a fixed-pie where their interests are incompatible with the interests of the other side (e.g., Thompson, 1991; Thompson & Hastie, 1990; Thompson & Hrebec, 1996). When perceptual and active inference fail because of uncertainty and ambiguity, relying on negotiation priors results in perceiving the counterparty as a competitor and the situation as a zero-sum distribution. For example, while the above statement ('I really cannot offer you more than a thousand') may be motivated by either a collaborative or competitive intent, negotiation priors will color it as competitive. More generally, while an action in an interdependent situation can be collaborative or competitive, in *negotiation* it will likely be *perceived* as competitive.

Competitive priors affect mentalizing

The inference that is affected in the manner described above in negotiation is predominantly mentalistic inference. The critical inputs to the negotiators' decision making about optimal action to satisfy their own interests are the counterparty's mental states about their interests and their beliefs about optimal action (and iteratively their beliefs about the focal negotiator's mental states, and so on, for detail see *The model*, pp. 68-77). Mentalizing may be a critical skill in negotiation but is subject to a systematic competitive bias caused by the situational uncertainty and ambiguity, the difficulty of inferring the counterparty's intentions, and the consequent overweighing of the competitive priors.

Construing a situation as 'negotiation' triggers competitive priors

This brings us to an interesting observation: because the parties have no direct access to reality, but instead only to their individual construals of reality, the activation of competitive priors in mixed-motive interactions with value potential (that is, in situations the literature calls 'negotiations') should depend on the parties recognizing the context of the situation as 'negotiation', rather than on the objective features of the situation alone.

This is directly connected to the paradox of suboptimal results in negotiation that has been puzzling researchers for the past 50 years. Objective features of the negotiation situation allow the generation of joint value. Negotiating parties are presumably motivated to maximize these gains because their individual value is a direct function of joint value (even purely selfish negotiators would prefer a lion share of a large pie over a lion's share of a small pie). And yet, negotiators systematically and predictably fail to capture such value.

In other words, a significant part of the suboptimality problem may be not the asymmetric payoff matrix or the task structure, but the way people construe what the task they are facing is. In the next section we articulate this theoretical proposition that perceiving the situation as 'negotiation' triggers the competitive priors and outline the consequences, and present the hypotheses that are tested in the studies that follow.

The proposition

Our proposition is that the negotiation interaction that yields suboptimal results is the result of negotiators' competitive mindsets, which are driven by their competitive prior expectations, which are in turn triggered once the parties recognize that they are in a 'negotiation'. In other words, once the parties construe an uncertain and ambiguous interdependent mixed-motive situation as a 'negotiation', the associated competitive priors cause them to perceive that their *task is a competitive distribution of a fixed resource,* and that their *counterparty is an untrustworthy competitor* (epistemic mistrust).

Normative strategy is to contend

The perception that the task is a competition informs the parties' thinking and behavior. Given that the situation is zero-sum and the task is to divide, the relevant strategies are to contend, yield and compromise (Pruitt, 1983b): one needs to try to win as much as possible (contending), and if that is not possible, meet in the middle (compromise) or, in the worst case, give in (yield). The normative strategy in a competitive task is to contend, and yielding and compromising are not really active strategies as much as instances of ineffective contending and the plan B. It is important to stress that favoring these strategies results from the perceived task and the payoff matrix: if a dollar more for me is a dollar less for you, and we are only negotiating dollars, then these three strategies are the only applicable strategies there are. Problem-solving is irrelevant as there is no problem to be solved (except perhaps how to make the other side agree to give in as much as possible).

Value optimizing is out of focus

Bounded awareness of value potential

Emphasizing distribution leaves out of focus the information and strategies required to create value. The tendency not to be aware of things we are not actively looking for is well documented in the domain of general cognition. A well-known set of visual awareness studies showed that if people focus their attention on a specific, not particularly difficult, but attention-demanding task, such as counting passes between teams of basketball players, a large number of observers will simply fail to notice an unexpected event such as a chest-thumping gorilla walking through the basketball court (Simons, 2010; Simons & Chabris, 1999). Similar tendencies have been observed in general decision making, a phenomenon known as 'bounded awareness' (Bazerman, 2013; Bazerman & Chugh, 2006; Chugh & Bazerman, 2007). In addition to the 'winner's curse in strategic settings' (situations where an uninformed anchoring attempt in the presence of information asymmetry results in a losing outcome for the party who makes any offer (see the example in *Mentalizing in value claiming*, pp. 75-77), bounded awareness is also present and results in the winner's curse in competitive auctions (where the presence of multiple bidders bidding for a commodity of uncertain value virtually always results in the ‘winner’ substantially overpaying for the commodity, e.g., in auctions for leases for drilling oil; Bazerman & Neale, 1992) and M&A markets (e.g., where the share price of the 'winners' of takeover contests underperforms when compared to the share price of the 'losers' by up to 50%; Malmendier, Moretti, & Peters, 2018). Evidence for bounded awareness is also found in people’s performance in fictional decision making problems such as *Acquiring the company* (based on the 'adverse selection' phenomenon; see Akerlof, 1970, 1982), and spotting the difference between *Monty Always Opens* and its *Mean Monty* variation (Bazerman & Chugh, 2006; Chugh & Bazerman, 2007), among others.

A similar focusing failure likely occurs once a situation is understood as a ‘negotiation’: the parties’ judgments about the action that is an optimal fit to the situation (normative strategies) and the relevant information they should seek and provide, and their understanding (and prediction) of the behavior and the information provided by the counterparty, are processed in the *context of competing for a share in a fixed pie*. For example, a statement ‘I really care about the salary’ is much more likely to be understood (and intended) as posturing or an attempt to mislead (e.g., to later trade it for a concession on another issue, more valuable to the counterparty) than as a piece of information that may help in crafting a solution that optimizes dyadic gain through logrolling or identifying a compatible issue. Because the interaction is ambiguous and the information asymmetric, such beliefs are relatively resistant to falsification (e.g., the counterparty’s logrolling offers have no impact on the fixed-pie bias; Moran & Ritov, 2002).

Similarly, the focus on distribution is likely to frustrate the identification of issues where the parties’ preferences are identical. As highlighted by the negotiator’s dilemma (also 'information dilemma' or 'the tension between creating and claiming'; Lax & Sebenius, 1986b; Mnookin, 2000; Murnighan et al., 1999), it is dangerous to honestly provide information in competitive contexts. Truthfully disclosing one’s preference in compatible issues can, and often does, result in being taken advantage of. For example, Loschelder et al. (2014) showed that, contrary to the common wisdom that one should anchor in negotiation (Furnham & Boo, 2011; Gunia et al., 2013), the first-mover may face a disadvantage because the first offer may backfire when the information that the first offer carries can be taken advantage of by an astute recipient. For example, if both negotiators prefer the job location to be in San Francisco, and one of them discloses it, the other one can reciprocate by also truthfully revealing their identical preference, or instead respond by saying something along the lines of: 'I understand your preference for San Francisco, a lot of people wish to work there. It is unfortunately quite expensive for us to post you there, but we could work toward accommodating you against a concession on another issue.' The latter move makes a lot more sense in a competitive interaction, which means that the first negotiator, who can anticipate it, will be highly unlikely to expose herself to such risk with an honest statement of her preferences to start with. This is how lose-lose agreements, where the parties choose a deal where they are both worse off, take place.

fMRI studies

There is evidence in neuroscience supporting the idea that perceiving a mixed-motive situation as competitive or collaborative triggers different mindsets. Various fMRI studies show people engaged in identical tasks, but under different frames (cooperative versus competitive) activate different regions of the brain. Decety, Jackson, Sommerville, Chaminade, and Meltzoff (2004), for example, scanned the participants playing an identical computer game either in a cooperation or a competition mode with a counterparty and compared their scans to the participants playing the same game independently. They found that distinct and different brain regions were selectively recruited when individuals were in cooperation and competition (mainly the orbitofrontal cortex in cooperation and the inferior parietal and medial prefrontal cortices in competition). Lissek et al. (2008) had healthy participants reflect on the protagonists' mental states in cartoons showing instances of cooperation or deception, and a combination of these two in a situation where two characters cooperated to deceive the third. They found significant differences in activation patterns across the conditions. Among other things, deception recruited orbitofrontal and medial prefrontal brain regions. Finally, Tsoi, Dungan, Waytz, and Young (2016) investigated whether the theory of mind regions were recruited more in cooperative than in competitive settings. They scanned participants playing an identical mixed-motive game under the cooperative and competitive frames and found that while the ToM regions were recruited similarly across interaction contexts, the activated neural networks encoded information separating cooperation from competition. In other words, when people are motivated to think about mental states of others, the ToM networks encode different aspects of mental states during perceived cooperation than during perceived competition. Jointly, these studies offer indirect support to the proposition that different mindsets are activated during perceived cooperation and competition.

Summary

Just like people who do not expect to see a gorilla in the video fail to see it, negotiators who expect a competition fail to conceive of the possibility there might be value available in the situation, and do not act in value-seeking manner. For them, distributive competition is all there is.

Perceiving a situation as 'negotiation' impairs interpersonal (epistemic) trust

Perceiving a task to be a competition also has implications for interpersonal trust. A corollary of the default distribution-focused competitive mindset is the perception of the counterparty as an untrustworthy communicator. Value claiming in negotiation (e.g., anchoring, sequential offers, converging concessions, non-offer offers) mostly consists of attempts to influence the counterparty's perception of what is possible by more or less explicit misinforming (Lax & Sebenius, 2006). It is risky to disclose information and to trust the information provided by the counterparty (Lax & Sebenius, 1986b; Loschelder et al., 2014; Murnighan et al., 1999).

One aspect of the problem of inferring intentions underlying value-claiming/value creating statements such as 'I *really* cannot pay more than a thousand' (see *Exacerbated opacity of negotiators' intentions*, p. 199) is trust. Whether the recipient of this communication ought to believe its veracity can depend on whether the recipient trusts the communicator. In our chosen framework, trust is a prior expectation regarding the counterparty[18] that will affect the perception and action in instances of ambiguity. In other words, the trusting party will believe the communicator not because it can discern the veracity or intent from the counterparty's behavior and statement, but because its prior expectations are that the intent of the counterparty is likely to be benign and informative, rather than malign and competitive.

There are instances where the situation is ambiguous and the party possesses two priors that are in conflict, for example, when you are 'negotiating' with someone you trust. The priors associated with negotiating dictate that the focal negotiator is looking at a competitive task without any value to be created, and the priors associated with the counterparty suggest that they are trustworthy. Our take is that priors regarding the task and priors regarding the counterparty work to different effects. In other words, the perception of the counterparty as an (un)trustworthy communicator (epistemic mistrust) is relatively independent from the perception of the situation as a competitive, purely distributive zero-sum game. It is perfectly possible to trust the source of information in a competitive setting, i.e., in a situation that does not call for subtle problem-solving.[19] The task is still to divide, and the key strategy is still to contend, and while the process may be more courteous and less based on deception, it will still be aimed at dividing the pie rather than trying to generate value (see also p. 204). For these reasons, establishing epistemic trust alone should only partially mitigate the harmful effects of construing the situation as a zero-sum distribution. While it will likely facilitate a more efficient exchange of information and thus create the conditions under which verifying the zero-sum assumption is possible, it will not by itself modify the parties' perception of the situation as a competitive distribution nor the resulting competitive mindset.

[18] Other priors with similar effect are the prior relationship between the parties and the reputation of the communicating party.

[19] It is equally possible to distrust someone when facing a joint problem (e.g., Mnookin, 2010).

Competing abstract hypotheses are explained away

Abstract knowledge about negotiation does not seem to improve negotiation outcomes. There are many books on negotiation and the offer of negotiation training is vast, and yet the results do not seem to improve. The negotiators might well have read *Getting to Yes* (Fisher et al., 1991) and know, on some abstract level, that 'behind opposed positions lie shared and compatible interests, as well as conflicting ones' (p. 43), but this hypothesis may not be able to effectively coexist with or supplant the competitive (zero-sum) prior that already has traction with the sensory input.

The higher-level belief that negotiation can have hidden value is, when competing with the universally shared naïve hypothesis that human interactions involving conflicts of interest are predominantly about conflict (the zero-sum assumptions), may be too abstract to have effect. In Hohwy's words, it does not 'predict at the right fineness of spatiotemporal grain' and therefore 'cannot make predictive contact with the sensory input', and is, for that reason, 'probabilistically idle' (2014, p. 127). In other words, even if the negotiator knows at some abstract level that there may be value in negotiation, they cannot get this 'rather coarse (invariant) true prior' to make proper contact with the 'fine-grained (variant)' real-time negotiation interaction, involving things like posturing, more or less veiled threats and ultimatums, a range of emotions, subtle or less subtle questioning, and so on (Hohwy, 2012; 2014, p. 127; Hohwy & Rosenberg, 2005). In uncertain situations, a number of hypotheses may concurrently attempt to explain the sensorium, however once one emerges as a good fit, the activity of the others tends to dissipate (or rather, they are 'explained away'; Hohwy, 2014, p. 61). In short, new abstract knowledge may fail to influence perception if a competing prior hypothesis is already active and has good traction with the sensory input. It may be difficult for an abstract idea, however true it may be, to displace or even coexist with a prior expectation that has already been deemed good enough and explains away any competing hypotheses.

The important implication is that training needs to affect the 'fine-grained', low-level priors, rather than abstract knowledge. There is much support for this in the literature that shows this is no easy task (Patton, 2009). We will return to this question in the conclusion.

Overview of the present studies

The chapters that follow present a set of studies that experimentally test the proposition that perceiving an interdependent situation as ‘negotiation’, other things being equal, triggers a competitive, distribution-focused mindset and strategies that are appropriate for distributive tasks, but are poor at creating the available value or designing optimal agreements.

This proposition triggers many questions. For example, what causes the parties to recognize a situation as a ‘negotiation’? Is it triggered by perceiving the distributive aspects of issues, which are by far the most salient, less subject to information asymmetry, and the easiest to intuitively assess? Or is it driven by a somewhat paranoid loss aversion related to the quantum a party is interested in maximizing, and is so just an affliction of a cognitive miser? Or does it have to do with the cultural and semantic connotations of the term ‘negotiation’? These are questions for further studies. We limit ourselves to experimentally testing the impact of perceiving a task as ‘negotiation’ (compared with an alternative, more collaborative frame) on the strategies employed by negotiators and the outcomes they achieve.

We test this in three studies. In the first study (Study 4), we manipulate (i) the perception of an identical multi-issue task by labelling it alternately ‘negotiation’ and ‘problem-solving’, and (ii) the perception of the counterparty by increasing the dyadic levels of trust. In a supplemental Study 4b we investigate the impact of a personal conversation and active listening on trust and the perception of the task. In the second study (Study 5), we investigate which strategies participants choose to endorse in a multi-issue task labelled alternatively ‘negotiation’ and ‘problem-solving’. Finally, in the third study (Study 6), we investigate the impact of alternative construals (‘negotiation’ versus ‘deal design’) of a prospective corporate acquisition on both the (i) strategies used by the negotiating partners and (ii) negotiation outcomes, specifically on their ability to reach an optimal solution. This study replicates and extends the results of Study 4 using a different negotiation paradigm, and provides insight into how different construals affect the strategies used by the participants, and how these strategies contribute to the quality of negotiation outcomes.

Jointly, the three studies provide solid support for the idea that construing the task as ‘negotiation’ affects the strategies negotiators endorse and contributes to the suboptimality of bargaining outcomes.

STUDY 4: TASK CONSTRUAL AND OUTCOMES IN MULTI-ISSUE NEGOTIATION

This study tests the proposition that construing a mixed-motive interdependent situation as ‘negotiation’ triggers competitive mindsets that result in poorer joint gains than when an identical situation is construed as ‘problem-solving.’ Furthermore, increased trust only partly mitigates the harmful impact of the ‘negotiation’ construal, as the parties, while more trusting and trustworthy, are still engaging in a task that is perceived as competitive.

Task construal, trust and strategies in negotiation

Construal and strategic choice

Evidence shows that perception of an interdependent task influences the parties’ choices and outcomes in game theoretic tasks. Liberman et al. (2004) showed that labelling a prisoner’s dilemma a ‘Community game’ entailed almost double the levels of cooperation than labelling it the ‘Wall Street game’, while the players’ competitive and cooperative personality traits had no predictive value (also see Ward et al., 1997). Similar effects were reported in a prisoner’s dilemma by Batson and Ahmad (2001), Batson and Moran (1999), Eiser and Bhavnani (1974), Zhong et al. (2007), in social dilemmas by Larrick and Blount (1997) and Pillutla and Chen (1999), and in trust games by Burnham, McCabe, and Smith (2000). While suggestive, these findings do not automatically generalize to negotiation, because the interaction in strategic games is different from that of multi-issue negotiation tasks. In games, players know the payoff structure, and the interaction involves a limited number of available strategies for them to consider and predict. By contrast, multi-issue negotiation involves unrestricted social interaction, a wide range of strategies, opaque payoffs, and each party can veto any proposed agreement (Bartos, 1972; De Dreu et al., 2007; Rapoport, 1969).

While there are theoretical suggestions that the bargaining process may be influenced by how the parties mentally model, frame or approach the task (e.g., Bazerman et al., 2000; Menkel-Meadow, 1983; Olekalns & Smith, 2011), no study so far demonstrated a causal impact of task construal on negotiation outcomes. The only (unsuccessful) attempt was a study by Thompson and Deharpport (1998), where the effect of labelling was investigated as part of a complex 3 x 3 x 3 design. The authors found that, while the 'problem-solving' label affected the negotiators' self-reported expectations before the task, it had no effect on outcomes. The reason for the null result was likely the highly complex experimental design and an underpowered subsample testing the effect of labelling. Also, because the questionnaires were administered after the experimental manipulation but before the task, they might have interfered with the priming process.

Trust and epistemic trust

In social science and negotiation research, trust is usually understood as the willingness to be vulnerable to another person based on a perception of their integrity, ability and benevolence (Mayer, Davis, & Schoorman, 1995; Schoorman, Mayer, & Davis, 2007). Studies of trust in negotiation (for review see Brett & Thompson, 2016; Kong et al., 2014) show trust negatively correlates with distributive and positively with integrative behaviors (Kimmel et al., 1980; Kong et al., 2014; Yao, Zhang, & Brett, 2017), improves the duration of the relationship (Dadzie, Dadzie, & Williams, 2018) and contract implementation (Mislin, Campagna, & Bottom, 2011), and facilitates turning points associated with improved outcomes (Olekalns & Smith, 2005). Trust is also a necessary condition for deception (Yip & Schweitzer, 2015), and different types of trust encourage and inhibit different kinds of deceptive behavior (Elahee, Kirby, & Nasif, 2002; Olekalns & Smith, 2009; Zhang, Liu, & Liu, 2015).

On the question of the impact of trust on outcomes, the studies are inconclusive. While some have found a modest correlation between trust and joint outcomes (Gunia et al., 2011; Kong et al., 2014), others have not: in a study directly measuring the impact of trust on integrative outcomes, Butler (1999) found no effects; Kimmel et al. (1980) showed that trust increased cooperative behavior, but did not predict integrative gain; and Sinaceur (2010) found that wholly-trusting dyads did no better than distrusting ones.

Because information exchange is an essential component of both creating and claiming value, and is core to the tension between the two, a narrow concept of information-focused trust is adopted by this study: *epistemic trust*, a belief that the trustee will impart information that accurately reflects reality. Accounts in developmental psychology and evolutionary theory suggest that epistemic trust underpins mental health (Fonagy et al., 2015; Fonagy et al., 2017), intergenerational transfer of knowledge (Csibra & Gergely, 2009, 2011) and communication generally (Grice, 1957, 1975; Sperber & Wilson, 1986; Tomasello et al., 2005)

Differential impact of task construal and trust on strategic choice in negotiation

The proposition we extend and test in this study is that once the parties understand they are in a 'negotiation', they construe the task in front of them as a competition, the value is fixed (the zero-sum bias), and the counterparty is an untrustworthy opponent. This informs their action choice. Because they are competing for a fixed resource, the normative strategy is to contend. Alternative strategies, such as problem-solving, are irrelevant and out of focus.

These competitive mindsets are the consequence not of objective task payoffs, but of recognizing certain features of the situation to be what we call 'negotiation'. In other words, the parties behave more competitively and achieve lower joint outcomes when they understand their task to be a 'negotiation' than if they understand an identical task as, for example, 'problem-solving'.

It also follows that interpersonal trust only partially mitigates the effects of the competitive 'negotiation' construals. While trust may entice the parties to be somewhat more vulnerable and possibly exchange more information, they will still perceive the ‘negotiation’ task to be a competition.

Method

Design

Three groups of participants negotiated a scorable multi-issue negotiation task. In the negotiation condition, the participants negotiated a task labelled ‘negotiation’. In the problem-solving condition, the participants engaged in an identical task except that it was labelled ‘problem-solving’. In the trust condition, the participants completed an interpersonal exercise designed to enhance epistemic trust before completing a ‘negotiation’-labelled task.

Participants

The participants (N = 502) were graduate law students at a large university in the United Kingdom. They were split into three groups: the negotiation condition (n = 172, 63% female, ages 20-37 [mean 25]), the problem-solving condition (n = 180, 62% female, ages 20-36 [mean 25]) and the trust condition (n = 150, 60% female, ages 20-36 [mean 25]). Participants were recruited over a period of two years during the first week of their course to participate in a ‘professional skills exercise’. A policy not to refuse anyone who agreed to take part led to a minor difference in group sizes. The participants received no compensation.

On the basis of the only similar study available (Liberman et al., 2004)we expected a small to medium effect (r = .25). For an independent 3 group ANOVA with two variables, G*power (Faul et al., 2007) suggested a sample of 159 would achieve 80% power (alpha p = .05). Because the current study takes a significantly different approach, we aimed for 95% power and to recruit 150 participants (75 dyads) per cell, overall 450 individuals. Post-hoc analysis showed that the study had a 86% power to detect the results of a 3 group MANOVA with 2 predictors at alpha $p = .05$.

Procedure

In the negotiation and problem-solving conditions, the procedure was explained to the entire group. The participants were instructed to pair up with an individual they did not know. They had 20 minutes to prepare and 30 minutes to conduct a face-to-face negotiation. The procedure was identical in the trust condition, except that after the pairing and before the preparation and negotiation, the participants completed a trust task.

Ethics approval

The university ethics board provided the required approval for the study (UCL 8561/002 and amendments).

Negotiation task

The task was The New Recruit, an established experimental paradigm in negotiation research (e.g., Galinsky et al., 2008; Thompson, 1991; Thompson & Hastie, 1990). To reach an agreement, eight issues had to be agreed, each worth a different number of points to each party. Two issues are distributive, two compatible, and four integrative. The instructions explicitly stated that the payoff schedule must not be shown to the counterparty. A full description of the case is in the methods section to Study 1 (*Task: New Recruit*, p. 112).

Negotiation condition

In the negotiation condition we used the authors' original wording of the instructions. For the recruiter, the relevant part read:

> *This is a negotiation between a job recruiter and a job candidate. You will play the role of the recruiter. There are eight issues of concern in this negotiation…*
>
> *Your goal, as the recruiter, is to reach an agreement with the candidate on all eight issues that is best for you. The more points you earn, the better…*

Problem-solving condition

In the problem-solving condition the instructions had a title 'Problem-Solving Task' and read (changes underlined):

> *This is a joint problem-solving discussion between a job recruiter and a job candidate. You will play the role of the recruiter. There are eight issues of concern in this discussion…*
>
> *Your goal, as the recruiter, is to use this problem-solving session to find, together with the candidate, the best possible arrangement for you on these eight issues. The more points you earn, the better…*

All other text remained unchanged. We did not explicitly modify the goals (e.g. by suggesting the aim is to maximize joint or the counterparty's gain) or the participants' social orientation (e.g., by addressing the counterparty as a 'partner' or 'opponent'; Burnham et al., 2000; De Dreu et al., 2006). Apart from the differences in labels, the tasks were identical.

Trust manipulation

After pairing up, and before receiving any instructions regarding the negotiation task, the participants in the trust condition spent five minutes talking to their counterparties about three questions designed to facilitate trust and to make ostensible efforts to understand the counterparty. This manipulation has been validated in a separate sub-study 4b (pp. 227-235). A full description of the trust manipulation is in the Appendix (p. 297).

Dependent measures

Joint and integrative gain were measured as points achieved by dyads as per the task payoffs. *Integrative gain* was the sum of points negotiators achieve in bonus, vacation, moving expenses, and insurance issues (maximum value is 14,400, the split-down-the-middle compromise is 9,600 points). *Compatible gain* was the sum of points in job assignment and location issues (maximum is 2,400 points, compromise is -600 points). *Joint gain* was the sum of negotiators' points in integrative and compatible issues, which vary depending on dyadic performance, and distributive issues (salary and starting date) that are a constant -3,600 per dyad (maximum is 13,200, the compromise is 4,400 points). We measured the dyad's ability to correctly settle compatible issues on a 0–2 interval scale denoting how many issues were settled correctly (0 = no issues settled correctly, 1 = one issue settled correctly, and 2 = both issues settled correctly). An individual variable of percentage of joint value claimed by each participant was used to assess distributive equity.

Statistical analysis

The data was analyzed with Stata and R (R Core Team, 2013). Joint gain, integrative gain and settled compatible issues were analyzed at the dyadic level. For parametric analysis, we replaced joint and integrative gain (and for the purposes of the graph the compatible gain) with the minimum scores in the group. The joint gain variable was subjected to a power transformation. Where the assumptions for parametric analysis were violated, Wilcoxon-Mann-Whitney, Kruskall-Wallis and multivariate nonparametric tests using R packages *npmv* (Burchett & Ellis, 2017; Ellis, Burchett, Harrar, & Bathke, 2017) and *nparMD* (Bathke & Harrar, 2016; Kiefel & Bathke, 2018) were used on unadjusted data.

We fitted a multivariate omnibus test to assess the impact of the condition on the joint, integrative and compatible gains. We explored specific differences with means or median comparison tests. Distributive equity was assessed by comparing the difference in the percentage of joint gain claimed by recruiters and candidates.

Predictions

We expected that, compared with the control group engaged in a 'negotiation'-labelled task, the participants in the 'problem-solving' condition would (1) correctly settle more compatible issues and (2) achieve higher integrative gain, whereas the negotiators with increased trust (in a 'negotiation'-labelled task) would (3) only correctly settle a larger number of compatible issues. In addition, (4) any distributive inequity between the roles found in the control condition would be attenuated in the problem-solving and trust conditions (this expectation was based on the data provided by the authors of the exercise which shows that candidates claim more value than recruiters: in an MBA sample (*N* = 188), the candidates' mean gain was 5,070 and the recruiters' 4,252).

Results

Six dyads (7%) in the negotiation condition, two dyads (2.2%) in the problem-solving condition and one dyad (1.3%) in the trust condition failed to reach an agreement.

There were two dyads in the negotiation condition where the contracts the parties returned did not match. These were omitted from the dataset.[20]

Trust task manipulation check

The participants in the trust condition completed a 7-point scale (1 = strongly agree, 7 = strongly disagree) questionnaire before and after the trust exercise. The task increased the score of participants in all individual items assessing trust (items 1, 2, 3, 5 and 6, Cronbach's $\alpha = .88$); $t(1, 147) = 13.54, p < .001$. The effects were the strongest in the two items related to recognition of agency (1 and 5), the key driver of epistemic trust. The manipulation was also effective at the dyadic level; $t(1, 73) = 12.56, p < .001$.

The task lowered the participants' anxiety about the upcoming negotiation but did not reduce the expectation that the counterparty would be competitive. On the contrary, the participants expected their counterparties to be marginally *more* competitive after the task. For details refer to **Table 14**.

Table 14. Trust task manipulation check

	Questionnaire					
	before		after		*t*	*p*
I feel I have a sense of the person my counterparty is	4.10	(1.16)	5.26	(.98)	12.79	.0000
I expect my counterparty to be fair	5.46	(1.28)	5.83	(1.06)	4.75	.0000
I expect my counterparty to be reasonable	5.63	(1.13)	5.88	(1.02)	3.29	.0006
I am nervous regarding the upcoming negotiation	3.72	(1.92)	3.11	(1.70)	5.74	.0000
I feel that my counterparty has a sense of who I am	3.81	(1.24)	4.87	(1.12)	10.51	.0000
I feel I can trust my counterparty	4.70	(1.37)	5.42	(1.22)	7.79	.0000
I expect my counterparty to be competitive	4.89	(1.23)	5.02	(1.45)	1.46	.0734

We also conducted a separate control study of the effect of the trust task where the trust was assessed by established trust questionnaires (Lewicki, Stevenson, & Bunker, 1997; Mayer & Davis, 1999). The task was effective in increasing trust in all scales; $F(1, 53)$

[20] Including these as no-deals makes the results related to the hypotheses more conservative.

= 6.87, p = .011 for affect-based, F(1, 53) = 13.49, p < .001 for cognition-based, F(1, 53) = 7.13, p =.01 for ability-based, F(1, 53) = 4.05, p =.049 for benevolence-based, and F(1, 53) = 4.97, p = .03 for integrity-based trust. The study is presented in full following the present study (see Study 4b, p. 227).

Joint gain

As predicted, joint gain was a function of task perception and trust. One-way ANOVA showed the effect of the condition on adjusted joint gain; F(2, 248) = 5.11, p = .007, η^2 = .040, 95% CI [.003, .092]. Kruskal-Wallis of the unadjusted variable was significant too; $\chi^2(2)$ = 9.86, p = .007.

Mean joint gains were the lowest in the negotiation condition (M = 8,393, SD = 3,122). The trust group on average (M = 9,280, SD = 2,200) outperformed the negotiation group by 10.6% and the problem-solving group (M = 9,633, SD = 2,593) outperformed it by 14.8%. These differences were significant for the transformed variable; t(159) = 1.74, p = .042, d = .27, 95% CI [.04, .59]; and t(174) = 3.08, p = .001, d = .46, 95% CI [.16, .76]. For nonparametric tests see **Table 15**.

Table 15. Joint gain in Negotiation, Problem-solving and Trust conditions

Joint gain			1				2			
	M	Med	*U*	*z*	*p*	*r*	*U*	*z*	*p*	*r*
1. Negotiation	8,393	9,000								
2. Problem-solving	9,633	10,200	2,822	3.12	.002	0.23				
3. Trust	9,280	9,600	2,812	1.41	.160	0.11	2,877	-1.64	.101	0.13

Note. The test statistic is Wilcoxon-Mann-Whitney. The effect of the problem-solving manipulation remains significant when adjusted for multiple comparisons.

Compatible and integrative gain

A nonparametric multivariate omnibus test showed a significant effect of the condition on dyadic integrative gain and correctly settled compatible issues, F(4, 494) = 4.03, p

= .003. Because the sample size and the absence of sharp differences in group sizes have likely mitigated the violations of assumptions for parametric models (Pituch & Stevens, 2015), we also calculated a one-way parametric MANOVA of the effect of the condition on adjusted integrative gain and correctly settled compatible issues. The results were very similar to the nonparametric test; $F(4, 494) = 3.89$, $p = .004$, $\Lambda = .94$.

As expected, negotiators in the problem-solving condition achieved higher integrative gain and correctly settled more compatible issues than in the negotiation condition. The effect of the condition was modest but significant on integrative gain; $F(2, 248) = 3.15$, $p = .045$, $\eta^2 = .025$, 95%, CI [.00, .07], and more robust on compatible issues; $\chi^2(2) = 11.19$, $p = .004$. Details are in **Figure 23**, **Table 16** and **Table 17**. Negotiators in the trust condition outperformed the negotiation controls in compatible issues, but not in integrative gain.

Figure 23. Dyadic gain in Negotiation, Trust and Problem-solving conditions as percentage of maximum

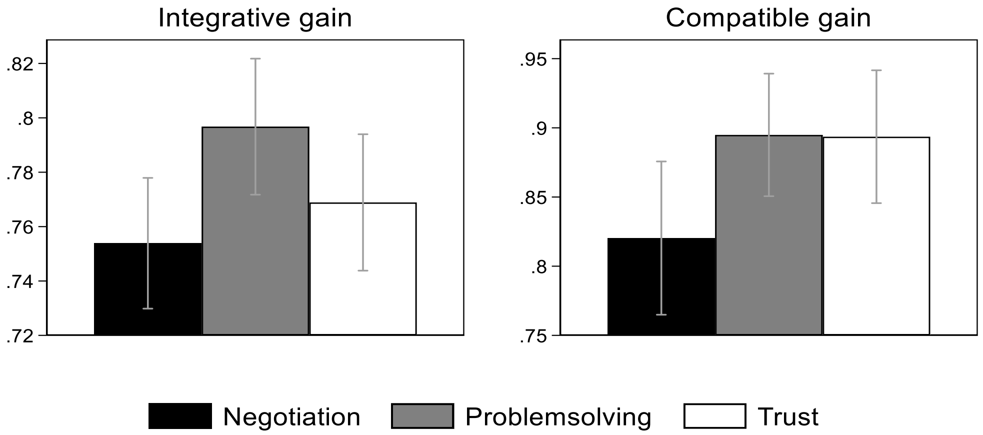


Table 16. Integrative gain in Negotiation, Problem-solving and Trust conditions

Integrative gain					1			2	
	M	*SD*	*.95 CI*		*t*	*p*	*d*	*t*	*p*
1. Negotiation	10,856	1,630	10,506	11,205					
2. Problem-solving	11,473	1,734	11,110	11,837	2.43	.008	.37		
3. Trust	11,072	1,588	10,707	11,437	.85	.198	.13	-1.54	.063

Note. P-values are two-tailed. The effect of the problem-solving manipulation remains significant if adjusted for multiple comparisons.

Table 17. Identified compatible issues in Negotiation, Problem-solving and Trust conditions

Compatible issues			1				2			
	M	*Med*	*U*	*z*	*p*	*r*	*U*	*z*	*p*	*r*
1. Negotiation	1.29	1.50								
2. Problem-solving	1.60	2.00	3,047	2.79	.005	0.21				
3. Trust	1.63	2.00	2,490	2.84	.005	0.22	3,335	0.17	.869	0.01

Note. Data was unadjusted. The test statistic is Wilcoxon-Mann-Whitney. Reported *p* values are two-tailed. The effects of the problem-solving and trust manipulation remain significant if adjusted for multiple comparisons.

If we exclude no- from the analysis, the difference in integrative gain between the negotiation (*M* = 11,085, *SD* = 1,448) and trust groups (*M* = 11,116 , *SD* = 1,552) virtually disappears, suggesting that any improvement in integrative gain in the trust group was due to avoided impasses rather than effective logrolling.

Distributive equity (individual)

The distribution of gains between the parties was unequal in the negotiation condition and not in the problem-solving and trust conditions. In the negotiation group, candidates claimed significantly (11%) more than recruiters (Wilcoxon-Mann-

Whitney $U = 2193.5$, $z = 3.44$, $p < .001$, $r = .27$). These differences were not significant in the problem-solving ($p = .85$) and trust ($p =.15$) conditions.

Unexpected effect of gender on gains

One-way ANOVAs showed an unexpected main effect of dyadic gender (all female, all male or mixed) on integrative gain, $F(2, 237) = 3.92$, $p = .02$, $\eta^2 = .032$, but not on identified compatible issues ($p = .23$). There was no interaction with the condition ($p = .65$). Neither parametric (two-way MANOVA, $p = .69$) nor nonparametric multivariate models using unadjusted data (employing nparMD package in R, Bathke & Harrar, 2016; Kiefel & Bathke, 2018; p = .53) found no significant interaction.

Further analysis showed that dyadic gender may have been a suppressor variable. A two-way ANOVA and MANOVA controlling for dyadic gender showed an enhanced effect of condition; $F(2, 235) = 4.04$, $p = .019$, $\eta^2 = .033$; and $F(4, 466) = 4.56$, $p = .001$, $\Lambda = 0.93$. Because this was not predicted, we do not discuss it further.

Discussion

In negotiations with identical payoff structures, a labelling manipulation encouraging the perception that the task is to maximize individual gain by 'problem-solving', rather than by 'negotiating', improved the mean joint gain by 14.8%. This increase was due to higher integrative and compatible gains. Increasing dyadic (epistemic) trust improved the mean joint gain by 10.6%. Unlike the problem-solving manipulation, trust only increased the number of correctly settled compatible issues; there was no effect on integrative gain, which is consistent with prior studies (Butler, 1999; Kimmel et al., 1980). This was particularly evident when no-deals were excluded. Both manipulations also improved distributive equity.

Why did the participants who were 'negotiating' this achieve inferior joint outcomes to the participants who were 'problem-solving'? Our findings lend indirect support to the idea that understanding a task as ‘negotiation’ results in competitive mindsets, centered around the zero-sum assumptions. These cognitions favor strategies focused on value claiming and leave out of focus (in the sense of focusing failures discussed by Chugh & Bazerman, 2007) the possibilities and strategies for cooperative optimization (i.e., if the task is a distribution, we need to compete).

The difference in outcomes was driven by the difference in the label, and in the participants who engaged in ‘negotiation’, some of the processes required to maximize joint gain – e.g., some level of trust, a truthful information exchange, taking perspective of another with a cooperative goal in mind (e.g., the joint intentionality required for stag hunt; Tomasello, 2014) – must have been poorer than in their ‘problem-solving’ peers.

The partially superior performance of the trust group vis-à-vis the negotiation group suggests that the trust manipulation adjusted one of the consequences of the negotiation construal but did not modify the perception that the task is to compete. We understand this as follows. In competitive situations, extending trust is a precarious courtesy because virtually all distributive tactics consist of some sort of deception or misinformation (e.g., 'shaping perceptions of what is possible'; Lax & Sebenius, 2006, chapter 12). It is risky to trust or tell the truth (Lax & Sebenius, 1986b; Loschelder et al., 2014; Murnighan et al., 1999). However, the mistrust is relatively independent of the perception of the task as a distribution. It is perfectly possible to distrust someone when facing a joint problem (e.g., Mnookin, 2010) or trust someone in a competitive setting (e.g., distributive negotiation with a close friend, or competition between athletes). The increased trust had some effect on information exchange that benefitted compatible gain (as the parties’ preferences are not in conflict, all that is necessary for optimal settlement is unilateral truthful disclosure of rudimentary preferences) but failed to generate more sophisticated strategies required to logroll.

The current research has several limitations. First, although the findings lend general support to the proposition that the ‘negotiation’ construal triggers distribution-focused mindsets, the study measured outcomes rather than the hypothesized cognitive processes. We abstained from administering questionnaires about task-perception to avoid disrupting the priming process and because the participants’ self-reports of predominantly implicit construal processes may not be reliable (e.g., Nisbett & Wilson, 1977). However, the limitation remains: we assume, but have not measured, there being thoughts and beliefs (that underpin action) that are more adversarial and competitive in the 'negotiation' than in the 'problem-solving' condition. In other words, our experiment assessed the effect of construal on outcomes, not on the mental states that drive such outcomes. We address this in Study 5 and 6, below. Second, generalizability of the findings is limited because the tasks deviate from real-life bargaining (e.g., the interests and priorities in New Recruit are arbitrarily assigned, measured in points and have no personal value for the participants). However, multi-issue scorable tasks have been an effective research tool as they allow inference about underlying cognitive processes (De Dreu & Carnevale, 2005b). While the magnitude of the observed effects may not be generalizable, we can expect the effects themselves – that framing a situation as 'negotiation' results in inferior outcomes than 'problem-solving' – to take place in natural settings, too, particularly given the solid grounding of the hypotheses in theory and practice. Nevertheless, ideally, the manipulations should be studied in an ecologically more generalizable context. Third, while the task was labelled 'problem-solving', it took place in an employment context, which in many cultures is *a priori* primed for distributive bargaining. Also, our participants were students in an adversarial and competitive field (law). It is difficult to say whether this enhanced or suppressed the effects of our manipulations. Fourth, most participants lacked formal training in negotiation. It is possible that trust can improve integrative gain in a population of negotiators familiar with logrolling techniques. Fifth, there is an alternative explanation for why increased trust did not improve integrative gain. The relationship between trust and integrative gain may be dose-respondent such that a certain threshold level of trust leads to improved integrative gain, and the trust manipulation may have been too weak to reach this threshold. Finally, and perhaps most importantly, the trust manipulation check questionnaire was not a standard, validated measure of trust, and the perception of cooperativeness and competitiveness was a single item in the manipulation check questionnaire. A study assessing the effects of trust manipulation on trust and competitiveness using standard assessment measures follows below.

STUDY 4b: IMPACT OF TRUST TASK ON TRUST AND PERCEIVED COMPETITIVENESS

To further evaluate the effect of the trust manipulation on trust and perceived competitiveness, we conducted a study assessing the effect of the trust manipulation from Study 4 using a series of trust questionnaires established in negotiation trust literature (Lewicki et al., 1997; Mayer & Davis, 1999).

The aim of this study was twofold. First, the goal was to assess the effect of the trust task on interpersonal trust and exclude the possibility that the effect was due to dyadic interactions that did not target trust (e.g., increased familiarity with the negotiation partner driven by exchanging information on other topics). We therefore examined if joint activity with the counterparty alone could account for the difference in trust using an independent sample of participants randomly assigned to either reading to each other or exchanging information in relation to themselves as in the trust generation task.

Our second aim was to investigate the expected differential effects of the trust task on the perception of the counterparty's trustworthiness and the perception of the counterparty's competitiveness. Our theoretical prediction, supported by the results of the trust task manipulation check in Study 4, was that the trust task would increase interpersonal trust, but not the expectation of cooperativeness and competitiveness of the counterparty.

Method

Participants

The participants (*N* = 55) were students enrolled in an intensive course in negotiation, split into the trust group (*n* = 30) and the control group (*n* = 25).

Post hoc analysis with G*power (Faul et al., 2007) showed that the study had larger than 80% power for all its significant findings for repeated measure ANOVA with 2 groups and 2 measurements at alpha *p* = .05.

Procedure

Each participant was paired up with a person they did not know. They were told that after the 'initial exercise' they would negotiate a negotiation task, which they did. Both the trust and control groups completed a set of trust questionnaires before and after their respective tasks. The trust group performed the trust task from Study 4 (see p. 218). The control group spent the same amount of time exchanging information by taking turns reading to each other excerpts from a text on social cognition and everyday mind-reading (Lieberman, 2013). The instructions required them to complete the task but did not restrict their communication, and most of them engaged in social pleasantries (e.g., introducing themselves, shaking hands) before commencing the task. No participant in this control study participated in any other study conducted as part of this thesis. The participants received no compensation for their participation.

Ethics approval

The university ethics board provided the required approval for the study (UCL 8561/002 and amendments).

Dependent measures

The pre- and post-task trust questionnaires included eight different instruments for measuring trust. The first set included scales assessing cognition-based trust (based on calculus- and knowledge-based trust) and affect- (identification)-based trust (Lewicki et al., 1997). The second set consisted of the classic scales assessing ability-based trust, benevolence-based trust and integrity-based trust (Mayer & Davis, 1999). Both are commonly used in negotiation research (e.g., Mislin et al., 2011; Olekalns, Lau, & Smith, 2007). We also employed a questionnaire assessing perceived cooperativeness and trust (used in e.g., De Dreu et al., 2006), measured on two three-item scales assessing perceived trustworthiness and cooperativeness/competitiveness of the counterparty, respectively. These were assessed with six semantic differentials on a scale 1-5 (e.g., 'The other party could not be trusted at all (1)' to 'could be trusted very well' (5) for trust and 'I see the other as very competitive' (1) to 'I see the other as very cooperative' (5) for the perception of the counterparty.

Predictions

We predicted that the trust manipulation would increase trust across all trust scales, i.e., that it would boost cognition-based trust (based on calculus- and knowledge-based trust), affect-based trust, ability-based trust, benevolence-based trust, integrity-based trust.

For the perceived cooperativeness and trust scale, we predicted that the trust task would increase the trust component, but not the perceived cooperativeness subscale.

Results

As expected, the trust manipulation was effective in increasing trust. We used a two-way repeated measures ANOVA with time and condition as predictors. In all trust questionnaires, there were no main effects but significant interactions between trust condition and time, and tests of simple effects revealed significant effects for the trust condition but not for control condition.

Affect-based and cognition-based trust

The trust task increased both affect- and cognition-based trust (**Table 18**, **Figure 24** and **Figure 25**). The significant effect on the overall cognition-based trust was driven by the highly significant effect on the calculus-based subscale. The cognition-based second subscale, the knowledge-based trust, was not affected, presumably because a five-minute conversation on specific personal questions is unlikely to provide sufficient information to the participants to make confident predictions regarding the counterparty's future behavior, which is what the subscale measures (Lewicki, Bunker, & Research, 1996; Lewicki et al., 1997). Conversely, such conversation seems effective in generating the feeling of reliability that the counterparty might keep their promises, which is the key driver of calculus-based trust.

Table 18. Effect of the trust task assessed by cognitive and affect based trust scales

	Cognitive and affect-based trust scales							
	Knowledge-based		Calculus-based		Cognitive-based (total)		Affect-based	
	F	*p*	*F*	*p*	*F*	*p*	*F*	*p*
Condition	0.60	.441	5.80	.000	1.53	.221	0.46	.502
Time	3.83	.056	27.97	.000	24.13	.000	6.44	.014
Condition * Time	0.92	.343	16.26	.000	10.76	.002	6.87	.011
Model R^2	.570	.000	.712	.000	.747	.000	.490	.000
Simple effects of Condition								
Control group			0.72	.399	1.22	.274	0.00	.955
Trust group			47.79	.000	36.91	.000	14.64	.000

Figure 24. Effect of the trust task assessed by cognitive and affect based trust scales with .95 CI

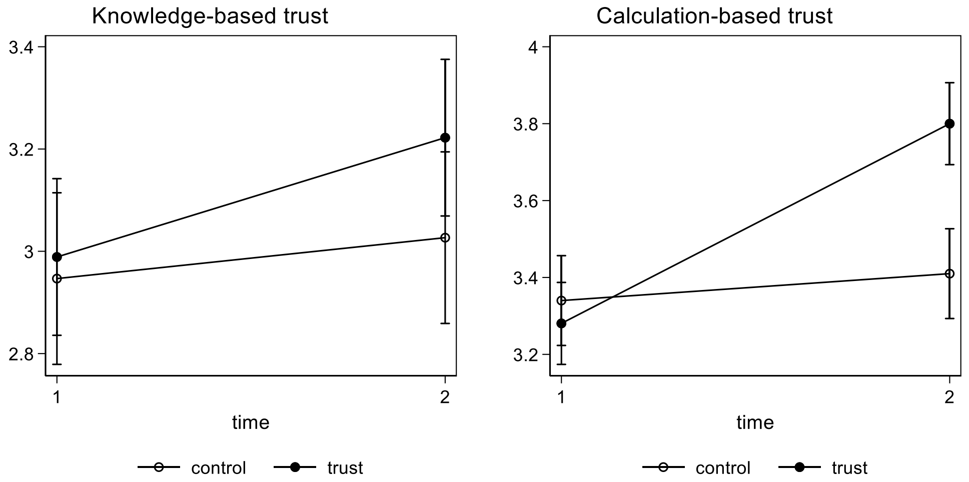
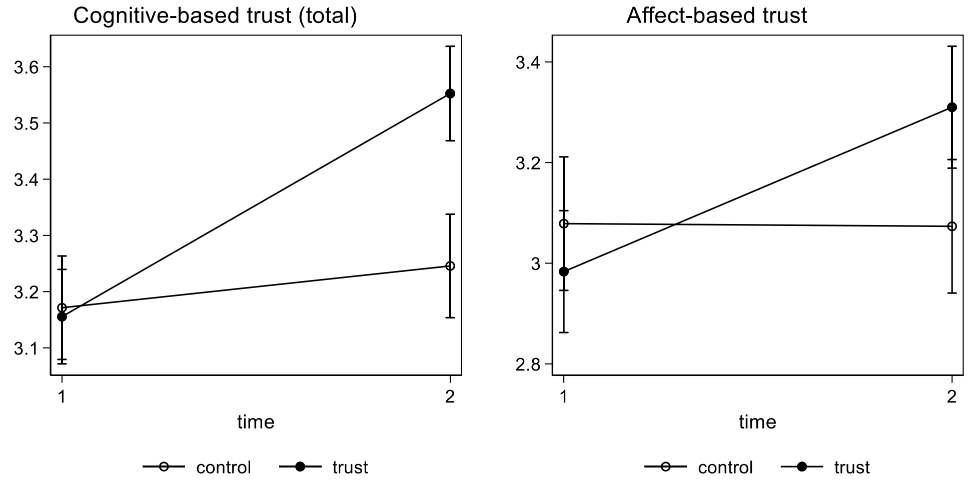


Figure 25. Effect of the trust task assessed by cognitive and affect-based trust scales with .95 CI



Ability, Integrity and Benevolence-based trust

As expected, the trust task increased ability-based trust; *F*(1,53) = 7.13, *p* =.01, benevolence-based trust; *F*(1,53) = 4.05, *p* =.049, and integrity based-trust. *F*(1,53) = 4.97, *p* = .03. Details are in **Table 19**, **Figure 25** and **Figure 26**.

Table 19. Effect of the trust task on ability-, benevolence- and integrity-based trust scales

	Ability, Benevolence and Integrity based trust					
	Ability-based		Benevolence-based		Integrity-based	
	F	*p*	*F*	*p*	*F*	*p*
Condition	0.30	.584	1.83	.180	0.51	.480
Time	19.76	.000	15.80	.000	20.61	.000
Condition * Time	7.13	.010	4.05	.049	4.97	.030
Adjusted model R^2	.516	.000	.817	.000	.639	.000
Simple effects of Condition						
Control group	1.44	.235	1.60	.211	2.45	.124
Trust group	27.85	.000	19.12	.000	25.21	.000

Figure 26. Effect of the trust task on ability- and benevolence-based trust scales with.95 CI

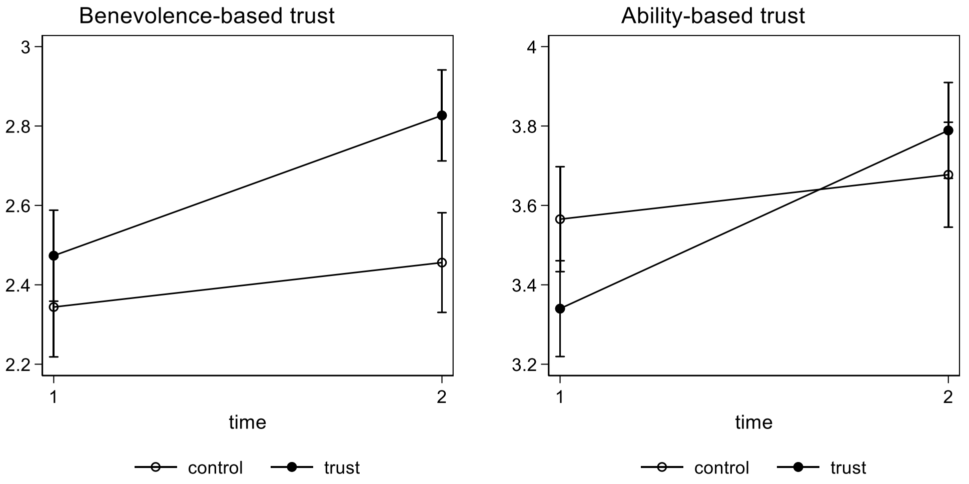
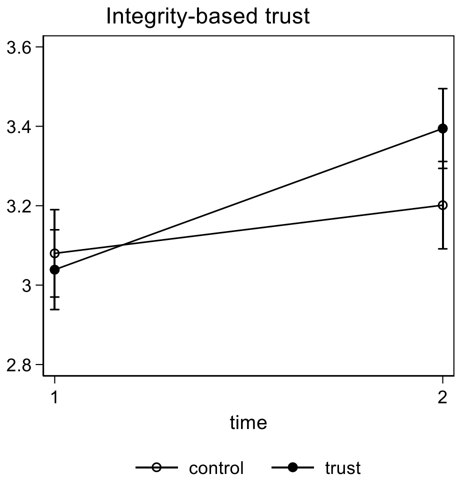


Figure 27. Effect of the trust task assessed by integrity-based trust with.95 CI



Perceived trustworthiness and cooperativeness scales

As expected, the trust task had an effect on the 3-item trust scale, but not on the 3-item perceived cooperativeness scale (De Dreu et al., 2006); the marginally significant effect on the overall trust-cooperativeness scale was driven mainly by the scores in the 3-item trust scale. Details are in **Table 20**, **Figure 28** and **Figure 29**.

Table 20. Effect of the trust task on the perceived cooperativeness and trust 3-item scales

	Trust and cooperativeness 3-item scales					
	Trust		Cooperativeness		Total	
	F	*P*	*F*	*p*	*F*	*p*
Condition	0.59	.448	0.70	.405	0.71	.403
Time	38.63	.000	45.41	.000	53.57	.000
Condition * Time	5.63	.021	0.90	.346	3.58	.064
Adjusted model R^2	.709	.000	.761	.000	.771	.000
Simple effects of Condition						
Control group	6.77	.012				
Trust group	40.56	.000				

Figure 28. Effect of the trust task on the perceived cooperativeness and trust 3-item scales with .95 CI

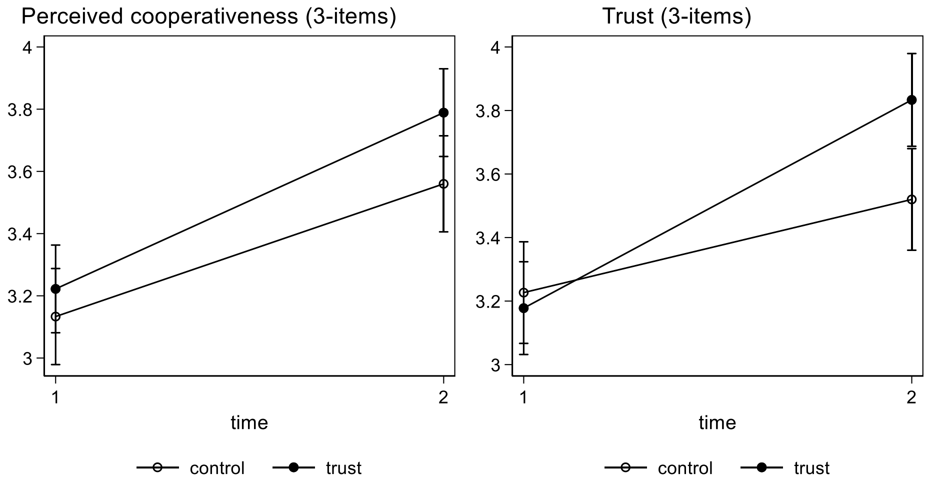
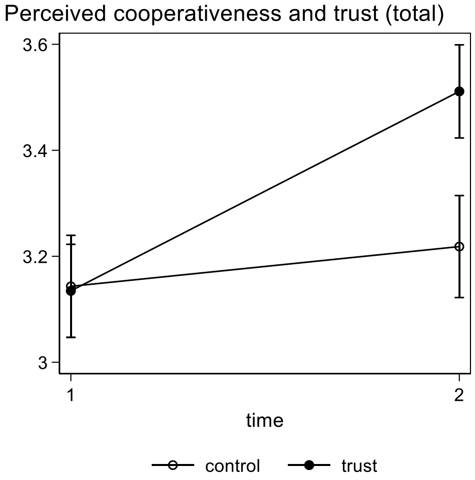


Figure 29. Effect of the trust task on the total perceived cooperativeness and trust with .95 CI



Discussion

As expected, the trust manipulation was effective in increased cognition-, affect-, ability-, benevolence- and integrity-based trust, however it did not increase the perceived cooperativeness of the counterparty.

These findings, jointly with the results of Study 4, strongly suggest that trust only partially mitigates the effects of the competitive construal of negotiation tasks. As elaborated in detail above (p. 224 and after), increased trust does not correct the perception that the task in a 'negotiation' is to compete. The parties can trust, not trust or trust something in-between in a task that is a cooperative or competitive venture.

From a practical perspective, the results of Study 4 and Study 4b lend empirical support to the long-standing prescriptive advice that parties should problem-solve in negotiation (Menkel-Meadow, 1983; Pruitt, 1983b; Raiffa, 2002) and use techniques such as joint fact-finding, brainstorming and the single-text procedure (Carter, 1982; Fisher et al., 1991). Active listening and expressing empathic understanding (see also Mnookin, 2000) are effective because they facilitate epistemic trust and increase compatible gain. The study also suggests that epistemic trust and a problem-solving orientation of negotiators attenuate inequity in the distribution of joint gain.

STUDY 5: TASK CONSTRUAL AND NEGOTIATORS' STRATEGIES IN A MULTI-ISSUE TASK

Studies 4 and 4b provided evidence for the idea that construing a task as 'negotiation' leads to poor dyadic outcomes but did not examine the intermediate step: the mindsets and behaviors of the parties that are the result of such construal. In other words, we know that understanding an interaction as 'negotiation' (as opposed to an alternative, less contentious, 'problem-solving' frame) causes suboptimal outcomes, but not the thoughts and behaviors (strategies) of the parties that lead to such outcomes.

In the following study we sought to address this issue. In Study 4 we assumed, but had not measured, different mental states that differentially drive negotiation outcomes associated with the 'negotiation' and 'problem-solving' construals. If our hypothesis is correct, these mental states ought to be relatively stable and able to be triggered without the participants actually doing a negotiation task, but rather by being presented with one and asked what strategies they believe they would employ in such a task.

Two groups of participants were described an identical multi-issue situation and tasked to get the best outcome for themselves. They were then asked to evaluate which strategies they would use as an effective way of achieving this goal. We expected that understanding the situation as 'negotiation' – as opposed to 'problem-solving' – would result in behaviors and strategies more appropriate for wining a zero-sum contest than for maximizing gain by exploiting integrative and compatible issues.

That people think negotiation is mainly a competition – and that they choose their behaviors accordingly – is not a new idea (e.g., O'Connor & Adams, 1999). However, the proposition that what makes people think a situation is mainly a competition depends not on the task structure, but on negotiators' construal of the task, is. In this study we investigate the impact of alternative construals of an identical situation on the strategies the parties tend to endorse. This research extends the findings of Studies 4 and 4b that situational construal affects integrative and compatible gain in identical tasks to the strategies that are triggered by the differential construals and, ultimately, underpin the differences in negotiation outcomes.

Method

Participants

The participants (*N* = 116, 56% female) were students at a large UK university recruited during two regularly scheduled sessions of a negotiation course. Our cut-off policy was class size and the students' consent to participate in the experiment. Post hoc analysis with G*power (Faul et al., 2007) showed the study had <80% power to detect the effects of an omnibus MANOVA with two groups and two predictors at alpha *p* = .05.

Procedure

The participants conducted the task using the university's electronic platform as a regular course assignment. Before and after the task the students were informed they can allow us to use their anonymized data for research purposes if they wish to do so, and provided with all information about the research project and the consent form. The participants who provided consent were included in the dataset. They received no compensation.

Ethics approval

The university ethics board provided the required approval for the study (LRS-19/20-20988).

Task

Two groups of participants were asked to imagine they were about to enter a negotiation situation based on the bartenders task by De Dreu et al. (2006), modelled on past research using scorable multi-issue paradigms (e.g., Pruitt & Lewis, 1975). They were told that they were starting work as one of two bartenders in a pub and that they need to address three issues: (i) how to split the tips between each other, (ii) who would work evenings on different days during the week, and (iii) how often each would clean the floors and toilets of the pub. Both groups had an *identical* description of the task and were explicitly instructed to maximize individual gain ('[y]our task is to get the best outcome for yourself').

Negotiation condition

The experimental manipulation was in the label. The 'negotiation' group was asked to imagine that they are 'about to conduct a negotiation with another person,' that includes three issues they needed 'to negotiate'. They were then asked to consider a number of tactics they could 'use in negotiation' and tell us which ones they 'would use as an effective way of negotiating'.

Problem-solving condition

The problem-solving group was asked to imagine that they were 'about to conduct a problem-solving session with another person' that included three issues they 'need[ed] to solve'. They were instructed to consider a number of tactics they 'could use in solving such a problem' and to tell us which they 'would use as an effective way of problem-solving in this situation'.

Dependent measures

The participants' strategies were measured by an adapted Dutch Test for Conflict Handling (De Dreu, Evers, Beersma, Kluwer, & Nauta, 2001; Janssen & van de Vliert, 1996). The lean version of the test comprises four scales: contending (forcing), problem-solving, avoiding and yielding, scored on a 1-7 scale (1 = Strongly disagree, 7 = Strongly agree). We used the problem-solving (e.g., 'I tried to work out a solution that serves my own as well as other's interests as well as possible') and the contending ('forcing', e.g. 'I did everything to win') scales. Cronbach alphas were .71 for contending and .80 for problem-solving.

Statistical analysis

We analyzed the data with Stata. There were two outliers in the problem-solving scores, one in each condition. They were replaced with the minimum scores in the group.[21] We first conducted a multivariate omnibus test and followed up with t-tests to highlight the differences between conditions and determine effect sizes.

Predictions

We predicted that the problem-solving group, compared with the negotiation group, would score higher on the problem-solving scale and lower on the contending scale. We made no predictions about yielding because the inclination to yield ought to be low in a competitive interaction ('negotiation') and irrelevant in cooperative 'problem-solving' (i.e., yielding can facilitate either integrative or suboptimal outcomes; Pruitt, 1983b). Similarly, we made no predictions about avoiding because it signifies inaction (Pruitt, 1983b) which is equally unhelpful to both competitive and collaborative approaches.

[21] Leaving the values unchanged does not alter the results in any meaningful way.

Results

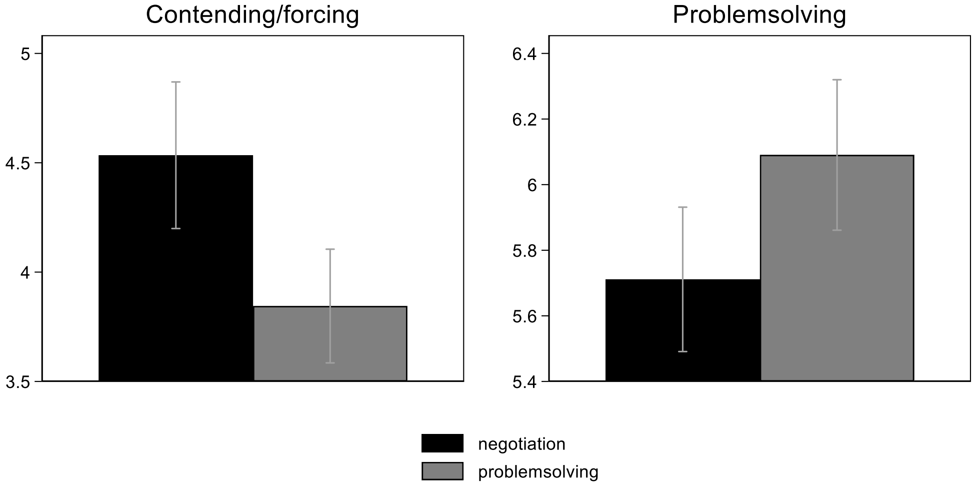
One-way MANOVA indicated an effect of the condition on dyadic problem-solving and contending; *F*(2, 113) = 6.34, *p* = .003, Λ = .90. As predicted, the participants in the problem-solving condition used more problem-solving and less contending strategies than their 'negotiation' peers. Details are in **Table 21** and **Figure 30**.

Table 21. Means and test statistics for problem-solving and contending in Problem-solving and Negotiation conditions

	Problem-solving		Contending	
Problem-solving	6.09	(.882)	3.84	(1.00)
Negotiation	5.71	(.847)	4.53	(1.29)
t		-2.36		3.22
p		.001		.001
d		.439 [.069, .806]		.598 [.225, .969]

Note. T-tests are one-tailed. Effect size 95% CIs are provided in square brackets. All relationships were also tested with Wilcoxon Mann Whitney and remained significant.

Figure 30. Contending and problem-solving in Negotiation and Problem-solving conditions with .95 CI



Discussion

The results of the study indicate that understanding an identical situation as 'problem-solving' rather than 'negotiation' impacts the strategies endorsed by the participants. Participants who understood an identical task to be 'negotiation' endorsed significantly more competitive (contending) strategies and were significantly less interested in problem-solving than their peers who understood the task to be 'problem-solving'.

These results are consistent with the findings of Studies 4 and 4b – showing that labelling an interaction 'negotiation' produces lower integrative and compatible gain than labelling it 'problem-solving' – and provide evidence of the strategies that lead to such inferior outcomes. It also supports the proposition that participants who construe an interdependent mixed-motive situation as 'negotiation' understand it as a predominantly competitive, zero-sum affair, and make their choices accordingly.

Contending, or 'forcing', involves trying to convince the counterparty to yield to one's demands via threats, positional commitments, ultimatums and advocacy, and attempts at gaining information while concealing one's own (Lax & Sebenius, 2006; Pruitt, 1983b). Such strategies are appropriate for zero-sum tasks where the only objective is to claim value. At the same time, the problem-solving strategies are considered pointless and are out of focus, which is reflected in significantly lower problem-solving scores of the negotiation group vis-à-vis their problem-solving peers.

The limitations of this study are predominantly related to the task. The task required the negotiators to imagine an interpersonal situation and report what kind of strategies they would employ in such fictional scenario. What people report they would do and what they do may be different things. However, we are investigating the participants' beliefs about what kind of strategies would make sense in 'negotiation' and 'problem-solving'. However, the study achieved our research goals. We were interested in what kind of strategies the participants consciously endorse for the two respective tasks, and the findings are legitimate in that context: they tell us which strategies the participants thought were most effective for 'negotiating' and 'problem-solving' a multi-issue task.

Jointly, studies 4, 4b and 5 show that understanding a multi-issue interdependent mixed-motive task with value potential as 'negotiation' compared with understanding it as ‘problem-solving’ triggers more competitive and less problem-solving behaviors, and lead to lower integrative and compatible gains. The major limitation of this set of studies is that while they show the impact of construal on outcomes on one hand, and the impact of construal on strategies on the other, they do not demonstrate the suspected mediation between construal, strategies and outcomes. In the following study, we present the participants with a different negotiation paradigm and slightly different primes, and measure both the strategies they employ and their impact on the outcomes they achieve.

STUDY 6: TASK CONSTRUAL, NEGOTIATORS' STRATEGIES AND OUTCOMES IN HIDDEN VALUE TASK

Studies 4 and 4b provided evidence for the idea that construing a task as 'negotiation' leads to poor dyadic outcomes in multi-issue tasks. Study 5 then examined the intermediate step: the impact of 'negotiation' construal on the processes, or more specifically on the behaviors of the negotiators. In the following study we sought to replicate the effect of construing the situation as a 'negotiation' versus an alternative, more collaborative frame, on both outcomes and strategies, this time in a different experimental task. As noted in the limitations section of Study 4, while the multi-issue scorable tasks have been an effective tool in negotiation research, they are relatively weak on the external validity front, mainly because the payoffs and preferences are not clearly connected to the parties' interests and are in that way somewhat arbitrary.

In the present study, we used a negotiation task situated in an M&A context, based on a reportedly real case, and investigated the strategies the participants used under different construals, and how these strategies contributed to the quality of outcomes.

Method

Participants

On the basis of the study observing the effect of construal on a competitive and collaborative interaction resulting in a binary outcome (Liberman et al., 2004), G*power (Faul et al., 2007) suggested a total sample of 50 dyads (100 individuals) for 80% power (alpha $p = .05$).

The participants (*N* = 120, 67% female, ages 20-38 [mean 26]) were graduate students at a large UK university recruited during two regularly scheduled sessions at the end of a negotiation course (our cut-off policy was class size).

Before they were recruited for this study, the participants had been trained in negotiation over a period of 11 weeks. They have completed and were debriefed on 15 different negotiation exercises. The topics covered included, in this order, the 7-elements framework based on the Harvard Principled Method, distributive tactics, objective criteria, interests and alternatives, the three tensions model (creating versus claiming, the agency tension, and empathy versus assertiveness), as well as the basic cognition in negotiation (mainly consisting of naïve realism and reactive devaluation). Most importantly, they have been exposed to material involving interests as motivators, information asymmetry, strategic interdependence in bargaining, the negotiator's dilemma, and have – among other things – completed a number of the negotiation exercises where optimal solutions were based on the notion of divergent rather than conflicting interests.[22] This implies the participants have known, at least at an abstract level, that interests can be in conflict, but also compatible (shared) or different in the sense they can be logrolled.[23]

Procedure

After pairing with someone they did not know, the participants had 10 minutes to prepare and 15 minutes to complete the task. They then filled out questionnaires assessing the strategies they used. They received no compensation.

[22] For example, in addition to the New Recruit (Study 1 and 4) and Hijacked Performance (Study 3), they had completed a settlement negotiation exercise where two parties on the brink of litigation about damages due to pesticide spray-drift need to, in addition to figuring out a settlement amount, also address the issue of future spray drift, where their interests are perfectly compatible.

[23] They have also heard, on many an occasion, a mantra that 'negotiators tend to assume that all interests are in conflict, when they are often not'.

Ethics approval

The university ethics board provided the required approval for the study (UCL 8561/002 and amendments).

Task

The task was a corporate acquisition based on a fixed-pie problem from Bazerman and Neale (1992, p. 17). The participants represented a prospective acquirer and target in a potential corporate acquisition. The prospective acquirer is a large diversified pharmaceutical corporation that has recently taken a strategic interest in the skin creams niche due to its projected growth and is pursuing expansion through acquisition. For that reason, the Acquirer is interested in the Target company, which is a medium-size cosmetics business with a small but strong presence in the niche the Acquirer is interested in. The Target has a startup history – it evolved from a stall in a local market into a respected medium-size business – and the two owners are in their late 50ties, thus potentially thinking of selling the company and retiring.

Each party was provided with a confidential set of instructions that contained their interests and a confidential company valuation (see **Table 22** for detail). They were further instructed that if any deal were to take place, the purchase price would have to be paid in cash (i.e., it could not include a deferred payment, an earn-out formula, equity interest, shares and such), and that no strategic alliances (e.g., share stakes, joint ventures or partnerships) were an option.

Table 22. Differential information in the M&A task

Acquirer	**Target**
Strategic interest in skin-creams. Values the company £16 million ▪ £10 million for skin-creams division ▪ £3 million for eyeliners division ▪ £3 million for lipsticks and glosses division Target's valuation likely significantly lower as they cannot realize the synergies the Acquirer can. The key interest is financial gain (including specific instructions how to calculate it).	Values the company £12 million, all divisions 'equally valuable' ▪ skin-creams division ▪ lipsticks and glosses division ▪ eyeliners division (develops new products) Acquirer's valuation is likely significantly higher due to the synergies they can utilize. The key interest is financial gain (including specific instructions how to calculate it). In M&A transactions profits of up to 40-50% are not uncommon. The acquirer is interested in the Target as part of their move into the skin-cream business. The manager-owners like developing new products.

The participants were specifically instructed that their key interest was individual financial gain. There are essentially two ways the parties can agree this deal. They can close an acquisition of the whole business, resulting in the joint gain of £4 million. However, the less obvious but superior solution is to agree an acquisition of only a single division (skin-creams). This solution provides higher dyadic financial gain (£6 million), satisfies the Acquirer's interest in expansion through acquisition of the Target's niche skin-creams operation, leaves the Target's manager-owners the eyeliner development division to manage, and is the only way of ensuring the aspired premium of 40-50% to the Target (see **Table 23** for details of the valuations and the bargaining range).

Table 23. Valuations, bargaining range and premium in the M&A task

	Value to Target	Value to Acquirer	Dyadic financial gain*	Average acquisition premium†
Whole company	12	16	4	16.7%
Division A	4	3	-1	-12.5%
Division B	4	3	-1	-12.5%
Division C	4	10	6	75.0%

Note. Values are in million. *The goal of negotiators is to maximize their individual financial gain, equivalent to their share in the dyadic financial gain. † Premium if the dyadic financial gain is split evenly (e.g., the whole company is acquired for 14 million).

This is not an easy task to get right. We use it in graduate and undergraduate negotiation education to illustrate the difficulties posed by the negotiator's dilemma and the fixed-pie bias. Anecdotal evidence based on its use during the past 7 years suggests that even in cohorts of students who have attended lectures on the fixed pie assumption and cognitive challenges in negotiation, only about 10% of negotiators manage to close the optimal deal.[24]

Negotiation condition

In the negotiation condition (*n* = 60) the participants were told that they were entering a 'negotiation about a potential corporate acquisition', and that their job was to 'to maximize the interests of [their] company in accordance with [their] instructions.'

Deal-design condition

In the deal-design condition (*n* = 60), the participants were entering 'a pre-negotiation creative deal-design discussion' and their task was 'to explore, jointly with the other party, the possibility of creating a deal that maximizes value and the interests of your company.' They were encouraged to try to find solutions 'by employing creativity and problem-solving.' The body text of the task, including the background, the interests of the parties, their valuations and the directive to maximize individual financial gain, was identical in both conditions.

Dependent measures

A binary variable measured whether dyads reached an optimal arrangement or not. An ordinal variable measured the quality of the agreement (0 = no deal, 1 = deal exceeding instructions and resulting in a loss to one party, 2 = suboptimal deal, 3 = optimal deal).

[24] The results of this study are remarkably similar, see p. 248, below.

A deal was deemed to exceed instructions when both parties reported a concluded agreement, but a buyer (seller) has agreed a price that was above (below) their reservation value (i.e., they incurred a loss). We also used a variable where we deemed such deals that exceeded instructions as no-deals (0 = no deal, 1 = suboptimal deal, 2 = optimal deal). The participants' strategies were measured by an adapted Dutch Test for Conflict Handling (De Dreu et al., 2001; Janssen & van de Vliert, 1996). The 'lean' version comprises four scales: contending (forcing), problem-solving, avoiding and yielding. Each scale features four items scored on a 1-7 scale (1 = Strongly disagree, 7 = Strongly agree). We used the same scales as in Study 5: the contending ('forcing', e.g. 'I did everything to win') and problem-solving (e.g., 'I tried to work out a solution that serves my own as well as other's interests as well as possible'). Cronbach alphas were .70 for contending and .74 for problem-solving.

Statistical analysis

We analyzed the data with Stata. Approximately 5% of the questionnaires was missing data. We performed the analysis on an incomplete dataset (listwise deletion). These results were checked, where possible, against the dataset where the missing values were imputed using the MVN algorithm (*mi impute* function in Stata). We tested the overall deal rate with Pearson chi squared, Fisher's exact test and proportion tests. To investigate the impact of condition on strategies, we conducted a multivariate omnibus test and followed up with t-tests to highlight the differences between the conditions, and with a hierarchical regression with a dyad as a random effect. We then tested the impact of strategies on the odds of optimal outcomes by fitting logistic and ordered logistic regression models. Finally, we estimated the indirect effect in logistic regression using the PROCESS macro (Hayes, 2020) by bootstrapping 5,000 samples.

Predictions

We expected that, compared with a control group understanding the situation as a 'negotiation', the participants perceiving the situation as 'deal-design' would (1) complete more optimal agreements and (2) achieve deals of higher quality. We further expected that they would (3) use more problem-solving and contending strategies, and that the differences in contending and problem-solving would contribute to (4) higher deal rate and (5) better deal quality. Finally, we expected that the differences in their strategies would mediate the effect of the experimental condition on (6) the odds of optimal agreement and (7) the deal quality.

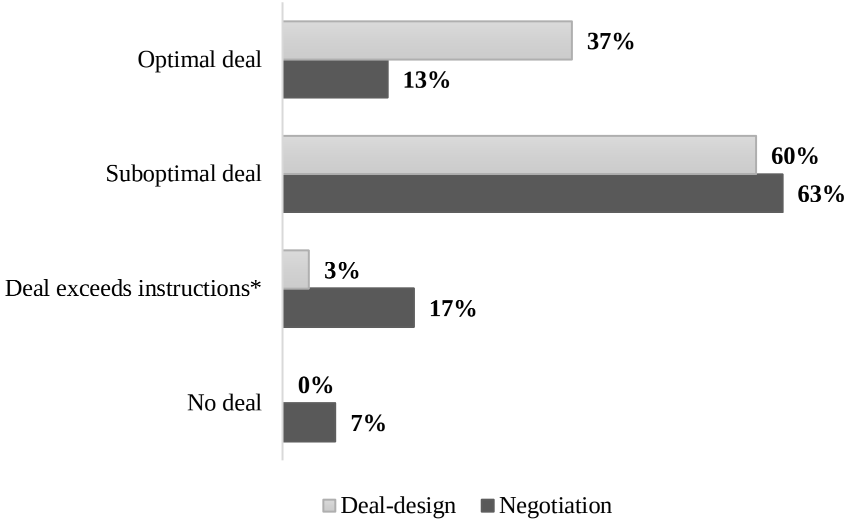
Results

Impasse rate and quality of agreements

As predicted, the deal-design group concluded more optimal agreements (37%) than the negotiation group (13%); Pearson $\chi^2(1, 60) = 4.36$, Fisher's exact $p = .036$; proportion test $z = 2.09$, $p = .018$. The deals were also of better quality in the Deal-design than in the Negotiation condition; $F(1, 58) = 8.76$, $p = .005$, $\eta^2 = .13$, 95% CI [.014, .29]; Kruskal-Wallis $\chi^2(2) = 7.42$, $p = .006$ (**Figure 31**).

The representatives of the Target were more successful in value claiming than the representatives of the Acquirer and claimed on average 57% (*SD* = .40) of value; $t(114) = 1.93$, p = .056, $d = .36$.

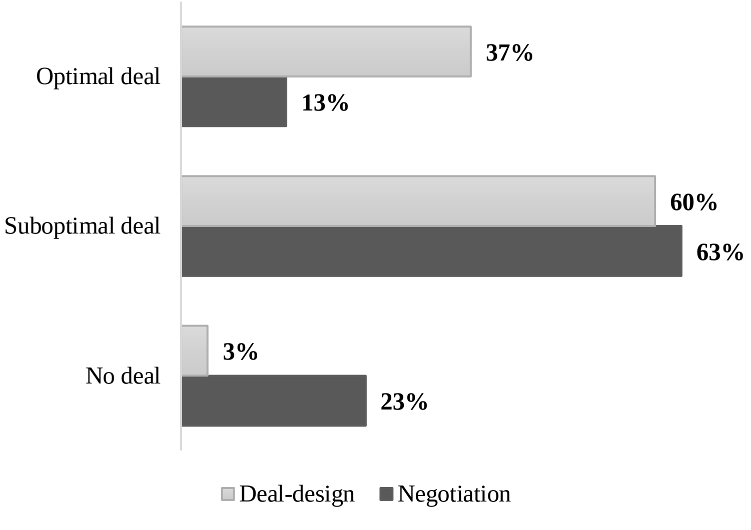
Figure 31. Quality of deals in Negotiation and Deal-design conditions



Note. Instances where the parties reached an agreement, but one of the parties exceeded instructions and incurred a loss, e.g., the acquirer paying a purchase price higher than 16 million.

If deals that exceeded instructions were treated as no-deals, the situation was the same; $F(1, 58) = 8.44$, $p = .005$, $\eta^2 = .13$, 95% CI [.01, .29]; Kruskal-Wallis $\chi^2(2) = 7.34$, $p = .007$ (**Figure 32**).

Figure 32. Quality of deals in Negotiation and Deal-design conditions



Note. No deals include instances where the parties reached an agreement, but one of the parties exceeded instructions and incurred a loss.

In the rest of the paper, we only report results for the variable that treats deals exceeding instructions separately.

Impact of condition on strategies

One-way MANOVA indicated an effect of the condition on problem-solving and contending; $F(3, 110) = 3.94$, $p = .010$, $\Lambda = .90$. As predicted, participants in the deal-design condition used more problem-solving and less contending than their 'negotiation' peers. See **Figure 33** and **Table 24** for details.

Figure 33. Contending, avoiding and problem-solving in Negotiation and Deal-design conditions

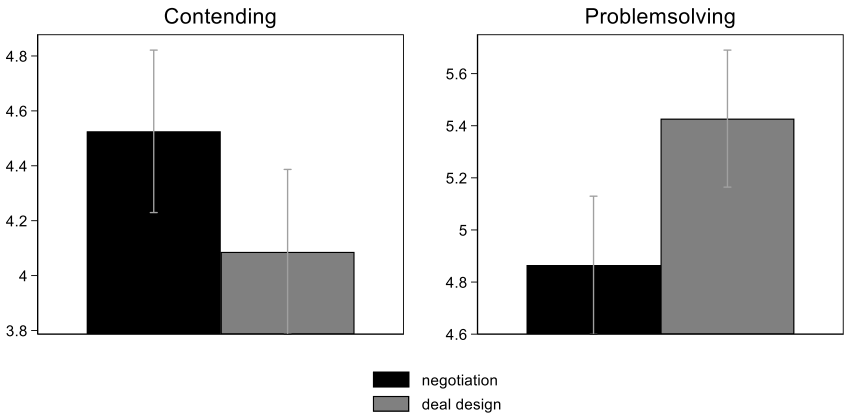


Table 24 Means and test statistics for individual problem-solving and contending in Deal-design and Negotiation conditions

	Problem-solving		Contending	
Deal-design	5.43	(.98)	4.09	(1.12)
Negotiation	4.86	(1.03)	4.53	(1.15)
t		2.98		-2.06
p		.002		.021
d		0.56 [.243, .871]		0.39 [.074, .697]

Note. T-test *p* values are one-tailed. Effect size 95% CIs are provided in square brackets.

Mixed-effects regressions with dyads as a random effect in a sample using multiple imputation for missing values showed that the condition significantly predicted problem-solving (*b* = .563, 95% CI [.17, .96], *t* = 2.81, *p* = .005) and contending (*b* = -.480, 95% CI [-.93, -.027], *t* = -2.08, *p* = .038).

Impact of strategies on odds of optimal deal and deal quality

Negotiators' strategies and odds of optimal agreement

As expected, contending strategies significantly reduced the odds of reaching an optimal deal. The interaction term was significant, too. The model explained 26% of variance in the odds of optimal agreement (**Table 25**). Contrary to predictions, problem-solving scores were not predictive.

Table 25. Impact of Acquirer's and Target's contending on the odds of optimal deal

	b	*SE*	*z*	*p*	*.95 CI*		*OR*
Contending (A)[a]	-2.24	1.23	-1.82	.035	-4.26	-0.22	0.107
Contending (T)[c]	-4.09	1.53	-2.68	.004	-6.60	-1.58	0.017
Interaction term	0.64	0.29	2.18	.029	0.06	1.22	1.900
Model χ^2	14.81	p = .002					
Pseudo R^2	.26						

Note. [a] Contending score of the Acquirer. [b] Contending score of the Target. The results of the full set using multiple imputation was virtually identical.

The interaction between the contending of the Acquirer and the Target (see **Figure 34**, **Figure 35** and **Figure 36**) indicated that high levels of the Target's contending (e.g., mean or higher) always resulted in a low probability of an optimal deal. However, the Acquirer's reciprocation of high contending marginally improved the odds of optimal outcome rather than further decreased them. This is consistent with suggestions in the literature that unilateral contending yields suboptimal outcomes (Carnevale & Pruitt, 1992; Pruitt, 1983b). We examine it in the discussion section.

Figure 34. Interaction effects between forcing scores of Target and Acquirer on the probability of optimal deal

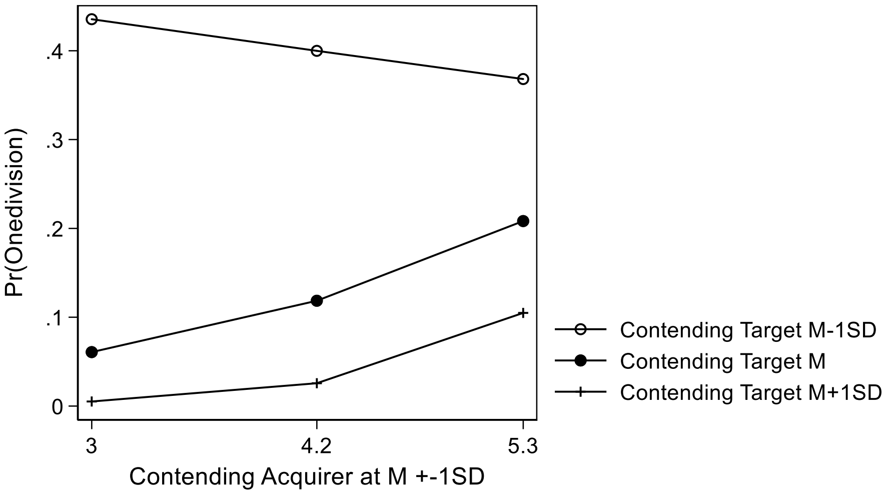


Figure 35. Interaction effects between forcing scores of Acquirer and Target on the probability of optimal deal

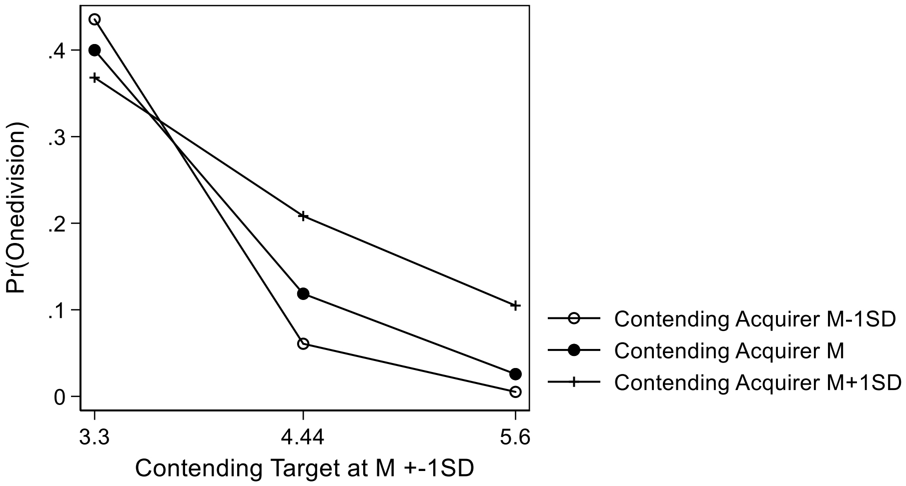
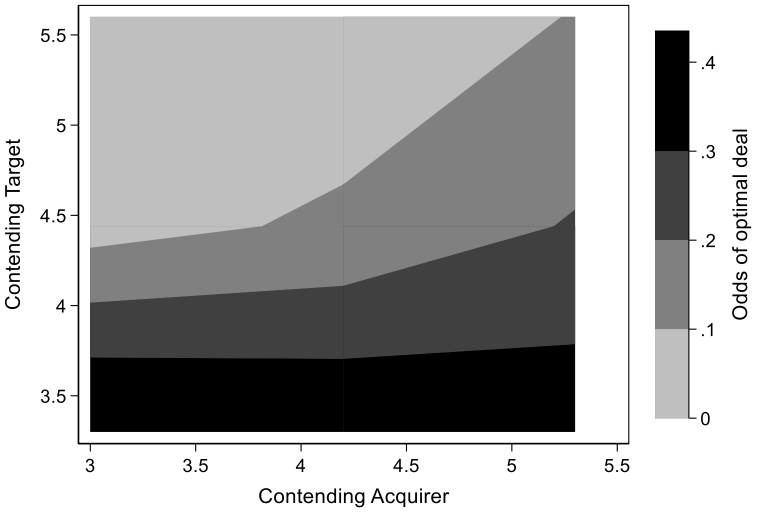


Figure 36. Contour plot of the impact of the interaction between contending scores of Acquirer and Target on the probability of optimal deal



Negotiators' strategies and deal quality

As predicted, the Target's contending was negatively associated with deal quality while their problem solving had a positive impact (**Table 26**). This was not true for the representatives of the Acquirer. Interactions were not significant.

Table 26. Impact of Target's contending and problem-solving scores on the quality of the deal

	b	*SE*	*z*	*p*	*.95 CI*		*OR*
Contending (T)[a]	-0.88	0.28	-3.15	.001	-1.33	-0.42	0.416
Problem-solving (T)[b]	0.54	0.28	1.94	.027	.082	1.01	1.723
Model χ^2	16.34	$p < .001$					
Pseudo R^2	.14						

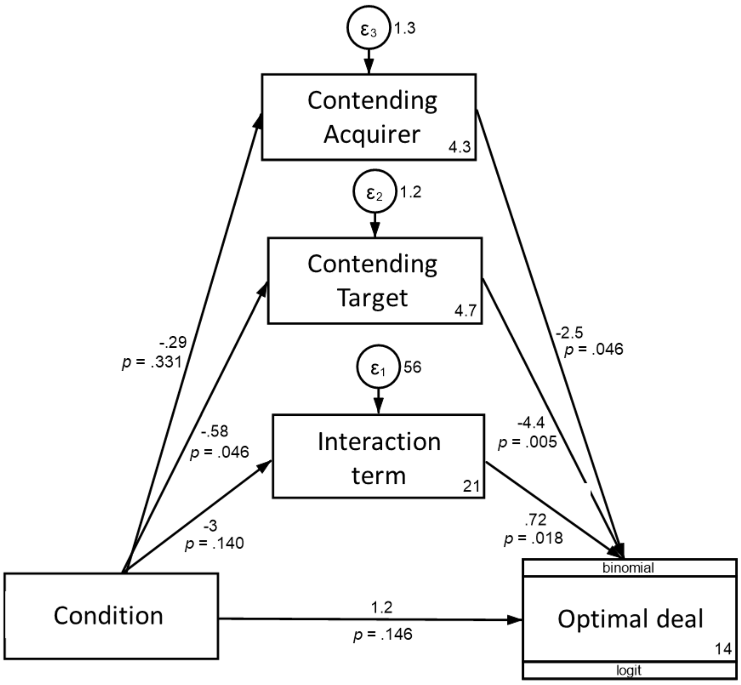
Note. [a] Contending score of the Target. [b] Problem-solving score of the Target. When missing values were imputed the results were virtually identical.

Mediating effect of strategies

Optimal deal

Structural equations model showed that while the contending of the Acquirer and the Target as well as the interaction term influenced the odds of reaching an optimal deal, only the (reduced) contentiousness of the Target had been both a consequence of the experimental manipulation (see **Figure 37** for details) and mediated the effect of the experimental manipulation (which as the result became marginally significant at *p* = .073). The indirect effect was significant and explained 64% of the total effect. It is worth noting that the impact of condition on both contending of the Acquirer and the interaction term (at a directional *p* = .15 and *p* = .07, respectively) were approaching significance, and both contending of the Acquirer and the interaction term predicted the odds of optimal deal (at *p* = .023 and *p* = .009, respectively).

Figure 37. Impact of contending of Acquirer, Target and the interaction term on the probability of optimal deal

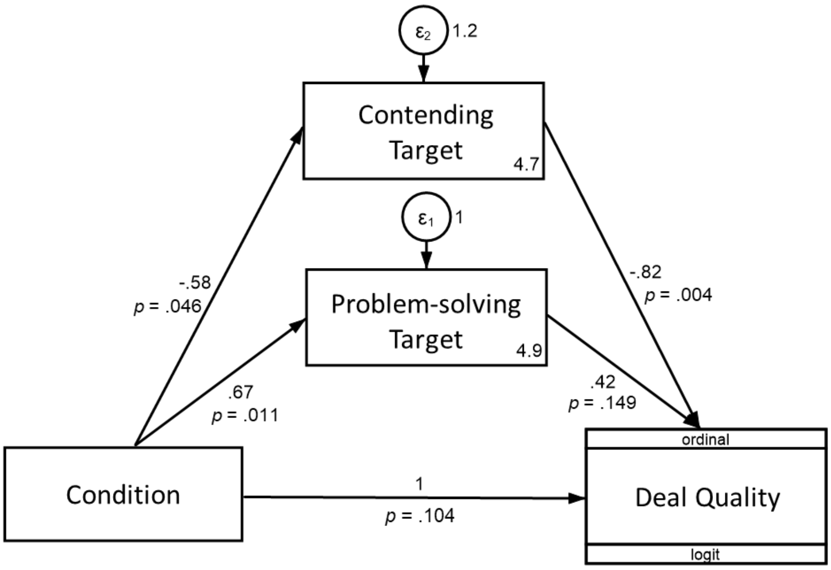


Note. The *p*-values are two-tailed (non-directional).

Deal Quality

The effect of the Target's contending on deal quality was significant. The Target's problem-solving was marginally significant (p = .075). The Target's contending mediated the effect of the experimental condition (that became marginally significant at p = .052, see **Figure 38** for details).

Figure 38. Impact of contending and problem-solving of Target on the quality of the deal



Note. The *p*-values are two-tailed (non-directional).

Summary

As predicted, the participants in 'negotiation' condition, compared to their 'deal-design' peers, reached inferior outcomes and employed more competitive and less problem-solving oriented strategies. Further, as per our hypotheses, the odds of optimal outcome and deal quality were adversely affected predominantly by the contending of the negotiator representing the Target, explaining 26% of the variance in odds of optimal deal. Contrary to our hypotheses, there was no such effect in Acquirers and problem-solving did not increase the incidence of optimal deals. However, problem-solving of the Target contributed, jointly with contending of the

Target (which contributed negatively), to the overall deal quality. Again, the Acquirer's scores were not significant. Finally, the contending, particularly of the Target, was both triggered by the 'negotiation' (and 'problem-solving') frames, and contributed to lower (higher) odds of optimal deal and deal quality.

Discussion

The findings of this study are consistent with previous studies we presented in this thesis. The 'negotiation' frame, compared with an alternative, collaborative frame, resulted in inferior outcomes (Study 4) and in more competitive (contending) and less collaborative (problem-solving) strategies (Study 5). This study replicates and extends the findings of these previous studies and suggests that it is increased contending associated with 'negotiation' construals that is particularly harmful for bargaining outcomes.

Participants with the 'deal-design'-framed acquisition task completed better quality deals than their peers tasked with 'negotiation' and completed almost *three times* the number of optimal deals (37% vs 13%). They also used more problem-solving and less contending strategies when tackling the task. The probability of reaching an optimal deal was associated with the reduced contending of both parties, and the deal quality was positively associated with problem-solving and negatively with contending of the representatives of the Target only. Finally, the contending of the Target mediated the impact of the experimental condition on both odds of optimal outcomes and deal quality.

While the Target's problem-solving and contending were predictive of the overall deal quality, the strategy that drove optimization was reduced contending. In other words, contending was the sole strongest predictor of suboptimal outcomes. Furthermore, structural equation models showed that a 'negotiation' frame triggered increased contending (relative to the 'deal design' frame), and that such increased contending harmed both the odds of reaching an optimal deal and deal quality, thus fully mediating the effect of the experimental manipulation on outcomes. This suggests that the critical difference – i.e., the difference that made a difference so to speak – between the effects of the 'negotiation' and the alternative 'deal-design' frames was the level of contending generated in the participants.

Stubborn insistence and imposing one's own view blocks creativity, makes it hard to accept new options, and risks triggering escalation (Carnevale & Pruitt, 1992; Pruitt, 1983b), and turned out to be the *sine qua non* bottleneck for generating the optimal option. The interaction effect suggested that the Acquirer's reciprocation to the Target's high levels of contending marginally improved the odds of an optimal outcome rather than further decreased them. In other words, dyads where both parties were contentious did marginally better than dyads where only one party was, and the other one rolled over. This is consistent with classic theoretical accounts showing that situations where an aggressive party negotiates with a 'soft' opponent are particularly prone to suboptimal outcomes (Pruitt, 1983b).

The importance of these findings is not so much in showing that priming problem-solving improves baseline bargaining outcomes, but in highlighting the aspects of the default negotiation mindset that interfere with effective negotiation. The alternative 'deal-design' frame in this study – just like the 'problem-solving' primes in Studies 4 and 5 – did not introduce any novel strategies the participants might not have known (such as, for example, making simultaneous offers, exploiting preferences or attempting post-settlement settlements). Instead, the primes invited the participants to employ creativity and problem-solving and 'jointly with the other party' attempt to design a deal that serves their individual interests. These approaches mirror the cooperative approaches that the lessons from game theory and prescriptive negotiation literature suggests are superior to competition in complex negotiation settings (e.g., R. Axelrod & Dion, 1988; R. Axelrod & Hamilton, 1981; Fisher et al., 1991; Lax & Sebenius, 2006). The findings of this study highlight in which ways the default negotiation strategies deviate from the prescriptive model: excessive contending and underutilized problem-solving.

The results of this study, in conjunction with the findings of Studies 4, 4b and 5, provides support for our theoretical model stating that prior expectations drive competition in negotiation. Under Friston's free energy model, the brain is an inference, error minimizing machine explaining sensorium using prior expectations and the incoming sensorium (pp. 192-197). Because negotiation situations are particularly uncertain and ambiguous (p. 198), perceptual and active inference cannot determine the cause of sensorium, which results in cognitive penetrability of priors (p. 195 and p. 201). Such prior expectations in negotiation tend to be competitive (pp. 197-201). The parties therefore perceive that their task to be a competitive distribution of a fixed resource (p. 203). As the result, the normative strategy is to contend (p. 204), value optimizing is out of focus (p. 204), perceiving a situation as 'negotiation' impairs trust (p. 207) and competing hypotheses are explained away (p. 209), and the outcomes are poor.

In deal optimization (and in reaching overall deal quality), the effects of the level of contending of the Target (and problem-solving) were larger than those of the Acquirer. The reason is the asymmetry between the roles. The Target has more information about the critical aspects of the case (see **Table 22**, p. 245). While the Acquirer needs the acquisition to achieve horizontal integration, this interest can be met by the default option of buying the whole Target company. The Target's interest of M&A-level 40-50% margins, however, can be met if only the critical division is sold (as the ratio of value created to the division value allow such large margins, see **Table 23**, p. 245). Finally, the context in which the parties negotiate is somewhat of a sellers' market. The Target does not need to sell unless it is at the right price, whereas the Acquirer needs the skin-cream division as part of their horizontal integration strategy. In M&A, the sellers' market is not uncommon in situations where target companies have considerable brand equity, valuable technology or a market share in the relevant niche, which is part of the case context. Such sellers' market often manifests in better value-claiming performance of the targets, which is what we found in this study too. For these reasons, the Target is the more critical role when it comes to optimizing.

The current research has several limitations. First, the theory tested involves states of mind, but we tested hypotheses related to outcomes and strategies. The actual mindsets remain unmeasured. Future studies should consider tackling the challenge of recording the mainly implicit construals and mental processes. Second, while the generalizability of our findings is somewhat better than those of Study 4 alone because the M&A task used here has better ecological validity (it is based on a real-life case, and the priorities are clearly connected to the parties' underlying obvious conflicts and hidden compatibilities), it is still an artificial task with the key parameters outlined on one page, with parties whose interests are fictional rather than real. Third, our participants were students, lacking work experience, and studying a competitive topic of law. They had been recruited at the end of their semester-long course in negotiation, after having heard lectures and negotiated cases demonstrating the interest-based nature of negotiation (see the methods section, p. 242, above). However, as noted in the limitation sections of our previous studies, artificial tasks and student participants have been an effective research tool as they allow inference about underlying cognitive processes (De Dreu & Carnevale, 2005b), and while we perhaps cannot generalize the effects' magnitude, we probably can generalize effects: framing a situation as 'negotiation' results in competitive strategies and inferior outcomes. Fourth, the context of the case was M&A, a notoriously cut-throat and competitive environment. It is hard to evaluate the impact of that on the baseline 'negotiation' construals and the 'deal-design' prime. Finally, the participants in the 'deal-design' condition might have taken a hint from the prime that there was something possible that was more than a flat-out distributive deal, and that this was the key driver of the change in outcomes. We think this is unlikely as the results of this study are consistent with the findings of Studies 4 and 5 that had no such difference in primes. Also, the assumption that there is 'some value hidden' and that you can do something clever to capture is exactly what is missing in the 'negotiation' mindset.

Overall, Study 6 replicates and extends the findings of Studies 4 and 5 and provides support for our theory that construing an interdependent, mixed-motives task with value potential as 'negotiation' triggers competitive mindsets that result in competitive strategies and poor outcomes.

REMARKS AND CONCLUSIONS

Combined, our studies (Study 4, 4b, 5 and 6) show that construing an interdependent mixed-motive task with value potential as ‘negotiation’, compared with an alternative collaborative frame (‘problem-solving’ and ‘deal design’), results in competitive strategies that lead to suboptimal outcomes.

In the first and second study (Study 4 and 4b), the participants who engaged in a ‘negotiation’- labelled task achieved lower integrative and compatible gain than their peers who whose task was labelled ‘problem-solving.’ Increased trust only partially improved the outcomes (compatible gain, but not integrative gain) as it presumably affected only the parties’ perception of their counterparties while the task remained a competitive ‘negotiation’.

While these results clearly show that the outcomes under the ‘negotiation’ construal are suboptimal to ‘problem-solving’ one, the study did not investigate the difference in strategies that led to the difference in outcomes. The second study in this set (Study 5) addressed that question. The findings suggested that the participants whose multi-issue task was labelled ‘negotiation’ endorsed more competitive strategies than their ‘problem-solving’ peers. Specifically, they recommended more contending and less problem-solving strategies, and were more likely to engage in avoiding behaviors.

In the third study in our set (Study 6), we investigated which strategies, endorsed by the participants engaging in a ‘negotiation’- and ‘deal-design’ framed tasks, lead to optimization, and which of them are affected by the change in construal. The results were consistent with Study 4, 4b and 5; the ‘negotiation’-primed participants achieved worse outcomes and endorsed more competitive strategies (higher contending and lower problem-solving) than their ‘deal-design’ primed colleagues. Contending turned out to be the strategy that was both, first, affected by the task construal and, second, acted as the key inhibitor of optimization.

The four studies combined provide solid support for our theoretical proposition that construing an interdependent mixed-motive task with value potential as a ‘negotiation’ results in mindsets that are predominantly focused on distributing value rather than finding ways of creating it. First of all, the value-creating aspects of outcomes were significantly depressed in ‘negotiation’-labelled tasks compared with a more problem-solving-oriented frames; the participants who were ‘problem-solving’ achieved significantly higher integrative and compatible gain than the ones who were ‘negotiating’ (in the region of 30%, see Study 4), and the ‘deal design’ group concluded three times as many optimal agreements as the ‘negotiation’ group (Study 6). Second, the participants who were asked about efficient strategies in a ‘negotiation’ task endorsed more contending and less problem-solving than their ‘problem-solving’ peers (Study 5). This clearly reflects the competitive mindset that is appropriate for a zero-sum task. Finally, the strategy that was both differentially triggered by the ‘negotiation’ and ‘deal-design’ primes, and significantly suppressed deal optimization, was contending (Study 6). Contending consists of attempts to *impose one's will* on another with positional commitments, persuasive arguments, and threats and bluffs (De Dreu et al., 2001), a strategy appropriate for zero-sum, distributive tasks, characterized as ‘contests of will’ (Fisher et al., 1991). In short, value claiming is all there is, and ‘[t]o win at negotiating – and thus make the other fellow lose – one must start high, concede slowly, exaggerate the value of concessions, minimize the benefits of the other’s concessions, conceal information, argue forcefully on behalf of principles that imply favorable settlements, make commitments to accept only highly favorable agreements, and be willing to outwait the other fellow. The hardest of bargainers threaten to walk away or retaliate harshly if their one-sided demands are not met; they may ridicule, attack and intimidate their adversaries’ (Lax & Sebenius, 1986b, p. 33).

This is also consistent with our proposition that the likely mechanism behind this is the overweighing of priors in negotiation. Negotiation is an exceedingly complex, ambiguous and uncertain setting (see p. 87 and 198): the information/negotiator’s dilemma (Lax & Sebenius, 1986b; Murnighan et al., 1999), the observational opacity of other’s mental states and the exacerbated opacity of negotiators’ intent (see p. 199), and our inability to detect lies (Bond & DePaulo, 2006) render perceptual and active inference impotent. The perceptual processes then rely heavily on prior expectations (Friston, 2010), and these expectations are that negotiation is a conflict-ridden distributive contest (e.g., O'Connor & Adams, 1999) where the payoffs are zero-sum (Thompson, 1991; Thompson & Hastie, 1990) and our interests cannot possibly be compatible (Thompson & Hrebec, 1996).

In conclusion, our studies further demonstrate the specific impact of two kinds of interventions aimed at mediating the harmful effect of the ‘negotiation’ construal: increasing trust has a selective impact on compatible issues, and reframing a task as ‘problem-solving’ or ‘deal-design’ has a more robust presumed effect of triggering a collaborative mode of approaching the task. Finally, the studies provide empirical support for long-standing prescriptive advice that collaborative strategies are more likely to optimize outcomes than the default competitive ones. This has potentially considerable practical importance. Approaching a bargaining situation as a problem to be solved, rather than a distributive ‘negotiation’, and taking an interest in the counterparty’s agency thus establishing epistemic trust, may help reduce the paradoxically suboptimal outcomes in situations where good results are possible and the parties are motivated to reach them.

Social cognition in negotiation: concluding thoughts

IN CONCLUSION

The following sections summarize our theory and research, and discuss their backdrop and implications. We start by providing an overview of our proposition and the empirical results supporting it. We then outline the major implications and consider how the findings fit into our overarching theory of cognition in negotiation. We propose that a part of human automatic cognition is a particular conflict-endorsing heuristic which gets triggered in situations that we call ‘negotiations’. We describe this heuristic and the biases that emanate from it, before moving on to two critical outstanding questions, first, what exactly triggers the competitive heuristic in negotiation settings and, second, why is the heuristics competitive rather than collaborative. We continue with the implications of our research for training and outline questions for future research.

Mentalizing in negotiation

We developed and tested an original theory that mentalizing – representing and thinking about people in terms of motivational and epistemic mental states – underpins the key processes in negotiation. Decision making in bargaining is about choosing behavior that is the likeliest to maximize negotiators’ interests, and depends on what the negotiators want and what they believe is the best way of getting what they want. Because these wants and beliefs these are motivating and epistemic mental states that at least partially depend on the wants and beliefs of the counterparty, the ability to infer and represent them, in both oneself and the counterparty, is perhaps the critical capacity in negotiation. Mentalizing underpins bargaining.

The studies provided solid evidence supporting this claim. The capacity to mentalize predicted the individual and joint success (conditional on training) in a multi-issue task, in a distributive task and in a partisan-perception driven dispute-resolution task. Cognitive reflection, the metacognitive capacity to engage controlled cognition where necessary, assisted negotiators' gain. The interpersonal understanding of each other in terms of our thoughts and feelings, facilitated by mentalizing, is at the heart of interest-based negotiation.

In the second part of the thesis, we tested the theory that social cognition in negotiation crucially depends on how individuals construe the interdependent, mixed-motive task that they are facing. Because negotiation situations are highly uncertain and ambiguous, prior expectations play a decisive role in perception and action. In 'negotiation', the priors are competitive and bias social cognition accordingly.

The studies supported this theoretical proposal. In an identical multi-issue task, the negotiators whose instructions were labelled 'negotiation' achieved inferior integrative and compatible gain and endorsed more competitive and less collaborative strategies than the participants whose task was labelled 'problem-solving'. In a task with hidden value potential, the participants who were 'negotiating' used more contending and less problem-solving, and reached inferior outcomes than their 'deal-design' peers. The critical strategy associated with 'negotiating' turned out to be contending (it was both elicited by the 'negotiation' frame and harmed the outcomes the most, see pp. 254-255). This unequivocally supports our proposition that understanding an interdependent mixed-motive situation as a 'negotiation' results in negotiators adopting competitive mindsets.

Or research also contributes to understanding the paradoxical inefficiency of negotiators. We stated a few times in this thesis that while good outcomes are possible, we tend to fail at capturing the available value and instead waste opportunities and increase social conflict (e.g., p. 78). Our research indicates that a part of the reason is likely a combination of the individual differences in the social cognitive capacity and certain universal modifiers of how social cognition is applied in negotiation.

Our individual difference-based research indicates that the social-cognitive traits (the capacities for cognitive reflection and mentalizing) correlate with success in negotiation. However, the effect mentalizing on outcomes is conditional on training. This suggests that poor outcomes are partly due not only to the low levels of social cognitive capacities in parts of the population, but to the fact that some negotiators simply do not use mentalizing to negotiate at all. The upside is that this is treatable:

training improves the effect of mentalizing and increases cognitive reflection. We discuss this in more detail below (see p. 273).

Our experimental investigation of the impact of how negotiators construe (perceive) a task on their strategies and outcomes shows that it is not the features of the situation that drive competition and poor outcomes. The ambiguity and uncertainty of the negotiation situations may conspire to make solutions nonobvious and difficult to find, but the competitive interpretation of such a situation – and the resulting contending rather than problem-solving – is a uniquely human contribution. In other words, it is not the task, it is us. Perhaps the best place to start improving our poor negotiation outcomes is by recognizing this propensity for interpersonal competition in mixed-motive tasks.

Cognition in negotiation

Constraints on cognitive processing in the domain of general (non-social) cognition have been studied extensively. Since Nobel laureate Herbert Simon's proposition that people strive to be rational, but fail ('bounded rationality'; Simon, 1955, 1956), human judgment and decision-making have been the focus of a vast 'heuristics and biases' research focusing on systematic violations of rationality in a wide array of individual thinking tasks (e.g., incorrect probability assessments, faulty hypothesis testing, context dependency, framing; Kahneman, 2011). The failings in rational decision-making have been largely understood under the framework of the dual process theory (for an overview see Stanovich, 2011), which sees human judgment and decisions in terms of a tradeoff between computational power and expense. Controlled cognition that provides a higher chance of optimal decisions can only be engaged in a fraction of the situations the decision maker faces, because it entails a heavy computational cost and, more importantly, has a limited budget and is experienced as aversive (Frederick, 2005; Kahneman, 2011; Stanovich, 2011; Stanovich et al., 2016; Toplak et al., 2014). One of the consequences is that the vast majority of the available sensorium is rapidly processed by heuristic-based *implicit* automatic cognition ('bounded rationality'; Simon, 1955, 1956) or simply blanked out ('bounded awareness'; Bazerman, 2013; Bazerman & Chugh, 2006; Chugh & Bazerman, 2007; Idson et al., 2004; Simons, 2010; Simons & Chabris, 1999).

Bounded rationality has been studied in negotiation, too, and the research questions and paradigms unfortunately largely mirrored the questions and paradigms in the heuristics and biases research in non-social settings (e.g., Bazerman & Neale, 1991; Carroll, Bazerman, & Maury, 1988; Idson et al., 2004; Neale & Bazerman, 1992b). For example, in their popular *Negotiating Rationally*, Bazerman and Neale (1992) discussed how negotiation is affected by irrational escalation, framing, zero-sum bias, the winner's curse, overconfidence, information availability and anchoring (Bazerman & Neale, 1992). However, no research so far, to the best of our knowledge, systematically studied negotiation from the perspective of decision-making driven predominantly by social-cognitive inference and computation that involves the areas of the brain in charge of social decision-making (Lieberman, 2007). To the extent mentalizing underpins bargaining – and we have a solid theoretical basis and solid preliminary empirical support that it does – such research is necessary.

For negotiators, negotiation presents a chaotic computational overload. If nonsocial judgment and decision making are subject to information overload, this is *a fortiori* also true for judgment and decision making in negotiation that involve, in addition to nonsocial cognitive tasks, mentalizing minds of oneself and the counterparty under strategic interdependence, mixed motives and, often, time pressure and stress. Also, unlike in the heuristics and biases non-social tasks, both value claiming and value creating solutions depend not on the focal decision-maker (negotiator) alone, but are a function of the decisions of all negotiating parties. We discussed how negotiation situations are uncertain and ambiguous earlier in the thesis (*Uncertainty and ambiguity in negotiation* p. 198). Negotiators do not know the counterparty's interests and preferences, and consequently do not know the value potential of the situation. Also, their own interests and preferences need to be articulated and ranked. More importantly, counterparty-focused mentalizing – required for both collaborative and competitive purposes (see *Mentalizing in negotiation*, p. 68) – is especially difficult not only because mental states are twice removed from objective reality (one needs to infer behavior from the sensorium and then mental states from behavior), but because value-claiming negotiators mask their behavior to look like value-creating, and such misdirection is virtually impossible to detect from nonverbal cues (see *To catch a liar*, p. 200).

The excessive complexity and ambiguity of negotiation settings presents negotiators with a perception problem that translates to an action choice problem. We used the active inference framework to explain how in such situations the uncertainty is resolved by relying on prior expectations (see *Ineffective perceptual and active inference cause overweighing of priors*, p. 201) and suggested that the key prior in negotiation situations is that the task is a competition. While this resolves the perception and action problem (the task is a competition and the normative action is to contend), the bias primes, and oftentimes turns the negotiation interaction into contending match where the strategies required to generate value are out of focus. The competition-biased perception extends to mentalistic inference of others' actions to the extent that good-faith communication runs a risk of being (mis)perceived as value-claiming, and the real competitive moves are experienced more malignant than they are and understood in terms of the counterparty's personal traits (the fundamental attribution error, see pp. 84-87). Also, because both parties are subject to the same conflict heuristic and bias, this likely exacerbates the problem: perceived contending breeds real contending and results in an escalatory spiral of competitive moves and, first perceived and then real, zero-sum contest. In other words, the negotiator's dilemma (Lax & Sebenius, 1986b) that we outlined at the beginning of the thesis as one of the critical aspects of negotiation (*Tension between creating and claiming value*, p. 56) is exacerbated by priors-weighted mentalizing: the competitive bias paints ambiguous actions in competitive colors. This self-fulfilling prophecy explains why the zero-sum family of biases and suboptimal agreements are so obdurate and resistant to correction.

Under this view, the well-researched fixed pie (e.g., Pinkley et al., 1995) and incompatibility biases (e.g., Thompson & Hastie, 1990), as well as the competitive expectations in negotiation settings (e.g., O'Connor & Adams, 1999) are *corollaries*, or *consequences,* of the competitive construal. If the situation we are in is a conflict, then there is no way of satisfying both you and me (incompatibility and fixed pie biases), we can expect positional bargaining and frequent impasse, and the only relevant strategy is competitive.

Our proposition is that when a task is understood to be a ‘negotiation’, in order to deal with the information overload and the situational uncertainty and ambiguity, negotiators employ an automatic social-cognitive heuristic that we call the *conflict heuristic*: the task is to outcompete the opponent on the other side, the payoffs are zero sum, there are no compatible options, and the normative strategy is contending. Problem-solving (Menkel-Meadow, 1983), or creative design of options (Fisher et al., 1991) is not so much considered a poor strategy as it is simply not considered. The automatic negotiation mindset has been described by many authors, for example:

> *‘[t]o win at negotiating – and thus make the other fellow lose – one must start high, concede slowly, exaggerate the value of concessions, minimize the benefits of the other's concessions, conceal information, argue forcefully on behalf of principles that imply favorable settlements, make commitments to accept only highly favorable agreements, and be willing to outwait the other fellow. The hardest of bargainers threaten to walk away or retaliate harshly if their one-sided demands are not met; they may ridicule, attack and intimidate their adversaries' (Lax & Sebenius, 1986b, p. 33).*

No research so far, to our best knowledge, conceptualized the critical issues in negotiation in terms of heuristics. Such research seems necessary, especially given the similarities between the interaction of automatic and controlled cognition in social and non-social tasks (e.g., Lieberman, 2007; Lieberman et al., 2002; Ochsner & Lieberman, 2001; Satpute & Lieberman, 2006), and considering how important and impactful the non-social heuristics and biases research has been for the society.

Two questions regarding the conflict heuristic

What triggers the use of the conflict heuristic in negotiation settings? In other words, what causes the employment of competitive priors to resolve the perceptual ambiguities in mixed-motive, interdependent tasks we call ‘negotiations’? In our studies we triggered differential construals – including the ‘negotiation’ one that resulted in competitive strategies and suboptimal outcomes – by labelling manipulations in task instructions. However, in real life, priming is not so easily discernible. So, what is the trigger? There must be something about the interaction the individual is contemplating or engaging in that primes the mental model we call ‘negotiation’. We speculate that these aspects are most likely the competitive aspects of the interaction that are normally most salient, unlike the value potential which tends to be hidden. These features then trigger the mental model of ‘negotiation’, including the competitive priors and what we termed the conflict heuristic. The key aspect of such priming is ‘the passive acceptance of the formulation’ that is suggested (Kahneman, 2003, p. 703) where negotiators automatically take the conflict construal of the task as focal and all thought and action derives from it rather than from any alternative frames that remain out of focus, presumably because exploring them would require extra effort (Stanovich et al., 2016, p. 54). Within such a competitive frame, the negotiators’ cognition and action might be fully controlled and rational (e.g., making optimal choices to claim value), but will leave out of focus alternative collaborative task frames. This is an instance of what Stanovich et al. (2016) call ‘serial associative cognition with a focal bias’, a tendency to over-economize while engaged in controlled processes. What is biased is the frame, not the cognition within the frame. The focal frame that dominates all processing is normally the most easily constructed model, and it clearly appears that such models are what people already believe and have modelled previously.

This brings us to a key question: why are the prior expectations in negotiation competitive rather than cooperative? Why does the automatic heuristic suggest conflict rather than collaboration? Why does the ‘negotiation’ mindset, once triggered, contain competitive rather than collaborative expectations and endorses contentious rather than cooperative action? Two speculative explanations come to mind. First, the blueprints for both cooperation and competition can be the result of early years’ interpersonal experience. The oscillation between cooperative (friendly) and competitive (aggressive) modes resembles the two distinct ways of functioning – a constellation of anxieties, defences and internal and external object relations understood as a paranoid-schizoid position – in Kleinian psychoanalytic tradition. This splitting has roots in early infancy where the baby, in order to deal with instinctual pressures and frustration splits both self and object into ‘good’ (loved and felt to be loving) and ‘bad’ (persecuting and hated), with literally no integration between these two parts. This is accomplished by projective identification and, at times, evocative behavior that induces the recipient of projection to feel and behave as per the projected content (Klein, 1921, 1930). Omnipotent denial and idealization are also present. All that is bad is denied in the self and felt to be out there where it is projected, and the good experience is idealized and exaggerated (Roth, 2001). At a later point in development babies achieve what Klein called the depressive position, characterized by merging of the ‘good’ and the ‘bad’ and realizing that both self and object are both good and bad (Klein, 1935). However, oscillation between paranoid-schizoid and depressive functioning is something that persists throughout life. Particularly in stressful situations, or possibly under cognitive load, adults are likely to regress to paranoid-schizoid functioning, deny the bad and project it out where it is safe to be hated. It is very possible that the aggressive archetypal projection (the ‘bad’ breast) provides the blueprint for competitive mindsets in negotiation (where the counterparty is the ‘bad’ adversary also containing the split-off ‘bad’ part of self) whereas a depressive position is required to appreciate both the value creating and value claiming potential. Yet other possibilities are that the blueprints are provided by lateral experiences in love and hate with siblings (Mitchell, 2013) or through a ‘rough and tumble’ play acting out the ‘Play’ drive (Panksepp & Biven, 2012). More conceptual work is needed to elaborate these ideas.

The second explanation is evolutionary. The competitive and collaborative cognitive modes in negotiation bear marked resemblance to the different stages of human intentionality in interpersonal contexts provided by the cultural accounts of human evolution (Tomasello, 2014, 2016, 2019; Tomasello et al., 2012). The phylogenetically older mode of individual intentionality mobilizes all cognitive resources exclusively for a competitive pursuit of individual goals, and thus represents a sort of an archetype of zero-sum negotiation. This mode is still dominant today in great apes. For example, when chimpanzees observe a human reaching for a bucket, they are able to infer that the food is in that bucket in competitive settings, whereas if they observe the same in collaborative environments they remain completely oblivious (Bullinger et al., 2011; Hare & Tomasello, 2004). While such competitive cognitive deployment was effective in interactions that chiefly consisted of competition for food and mates, it became maladaptive in managing interpersonal situations with value potential (e.g., the 'stag hunt'; Bullinger et al., 2011). Evolutionary pressures selected joint intentionality, a cognitive mode that underpins joint attention, representation of common goals and formation of complex perspectival representations (e.g., simultaneously keeping in mind one's own and the other's perspectives on various issues). Individual and joint intentionality modes coexist in the mental repertoire of modern humans, reflecting a history where an individual's interests in interpersonal settings were satisfied in different ways at different times. Today the default mode is predominantly cooperative. For example, 18-month old children happily help non-kin achieve their goals (Warneken & Tomasello, 2006) and 5 year-olds actively coordinate to overcome a prisoner's dilemma (Sánchez-Amaro et al., 2019). Adults, too, seem to default to cooperation unless they think about it first. Rand et al. (2012) demonstrated an implicit cognitive preference for cooperation (also see Tomasello, 2012), which was reversed if people had the time or were instructed to think about the situation carefully, suggesting that the initial automatic response is collaborative, but can be reversed if people think about the task. Also, priming emotions triggers collaboration while priming reason triggers competition (Levine et al., 2018). It appears that understanding a situation as 'negotiation' has a similar effect and recruits the phylogenetically older, competitive implicit model that employs cognitive capacities for competitive purposes, thus reinforcing the zero-sum assumptions and failing to attend to the potential for joint gain. An intentional explicit cognitive intervention (or an automatized one; Bargh & Chartrand, 1999; Bargh, Schwader, Hailey, Dyer, & Boothby, 2012), is necessary to shift to the cooperative, joint intentionality mindset.

We are investigating this in an ongoing research program.

Mentalizing and active inference framework in negotiation

One of the implications of this thesis that invites further research is that the capacity to mentalize and the way we represent the world (the generative model driving perceptual and active inference), particularly in negotiation, are closely connected. In an inferring brain, a part of the higher cortical units' generative model of the world is a model of the minds of others. This model includes the parametrization of the validity of the lower hierarchical units' prediction error signal that comes from sampling of the sensorium as per the current top-down hypothesis. In other words, people differ in the sophistication of their models of others' minds, but also in the degree that the prediction error generated when the mentalized content mismatches prior expectations about others' minds is allowed to modify the higher-level model. In people whose mentalizing is effective, the model of others' minds is sophisticated and the prediction error that comes from a potential mismatch with the sensorium is trusted and allowed to drive model change. This manifests as a marked curiosity about mental states of others and a certain degree of explicit or implicit confidence in one's ability to infer these states accurately. We colloquially refer to this attitude as 'open mindedness': the willingness to change one's mind. In people whose mentalizing is poor, on the other hand, the high-level model is unlikely to both contain nuanced expectations about minds of others and harbor trust in the prediction error that can come from the sampled sensorium. This will manifest in an ineffective mentalizing such as psychic equivalence or pretend mode. We can consider the latter to be an unwarranted responsiveness to prediction error while the former may reflect the imposition of higher order guard model despite significant prediction error. The teleological model is more complicated. Here the individual feels obliged to change the signal from the sensorium in order to reduce prediction error which the generated model evokes with previous versions of reality. Obviously, this is just one part of a far more complex story where the reward circuits interfacing with the mentalizing network (particularly dorsal and ventral MPFC) bias model generation and inhibition of prediction error in favor of models generating greater reward.

Applied to negotiation, the default mentalistic model that depicts the counterparty as a competitor is likely to be more adversarial and resistant to change in somebody with limited mentalizing capacities than in somebody whose mentalizing is relatively effective (or to an extent in situations where mentalizing is difficult, e.g., stress and cognitive load, pp. 37-40). In such situations, negotiators' model of others' minds might be insufficiently nuanced to predict and test for collaborative moves. Even if the counterparty engages in collaborative action, for example by indicating a wish for a cooperative exploration of options, the relevant sensorium might be out of focus and thus not sampled (the chimp who is oblivious about the helpful pointing of the experimenter comes to mind, see p. 42). Even if there is a surprise that comes from an unexpected collaborative action, the prediction error signal would be dismissed as unreliable and the competitive priors would penetrate perception. This is a novel and interesting theoretical proposition that we intend to investigate in the future.

Implications for training

Negotiation seems easy to teach – Google reports 85 million hits to the query 'negotiation' 'training', more than seven times as much as reported by Bruce Patton in 2009 (p. 482) – but this ease is deceptive (Patton, 2009). It is one thing to organize training that the students enjoy; as long as one dishes out case simulations that introduce risk-free negotiation exercises and pepper the debriefs with referencing Getting to Yes (Fisher et al., 1991), it is virtually guaranteed the participants will enjoy the training and rate it positively in feedback forms. It is an entirely different challenge to run a course where the participants learn skills that will make a difference in their negotiation outside of the classroom. There are significant differences between teaching descriptive insights, prescriptive advice and analysis, and instilling in the students that illusive something that improves their negotiations. It is often said teaching negotiation is akin to training a sport like tennis; the history and strategy of the game is one thing, playing it quite another. Knowing is not the same as doing.

This is predominantly because some of the critical processes – construal (perception) and most of mentalistic sense-making – are implicit processes which are unavailable to conscious audit (Nisbett & Wilson, 1977). More importantly, they are difficult to correct. The key aspect of any kind of bias, social or non-social, is that the person is unaware of if not its existence then at least its extent. This is easy to demonstrate in non-social settings. For example, in the classic Shepard's tables task (see the Appendix, p. 302), the advice 'how you see the tables might be misleading' would not affect one's *perception* of the tables' length and depth. In (social) negotiation settings, prescriptive advice such as 'the fixed-pie bias is a fallacy' is unlikely to change a naïve negotiator's perception of the task, or their action. If it did, we would not be the poor negotiators that we are.

Teaching effectively is therefore about changing mindsets and priors. And while descriptive-prescriptive approach (Raiffa, 1982) does inform students about the pitfalls they are likely to encounter and how to address them, this is done at a very abstract level compared to the concrete, fine-grained experience people run into when negotiating. Abstract knowledge is simply too abstract to get any meaningful traction with the moment-to-moment complex negotiation interaction. The knowledge is thus 'idle' (see pp. 209-210).

This suggests that to be effective, teaching ought to affect the fine-grained hypotheses about the negotiation interaction. We believe that this can be accomplished by experiential teaching that focuses on the moment-to-moment negotiation interaction. For example, in addition to telling the students about the fixed-pie bias (abstract prior), the students need to negotiate and debrief a case where the fixed-pie bias is the main barrier to efficiency (experiential learning), observe how experts efficiently negotiate the same case (observational learning), and have an opportunity to negotiate another, slightly different case where the fixed-pie bias is a problem to apply the acquired skills (analogical learning). This is consistent with studies showing that experiential, observational and analogical trainings improve negotiation outcomes (Nadler & Thompson, 2003; Van Boven & Thompson, 2003).

Our training of undergraduate, graduate, and executive populations is based on a mixture of experiential, observational and analogical learning, and the descriptive-prescriptive approach advocated by the decision-perspective to negotiation. Our prescriptive-descriptive syllabus includes the general theoretical frameworks (Patton, 2005) and the three tensions model (Mnookin, 2000), negotiator cognition from the decision perspective to negotiation, and the basic theory of social cognition. The experiential, analogical and observational learning takes place in parallel to the abstract lessons: our students negotiate a number of negotiation exercises on a weekly basis, each designed to elicit a specific type of negotiation challenge. The students receive feedback on value claiming and value creating, quantified and ranked in exercises that have quantifiable outputs, and descriptively if they do not. In our debriefs we focus on both self- and other-focused mentalizing (without labelling it so) by explicating own and other party's interests, and generating options for value creating tasks, and own and other's reservation values, and crafting of effective distributive strategies for value claiming tasks. We often role-play effective strategies or show video recordings of negotiators effectively tacking the challenges in the task.

The evidence in our studies shows such training is effective (see pp. 121-122). Our empirical investigation indicated where some of these effects come from. Training made mentalizing an effective predictor of gains in multi-issue negotiation (training was a moderator of the effect of mentalizing on outcomes, see pp. 122-126). Training also increased the ability of the negotiators to engage controlled cognition where necessary (cognitive reflection partly mediated the effect of training on outcomes, see pp. 129-130). We believe this is because training, *inter alia*, provides negotiators with a framework for controlled thinking about negotiation, focuses their social-cognitive (mentalistic) capacities on the relevant collaborative aspects of the negotiation process, and sensitizes them to detect the areas in negotiation where they ought to exercise cognitive reflection and uncover hidden value.

Future research

Our theoretical propositions and empirical studies significantly extend the existing research. Our theory of mentalizing-based bargaining is the first comprehensive theory that suggests that the key negotiation processes (value creating and value claiming) are critically dependent on social-cognitive inference (mentalizing).

Our empirical studies in Part 1 used a brief self-report measure of mentalizing (the RFQ, see p. 114) and found support for the hypotheses that mentalizing supports both value creating and value claiming. This extends the research outlined in our literature review that showed that perspective taking and empathy, assessed by the Interpersonal Reactivity Index (the IRI; Davis, 1983) and some measures akin to reading the mind in the eyes test (Baron-Cohen et al., 2001), assists negotiation (Elfenbein et al., 2007; Galinsky et al., 2008; Galinsky & Mussweiler, 2001; Gilin et al., 2013; Neale & Bazerman, 1983). Our findings are also consistent with the artificial intelligence bargaining experiments (de Weerd et al., 2013a, 2013b; Weerd et al., 2017) that tested the impact of theory of mind to computer agents in formal mixed-motives games. Finally, our research of the impact of cognitive reflection (see p. 104) on negotiation outcomes is unique; no study, to the best of our knowledge, suggested and tested a theory that the ability to revise automatic responses underpins outcomes in negotiation.

In part two of our thesis, we proposed that the way negotiators construe an interaction exerts a major influence on their mindsets, strategies and outcomes in negotiation. In particular, 'negotiation' construals trigger an understanding of the task as a competition (conflict heuristic) which endorses contending rather than problem-solving strategies and results in suboptimal outcomes. We used the active inference framework to explain why in uncertain and ambiguous negotiation situations, prior expectations exert prejudicial influence on perception and action choice. We then tested whether construing a task as a 'negotiation' versus an alternative collaborative frame, leads to competitive strategies and depressed outcomes, which turned out to be correct. This significantly advances the knowledge in the area as studies so far neither proposed a similar theory nor tested similar hypotheses (the existing studies are limited to game theoretic tasks, see p. 212).

Future research needs to focus on investigating the finer aspects of the relationship between mentalizing and negotiation processes and outcomes. Specifically, it would be worthwhile to assess mentalizing for the purpose of a predictor in linear models with tools different from the RFQ, such as the short and long version of the Adult Attachment Interview (Steele & Steele, 2008), and employing negotiation paradigms covering different negotiation situations and hopefully possessing higher ecological validity. To explore the impact of the ability to engage controlled social cognition, developing a CRT-type performance test focused on social cognitive tasks, would be a worthwhile pursuit. Regarding training, future studies would need to consider which aspects 'focus' the reflective functioning and which aspects increase cognitive reflection in the participants, and how these relate to outcomes on a long-term basis, perhaps through a longitudinal study. Regarding the competitive construal, the key questions that warrant further research are why are 'negotiation' priors competitive rather than collaborative or neutral (i.e., why is the dominant heuristic conflictual) and what exactly triggers such priors. Finally, specific features of the conflict heuristic should be further explored in a variety of tasks and settings.

In addition, the studies we conducted had participants who were university students. While the sample was quite diverse – including participants from 24 countries with a good gender and age distribution – they were still all students and thus somewhat nonrepresentative of the general population, who engages in negotiation to address important and less important issues in everyday life. It is a long-standing debate whether the findings of studies using student samples can be generalized to the general population. The camp that questions such generalizability states that there are good reasons for doubt: most students have little professional experience, particularly in high-stakes negotiation, and their performance in class is likely not indicative of the performance of trained professionals. The opponents state that negotiation processes are highly prevalent and that everyone negotiates something every day, and that cognitive processes in students ought to mirror the cognitive processes in professionals, and hence the effects should generalize. The only empirical test of this question seems to support the conclusion that student-based studies provide valid inferences. Herbst and Schwarz (2011) investigated whether the results obtained from trained student samples are generally similar to those of professional negotiators, and found that students with some negotiation training and experience perform better than untrained student negotiators and that they are not significantly outperformed by professional negotiators. They conclude that many research questions can be validly tested using students. However, there is little doubt that our models would benefit from being replicated in samples consisting of professionals. This is a matter for future studies.

What lies ahead?

The findings have considerable practical importance because of the pervasiveness of negotiation in human affairs and the vast amount of value at stake, coupled with the evidence of value destruction and the risks inherent in competitive negotiation.

SARS-CoV-2 has damaged economies and livelihoods and widened existing inequalities between nations and individuals, particularly the gaps based on gender, race and income. The global economy is at the lowest level it has ever been in peacetime. The geopolitical challenges of the tensions between the East and the West, the divided America and the post-Brexit United Kingdom, the rising neo-statism (the G-zero mentality) and nationalism, cyber risks, and the ever more urgent climate change, suggest that we will have to negotiate many aspects of our lives. We will invariably find ourselves in situations involving conflicting, compatible and differing interests, and face a choice on how to address them.

Our research suggests that we should tread carefully, as our default approach to negotiation is likely to result in contending and outcomes that we might not be able to afford. It also shows that we are capable of doing well but are limited by the competitive mindset evoked by the ‘negotiation’ frame, the inability, or unwillingness, to consider the minds of others, and the reluctance to exert effort when thinking. Our research provides the negotiation field, among other things, tools for naming, confronting, and perhaps changing these phenomena.

APPENDICES

Glossary of negotiation terms

Bargaining range, also **zone of possible agreement**. The range of possible agreements that would be of economic benefit to both parties. For example, if in a zero-sum price negotiation the seller is willing to accept 80 and the maximum the prospective buyer is willing to pay is 100, the bargaining range is 20, between 80 and 100.

BATNA, also **best alternative to a negotiated agreement**, also **no-deal option**. The course of action available to a negotiator if the agreement under negotiation falls through. A walkaway option. For example, for a negotiator negotiating a purchase of a car, the alternative could be to rent a similar car.

Best alternative to a negotiated agreement. See **BATNA**.

Compatible issues. Issues in negotiation where the parties' preferences are perfectly aligned (e.g., both me and my counterparty desire to settle this lawsuit rather than litigate and wish this settlement to be finalized before the disclosure process is triggered because it entails a major legal expense). There is no conflict at all. The compatible option is the best outcome from the parties' joint and individual perspectives

Distributive issues, also **zero-sum issues.** Issues where the parties value the resources equally, so that the gain of a concession to one side is identical to the loss to the other.

Individual gain. The gain an individual negotiator gets in a negotiation. A direct function of joint gain and reservation values. For example, if a Seller sells a car she values $800 for $850 to a buyer who values it a $1,000, , the Seller's individual gain is $50 and the buyer's $150.

Information asymmetry. The difference in information the negotiators possess. For example, the buyer does not know the minimum price the seller is willing to accept.

Integrative issues. Issues in negotiation where the parties have different preferences from each other, which allows trading concessions on issues one cares about less for concessions one cares about more.

Integrative negotiation. Negotiations that carry the potential to create value, available when the benefit of a concession on an issue for one party does not equal the loss to the other. The first party's gain can correspond to either a non-equal loss, in which case the issue is *integrative*, or it can correspond to a gain, in which case the issue is *compatible*.

Interests. In negotiation theory, a party's basic needs, wants and motivations. In our theory, negotiator's motivating mental states. What is the negotiator trying to achieve by negotiating. In the library window example, the desire for fresh air and the desire to avoid draft are interests.

Joint gain. The total gain the parties generate in a negotiation. A function of reservation values of the negotiating parties. For example, if a Seller sells a car she values $800 for $850 to a buyer who values it a $1,000, the joint gain is $200. In non-zero-sum negotiation, a sum of gains from compatible and integrative issues.

No-deal option. See **BATNA**.

Options. A full range of things negotiators might possibly agree as part of the negotiated agreement (e.g., terms, conditions, procedures, contingencies, even deliberate omissions).

Positions. Possible ways of satisfying interests. There are many possible positions for each interest. In the library window example, the extent to which the window is open.

Price negotiation. See **distributive negotiation**.

Reservation value. Quantified best alternative to a negotiated agreement. For example, the minimum (maximum) the seller (buyer) is willing to accept (pay).

Value claiming. Part of the negotiation process, or negotiating parties' strategies, aimed at maximizing the individual gain of each negotiator.

Value creating. Part of the negotiation process, or negotiating parties' strategies aimed at maximizing the joint value that is available to both parties.

Zero-sum issues. See **distributive issues**.

Zero-sum negotiation, also **price negotiation** or **'zero-sum'** negotiation. A negotiation where the parties haggle over a resource they value *equally*. Equal valuations imply that a gain for one party corresponds to a loss of exactly the same value for the other. An archetype is a bazaar bargaining over the price of a rug. A dollar more for the seller is an exact same dollar less for the buyer; if I pay $900 rather than $1,000, I save the same $100 that the seller does not earn.

Zero-sum negotiation. See **distributive negotiation**.

Zone of possible agreement. See **bargaining range**.

Example of gains in multi-issue negotiation

Below we outline, for illustration purposes, the different types of gains in a (stylized and fully quantified) multi-issue negotiation.

Background

A seller and a prospective buyer are negotiating a sale of a new car. There are four issues to negotiate: price, warranty, equipment and color. The parties' preferences have been quantified in the payoff schedule below (column Value to seller/buyer). For example, the value the seller gets from a price of £78 thousand is 4,500 points whereas the buyer's value is lower at 1,500 points (she is paying a relatively high price).

Price

Option	Value to Seller	Value to Buyer
£80,000	6000	0
£78,000	4500	1500
£76,000	3000	3000
£74,000	1500	4500
£72,000	0	6000

Equipment and features

Option	Value to Seller	Value to Buyer
Maximum	4000	0
Most	3000	400
Moderate	2000	800
Minimum	1000	1200
None	0	1600

Warranty

Option	Value to Seller	Value to Buyer
6 months	1600	0
12 months	1200	1000
18 months	800	2000
24 months	400	3000
30 months	0	4000

Colour

Option	Value to Seller	Value to Buyer
Blue	2400	2400
Green	1800	1800
Black	1200	1200
Red	600	600
Yellow	0	0

Issues

Price is a **distributive (zero-sum) issue**. The gain each concession (change of option) brings to one party is identical to the loss it causes the other (e.g., if the price is £80 thousand rather than £78 thousand, the seller benefits 1,500 points and the buyer loses 1,500 points).

Warranty and Equipment and features are **integrative issues**. The seller cares more about equipment and features (e.g., an increase from ‘most’ to ‘maximum’ brings the seller 1000 but only costs the buyer 400). The buyer cares more about the warranty (e.g., an increase from 24 to 30 months benefits the buyer 1,000, but only costs the seller 400).

Color is **a compatible issue**. Both parties prefer the car to be blue. Any other option loses an equal amount of value for both and is an instance of a lose-lose agreement.

Dyadic gains

The parties’ **integrative gain** is the sum of the gain of both parties in integrative issues. If they compromise and ‘split the difference’ (i.e., they choose the midpoint option for both integrative issues), their integrative gain is 5600 (‘moderate’ equipment and features 2,800 + 2,800, and 18 months warranty 800 + 2,000). If they maximize integrative gain by logrolling (i.e., trading high value for low value issues) their integrative gain is 8,000 (‘maximum’ equipment and features 4,000 + 0, and 30 months warranty 0 + 4,000).

The parties **compatible gain** is the value derived from the compatible issue. If they choose the car to be blue, their compatible gain is 4,800 (2,400 + 2,400), which is also the maximum gain possible in this issue. Any other option is suboptimal (e.g., a green car results in compatible gain of 3,600, 800 points less than the preferred blue).

In **distributive** issues, the parties dyadic gain is a constant; irrespective of which option, the joint gain is 6,000. Distributive issues offer no opportunity to create value.

The parties **joint gain** is the sum of integrative, compatible and distributive gain.

Individual gain

Individual gain is the gain of each negotiator and is the sum of the value to the respective party on all issues. The sum of individual gains of the parties always equals joint gain. If the parties go for a price of £78 thousand, most equipment and features, 24 months warranty and a blue car, the seller's individual gain is 10,300 and the buyer's 7,300. The calculations are provided below (the parties agreed options are in bold).

Price

Option	Value to Seller	Value to Buyer
£80,000	6000	0
£78,000	**4500**	**1500**
£76,000	3000	3000
£74,000	1500	4500
£72,000	0	6000

Equipment and features

Option	Value to Seller	Value to Buyer
Maximum	4000	0
Most	**3000**	**400**
Moderate	2000	800
Minimum	1000	1200
None	0	1600

Warranty

Option	Value to Seller	Value to Buyer
6 months	1600	0
12 months	1200	1000
18 months	800	2000
24 months	**400**	**3000**
30 months	0	4000

Colour

Option	Value to Seller	Value to Buyer
Blue	**2400**	**2400**
Green	1800	1800
Black	1200	1200
Red	600	600
Yellow	0	0

Individual gain	
Seller	**10300**
Buyer	**7300**
Dyadic gains	
Joint gain	**17600**
Integrative gain	**6800**
Compatible gain	**4800**
Constant	**6000**

Questionnaires

Reflective Functioning Questionnaire

Please work through the next 8 statements, each time circling the one response that you feel describes you most clearly. Do not think too much about it - your initial responses are usually the best. Thank you.

Use the following scale:

Strongly	1	2	3	4	5	6	7	Strongly
Disagree								agree

1. __People’s thoughts are a mystery to me
2. __I don’t always know why I do what I do
3. __When I get angry I say things without really knowing why I am saying them
4. __When I get angry I say things that I later regret
5. __If I feel insecure I can behave in ways that put others’ backs up
6. __Sometimes I do things without really knowing why
7. __I always know what I feel
8. __Strong feelings often cloud my thinking

Cognitive reflection test

1. A bat and a ball cost $1.10 in total. The bat costs a dollar more than the ball. How much does the ball cost. ____ cents [Correct answer = 5 cents; intuitive answer = 10 cents]
2. If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets? ____ minutes [Correct answer = 5 minutes; intuitive answer = 100 minutes]
3. In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake? ____ days [Correct answer = 47 days; intuitive answer = 24 days]

Study 1

Impact of training in the reduced sample (N = 242)

As predicted, training increased all aspects of dyadic gain; one-way MANOVA showed a significant effect on both adjusted compatible and integrative gain; $F(2, 120) = 11.36$, $p < .001$, $\Lambda = .84$. Contrast analysis is below.

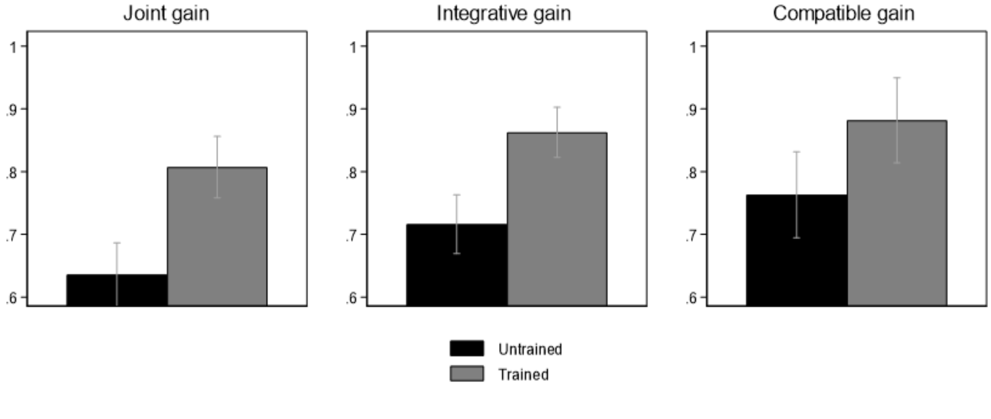
Contrast analysis of joint, integrative and compatible gain between trained and untrained samples.

	Joint gain		Integrative gain		Compatible gain	
	M	*SD*	*M*	*SD*	*M*	*SD*
Untrained	8,560	2,698	10,898	1,556	1,263	1,540
Trained	10,654	1,987	12,422	1,771	1,832	999
difference	2,094		1,524		570	
t	4.79		4.53		2.44	
p	<.001		<.001		.008	
d	.835		.939		.407	

Note. The test statistic is Welch-adjusted t-test (single-tailed). Effect size 95% confidence intervals are in square brackets. Results remain significant if adjusted for multiple comparisons.

Trained negotiators outperformed their untrained peers by 27% in (unadjusted) joint gain, resulting from a 20% improvement in (unadjusted) integrative gain and a 28% increase in compatible gain. Contrasts showing the improvements facilitated by training as percentage of gains are below.

Unadjusted joint gain, integrative gain and compatible gain as percentage of optimal outcome in untrained and trained groups



Comparison of models regressing negotiator's gain on RFQc and RFQc advantage

	1	2	3	4	5	6
RFQc Advantage	303.61	88.06				
	(206.00)	(220.44)				
RFQc Adv x training		583.19	671.25*			
		(432.79)	(372.45)			
RFQc				129.29	42.93	
				(227.21)	(222.39)	
RFQc x training					934.56***	943.99***
					(170.921)	(164.18)
Constant	4534.71***	4534.71***	4534.71***	4367.92***	4163.53***	4215.59***
	(137.77)	(137.77)	(137.77)	(313.29)	(302.51)	(159.54)
Wald	2.17	3.41	3.25	0.32	32.99	33.06

Note. *** p<.001, ** p<.01, * p<.05
Standard errors are in parentheses.

Part 2: Theory

Active inference framework

Perceptual and cortical hierarchy

In building an internal model of the world, the brain must represent regularities and objects (e.g., causal connections) on varying time-scales (e.g., movement of a ball versus changes of seasons) and of varying levels of detail (e.g., a fine-grained representation of mother's smiling face versus a more enduring representation of her face per se, versus a more abstract idea of a 'mother' or a 'face' or a 'smile'). These two are connected in that there is normally a tradeoff between the time scale and the level of detail: faster regularities are more fine-grained than slower ones.

This is important because the representations of such regularities are organized, from slower to faster, in the cortical hierarchy in the brain:

"[F]ast regularities are processed early in the sensory processing stream (for visual perception, this happens in area V1 at the back of the brain) and then increasing time scales are processed as the sensory signal works its way up through the primary sensory areas and into higher areas.

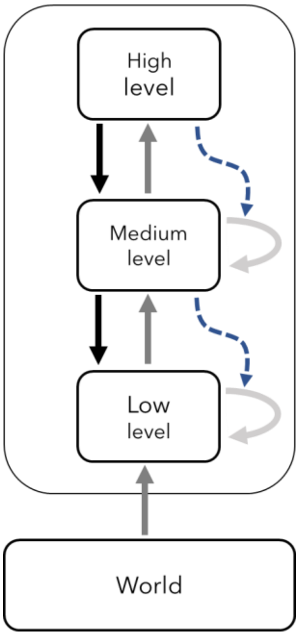
The hierarchy also has a spatial dimension, which sits naturally with the temporal focus we have had so far. The fast time scale regularities represented in low levels of the hierarchy (such as in V1) have small, detail-focused receptive fields of only a couple degrees whereas later areas of processing have wider receptive fields (e.g., 20–50 degrees in the temporal cortex). Receptive fields are also characterized by interconnections, such that wide receptive fields take in sets of smaller receptive fields processed lower down in the hierarchy. Perceptual inference happens in this highly interconnected, cortical hierarchy and can as such avail itself directly of its representation of myriad causal relations in its attempt to get the world right, in its construction of a first person perspective, and in its ability to orient itself for action in the world." (Hohwy, 2014, p. 28).

Slower regularities represented at higher cortical levels correspond to the relatively invariant aspects of perception. They are precise about the abstract parameters, but ignorant of fine-grained detail. Conversely, the faster regularities, represented at lower level units, correspond to the relatively variant aspects of perception; these units are relatively 'myopic': they are good at fine-grained detail but blind to context.

Prediction error and expected precisions

The traditional idea of perception is that human senses provide a rich stream of data that the brain makes sense of in a relatively passive way (Hohwy, 2014, p. 47). In these accounts, perception is predominantly driven bottom-up. Any top-down cognitive modulation is a mere feedback on the sensory signal that is coming the other way.

The active inference framework inverts this model. Perception starts with a set of expectations provided in a top-down fashion by higher cortical structures in the form of predictions about the expected sensory input (e.g., this is a cow). These predictions sample the world. If the sensory feedback fits what is expected, we are confidently perceiving in line with our predictions (it looks like a cow, moos like a cow, it is a cow). A signal that does not match the expectation (it walks on hind legs and growls threateningly) on the other hand results in a *prediction error*, which the brain addresses by either more intense or detailed sampling (by active inference, e.g., by looking more closely, changing a viewpoint to a different perspective, touching, smelling) or by changing the predictive hypothesis (e.g., it is not a cow, is it a bear?). In this model, the sensory signal is the feedback on the predictions provided by the brain, and the only thing propagated up and down the cortical hierarchy is the prediction error that assists in revising the model parameters (see below).

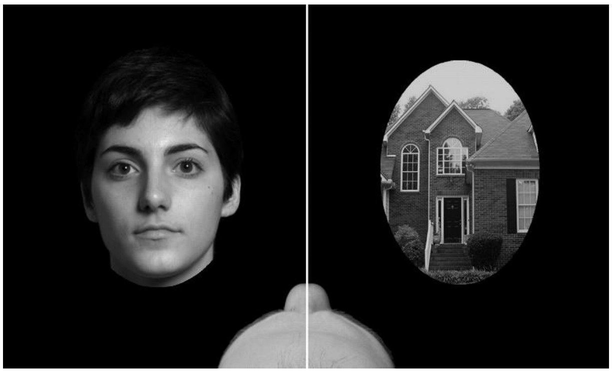


Perceptual hierarchy, prediction error and expected precisions. Higher cortical units send prior hypothesis-driven queries (black arrows) to lower units. The lower units answer by the prediction error signal (grey arrows). The higher-level units also perform contextual modulation by dictating expectations about prediction error precisions (dashed arrows), based on which the lower levels self-inhibit the prediction error signal (light grey curved arrows). The figure is a based on Hohwy (2014, p. 68).

Because prediction error drives the reality-model parameterization (active inference, further perceptual inference, and the revision or change of the hypothesis), our whole perception is hostage to the quality of the prediction error we can generate. Not all error signals are of the same quality. In some contexts, prediction error can be trusted more (e.g., encountering the large animal during a clear bright day in the Austrian alps) than in others (e.g., encountering it during heavy rain at dusk in the Austrian alps). To account for variance in the quality of prediction error, the brain must perform a second type of inference, this time about the expected precisions of the prediction errors provided by the lower hierarchical units. These expected precisions control the impact (the 'gain') of the reported prediction error and its ability to drive action or model change. For example, if the prediction error return on a higher unit's hypothesis-driven query is deemed to have high precision, the hypothesis might be updated (or active inference engaged) so that the error is explained away; if on the other hand the prediction error is estimated to be imprecise, its signal will be brusquely suppressed (i.e., not explained away but deemed less informative and dampened down).

Binocular rivalry

Binocular rivalry, a visual phenomenon that has captured the researchers attention for over 400 years (Alais & Blake, 2005), is example of Bayesian perceptual inference. Two different images, for example a face and a house, are presented one to each eye.

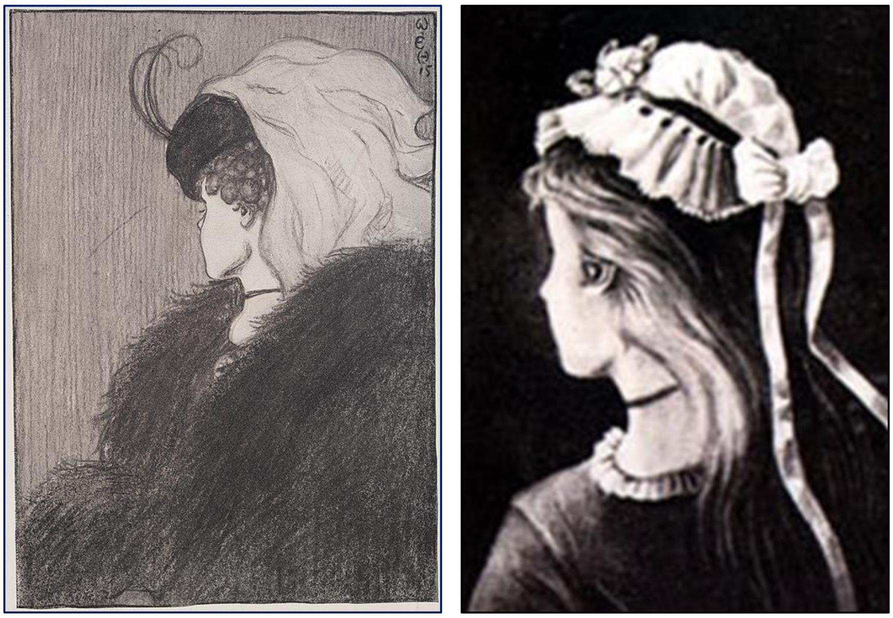


Binocular rivalry. Materials used to elicit binocular rivalry. The two images are presented to one eye each. Image from Anderson, Siegel, Bliss-Moreau, and Barrett (2011).

The person looking through a stereoscope sees, somewhat counterintuitively, not a mishmash of a face and a house, but *either* a house or a face, or one after another, each followed by a short period of fuzziness. The effect remains the same if each image is split in half and two composite images are presented to each eye (Diaz-Caneja, 1928), for example the left eye gets an image showing a half-house and a half-face, and the right eye a half-face and a half-house. Again, what is 'seen' is either a house or a face, not a 'houseface'. From the perspective of active inference, the hypothesis of a 'houseface' composite has a high likelihood (i.e., given it is indeed the underlying cause, it explains the sensorium really well), but a miniscule prior. Instead, the brain selects a 'revisionary' hypothesis of either a 'house' or a 'face' (Hohwy et al., 2008), each of which suppresses a significant part of the incoming sensorium (the 'house' hypothesis, for example, quells the incoming data related to the image of the face, and vice versa).

Example of cognitive penetrability of priors

An evocative example of the cognitive penetrability of prior expectations this are visual illusions such as *My wife and my mother in law*.

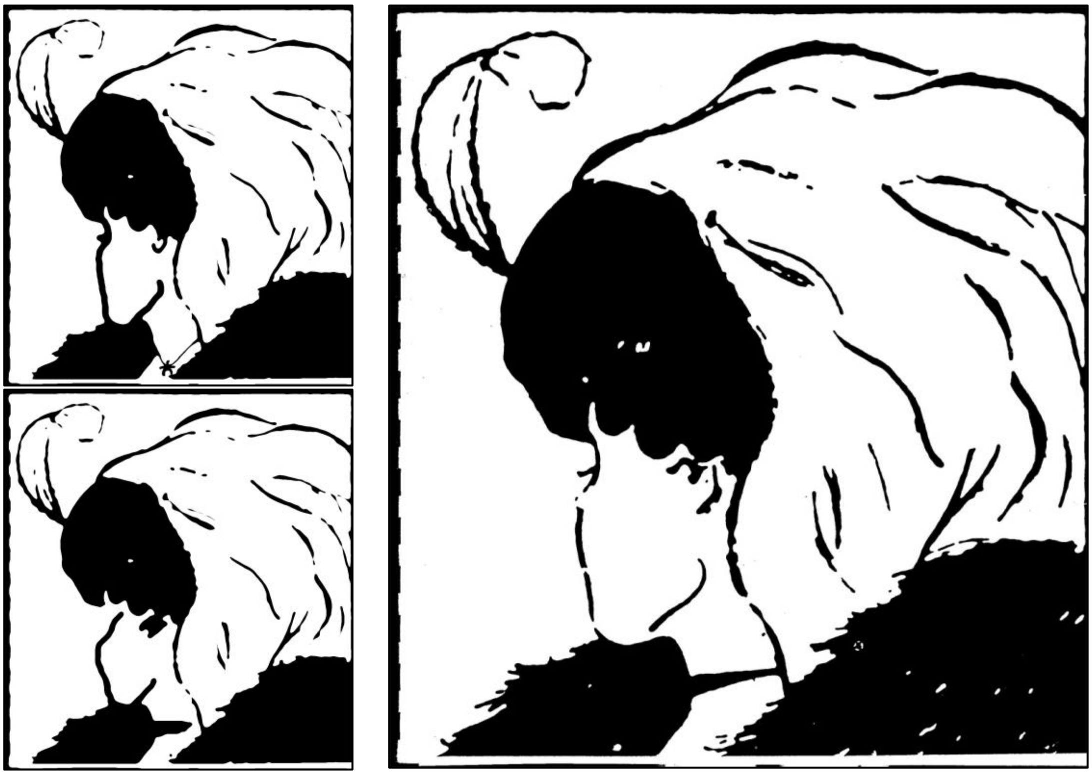


My wife and my mother in law. The original drawing by William Hill (on the left) that appeared in 1915 in an American humor magazine (Hill, 1915), likely based on a caricature from an older German postcard (on the right).

Perhaps the best example of the impact of priors is from the use of a version of this image in education.[25] As part of a lecture on partisan perceptions in disputes I hand out to half of the students printed cards with a young woman and the other half with pictures of the older woman. I then ask them to estimate the age of the woman in the picture and write it down (for anchoring purposes). I then collect the cards, show the ambiguous picture on a large screen, and tell the students that while this is not exactly the same picture as the one they just had seen, they need to tell me what they think the age is. This invariably leads to some participants reporting extreme opinions (high and

[25] I am indebted to Bruce Patton from the Harvard Negotiation Project who first showed me the use of this image to powerfully demonstrate how easily primed partisanship is and how it can cripple people's ability to reach an agreement although they are looking at identical 'facts' (the ambiguous image) and, more importantly, the critical importance of perspective taking in any attempt at dispute resolution.

low ages in the region of 15 and 95 are not uncommon). I pick a few of the most polarized participants to discuss the woman's age live in front of the class and try to come to an agreement about her age. What usually follows is an interaction along the lines of: 'she is ancient, look at her nose', 'what nose, there is no nose', 'what do you mean there is no nose, look at it, it is massive', and so on, and sometimes regressing into a heated argument where the participants employ personality attributions ('you are crazy'). Finally, there are invariably people in class who, even after being told there are two women in the picture, are not able to 'see' the woman they have not been primed with. This follows a path, predicted by the naïve realism theory, from miscommunication ('where is her nose?') to social hostility and erroneous personality attributions ('*you* are crazy'). See Study 3 in this thesis for a discussion and empirical test of the impact of mentalizing on revising naïve realistic perceptions.



Old lady, young lady partisan perceptions task. The top left picture is a version where the features of a young woman are emphasized. The bottom left picture emphasizes the features of an older woman. The picture on the right is the ambiguous version.

The *Old lady, young lady* example is particularly interesting because it simultaneously provides evidence for cognitive penetrability and impenetrability. Take an example of a participant who is heavily primed by an old woman, so much that even once told (by people into whom he presumably put his epistemic trust) that the picture is ambiguous in that it consists of images of both an old and a young woman, he is unable to see anything but the old one. His priors (largely a consequence of priming with the initial card), penetrate his perception and determine the reality of the otherwise ambiguous image.

The flip side is the cognitive impenetrability of the (later induced) prior that there is a young woman there, too: although our participant knows, at some level, that the picture also contains a young woman (based on the information provided by the instructor), he cannot perceive it. Our strongly primed participant thus possesses two sets of prior expectations: the earlier prior that generates the predictions of visual shapes that are readily met with the feedback signal resulting in a marginal prediction error (there is some because the second image is not exactly the same as the initial picture), and the later prior that generates predictions that remain unmet with the sensorium. The reason is likely that the later, higher-level belief (that there is *also* a young woman in the picture) is, when competing with an already primed old woman hypothesis, far too abstract to have effect. In Hohwy's words, it does not 'predict at the right fineness of spatiotemporal grain' and therefore 'cannot make predictive contact with the sensory input', and is, for that reason, 'probabilistically idle' (2014, p. 127). In other words, the participant knows at some abstract level that there is a young woman in the picture too, but cannot get this 'rather coarse (invariant) true prior' to make proper contact with the 'fine-grained (variant)' set of pixels in the image (Hohwy, 2012; 2014, p. 127; Hohwy & Rosenberg, 2005). Also, it is likely that when a particular hypothesis emerges as having a high probability, the other units at the same level of hierarchy are progressively inhibited from influencing inference. In uncertain situations, a number of hypotheses may concurrently attempt to explain the sensorium, however once one emerges as a good fit, the activity of the others tends to dissipate (or rather, they are 'explained away'; Hohwy, 2014, p. 61). From that we can cautiously draw a lesson which will be relevant in the negotiation context: new abstract knowledge may fail to influence perception if a competing prior hypothesis is already active and has good traction with the sensory input. A prior hypothesis that generates negligible prediction error may prevent a competing, more abstract hypothesis from being tested at all.

Study 4

Trust task

The following are the written instructions received by participants in the trust task:

Please spend the next 5 minutes talking with your counterparty. Below you have three questions that you should ask and answer. You should take turns in asking and answering the questions.

Your task is to **understand** the counterparty and to **make sure the counterparty knows you understand** them:

When your counterparty is talking, **listen carefully**, making sure you understand what they tell you. Try to **see and feel things from their perspective**. It is important that you understand them as well as you possibly can. You may need to ask questions of clarification to make sure you have understood what they told you.

Equally importantly, please make sure **your counterparty can see that you have listened to and understood** what they told you. You may wish to rephrase, nod, check if you understood (whatever feels appropriate).

When it is your turn to give your answers, **please be brief, frank and straight forward**. Do not make up stories (but also do not feel you need to give details about yourself that make you feel uncomfortable).

Remember, your task is to **UNDERSTAND** your counterparty and to **SHOW** them that you understand them.

Questions:

1. What is your name?
2. What does it mean for you to come to [this university] to do an LLM?
3. If you could have dinner with any person, living or dead, who would you choose (and why)?

If you have time:

1. What job would you be bad at (and why)?
2. What was the best piece of advice you can remember being given?

In addition, the instructors read out loud and showed on screen the following examples of active listening:

Example one

Speaker: I got that job!

Listener: Congratulations, what an achievement! How does it feel?

Speaker: It’s quite a relief, actually.

Listener: Was it very stressful?

Speaker: Yes, and now I finally don’t have to worry about my student loans

Listener: (nods) I get that… one thing less to worry about.

Example two

Speaker: I wish I could call my brother right now.

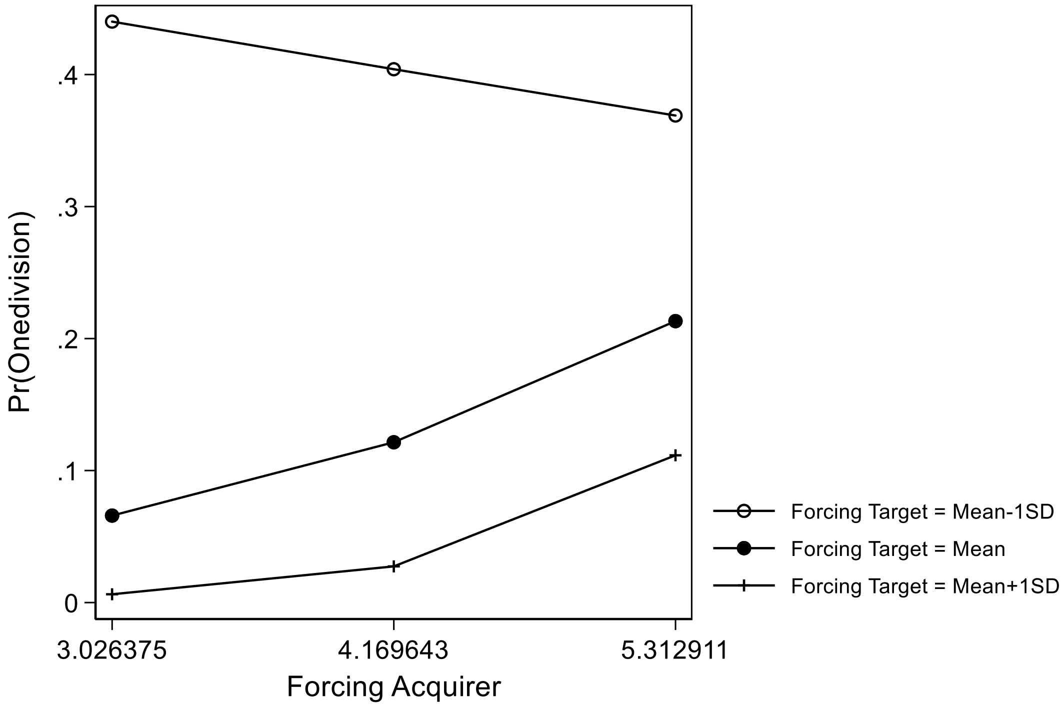
Listener: I see, there are things you would want to talk to him about.

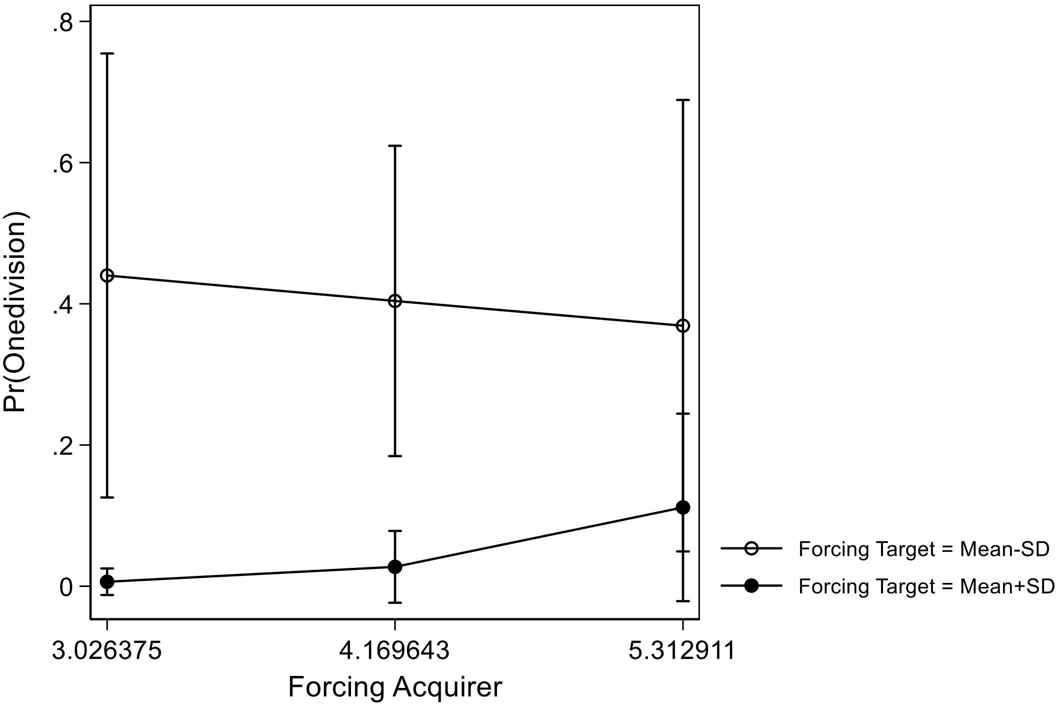
Speaker: Yes… and I miss him.

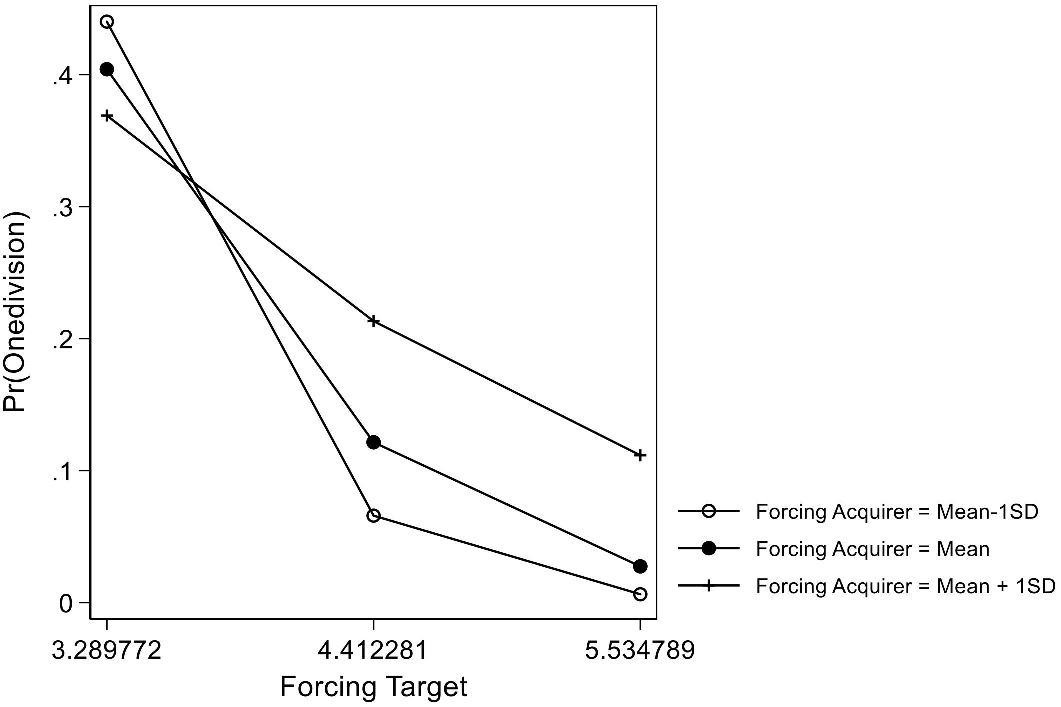
Listener: (nods) yes, I know what you mean. I miss my family too.

Study 6

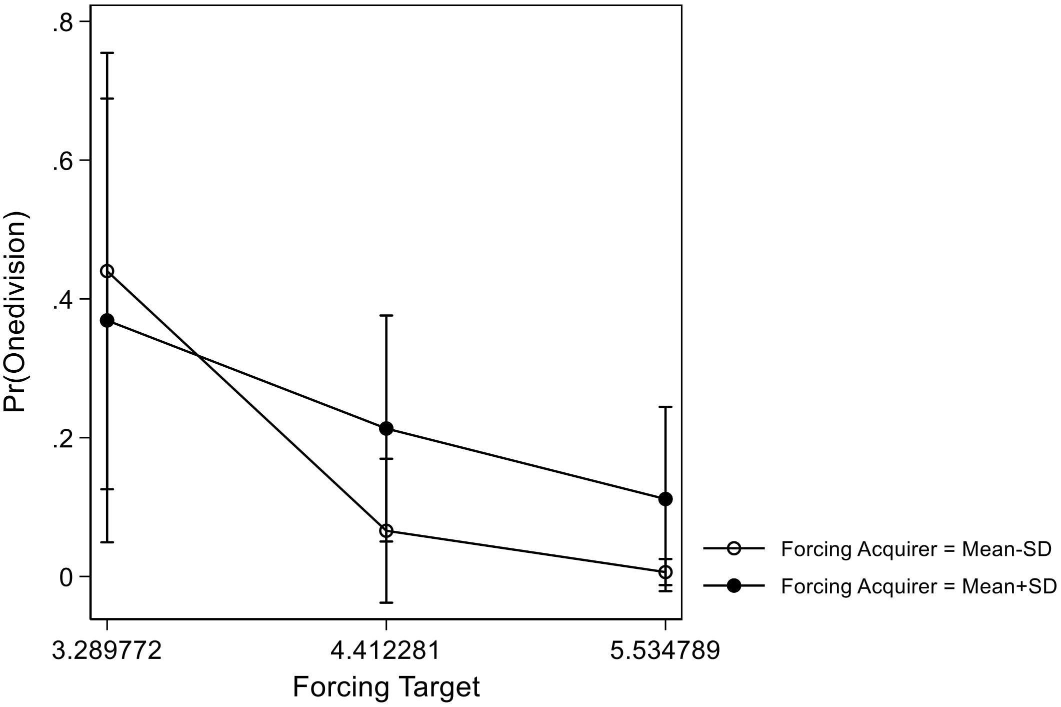
Interaction effects between contending of Target and Acquirer




.8
.6
.4
.2
0
Pr(Onedivision)
3.026375
4.169643
5.312911
Forcing Acquirer
Forcing Target = Mean-SD
Forcing Target = Mean+SD



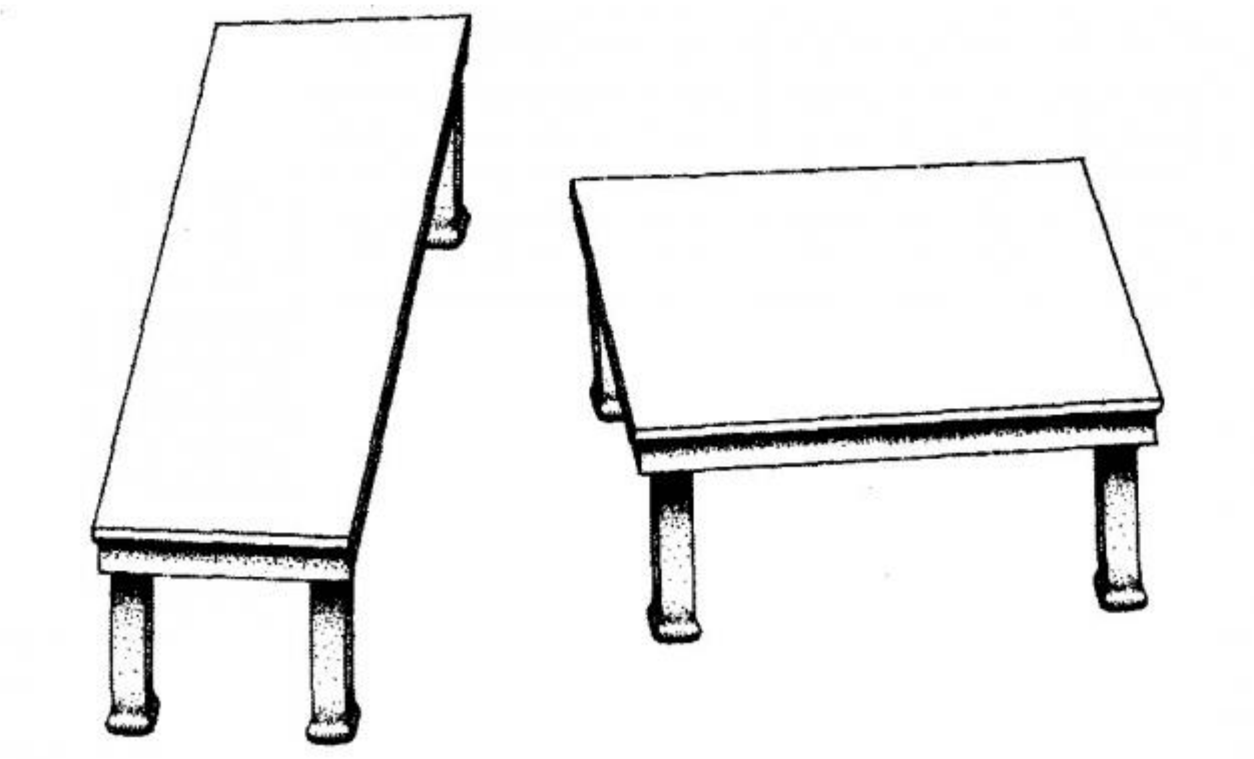
Pr(Onedivision)
.4
.3
.2
.1
0
3.289772
4.412281
5.534789
Forcing Target
Forcing Acquirer = Mean-1SD
Forcing Acquirer = Mean
Forcing Acquirer = Mean + 1SD



Pr(Onedivision)
.8
.6
.4
.2
0
3.289772
4.412281
5.534789
Forcing Target
Forcing Acquirer = Mean-SD
Forcing Acquirer = Mean+SD

Conclusion

Shepard’s tables



© RN(1990) Mind Sights: Original Visual Illusions, Ambiguities, and other Anomalies, NY: WH Freeman

REFERENCES

Ade, V., Schuster, C., Harinck, F., & Trötschel, R. (2018). Mindset-Oriented Negotiation Training (MONT): Teaching more than skills and knowledge. *Frontiers in Psychology, 9*, 907.

Aiello, L. C., & Wheeler, P. (1995). The expensive-tissue hypothesis: the brain and the digestive system in human and primate evolution. *Current anthropology, 36*(2), 199-221.

Akerlof, G. A. (1970). The Market for "Lemons": Quality Uncertainty and the Market Mechanism. *The Quarterly Journal of Economics, 84*(3), 488-500. doi:10.2307/1879431

Akerlof, G. A. (1982). *The market for "lemons": quality uncertainty and the market mechanism*: Cambridge University Press.

Alais, D., & Blake, R. (2005). *Binocular Rivalry*. Boston: MIT press.

Alexander, R. D. (1989). Evolution of the human psyche. In P. Mellars & C. Stringer (Eds.), *The human revolution: Behavioural and biological perspectives on the origins of modern humans* (pp. 455-513). Princeton, NJ: Princeton University Press.

Alexander, R. D., Mellars, P., & Stringer, C. (1989). Evolution of the human psyche. In P. Mellars & C. Stringer (Eds.), *The human revolution*. UK: University of Edinburgh Press.

Allen, J. (2006). Mentalizing in practice. In A. Bateman & P. Fonagy (Eds.), *Handbook of mentalization-based treatment* (pp. 3-30). Chichester, England; Hoboken, NJ: John Wiley & Sons Ltd.

Allen, J. (Ed.) (2013). *Mentalizing in the development and treatment of attachment trauma*: Karnac Books Ltd.

Allen, J., & Fonagy, P. (2008). *Handbook of mentalization-based treatment.* London: Wiley & Sons.

Allen, M., & Friston, K. J. (2018). From cognitivism to autopoiesis: towards a computational framework for the embodied mind. *Synthese, 195*(6), 2459-2482.

Allred, K. G., Mallozzi, J. S., Matsui, F., & Raia, C. P. (1997). The influence of anger and compassion on negotiation performance. *Organizational Behavior and Human Decision Processes, 70*(3), 175-187. doi:10.1006/obhd.1997.2705

Anderson, E., Siegel, E. H., Bliss-Moreau, E., & Barrett, L. F. (2011). The visual impact of gossip. *Science, 332*(6036), 1446-1448.

Austin, P. C., & Steyerberg, E. W. (2015). The number of subjects per variable required in linear regression analyses. *Journal of Clinical Epidemiology, 68*(6), 627-636.

Axelrod, R., & Dion, D. (1988). The further evolution of cooperation. *Science (New York, N.Y.), 242*(4884), 1385.

Axelrod, R., & Hamilton, W. D. (1981). The evolution of cooperation. *Science (New York, N.Y.), 211*(4489), 1390.

Axelrod, S., & May, J. G. (1968). Effect of increased reward on the two-person non-zero-sum game. *Psychological Reports, 23*(2), 675-678.

Babcock, L., & Loewenstein, G. (1997). Explaining bargaining impasse: The role of self-serving biases. *Journal of Economic Perspectives, 11*(1), 109-126.

Badoud, D., Luyten, P., Fonseca-Pedrero, E., Eliez, S., Fonagy, P., & Debbané, M. (2015). The French version of the Reflective Functioning Questionnaire: validity data for adolescents and adults and its association with non-suicidal self-injury. *PloS one, 10*(12), e0145892.

Baker, W. E. (1984). The social structure of a national securities market. *American Journal of Sociology, 89*(4), 775-811.

Bandura, A., & Locke, E. A. J. J. o. a. p. (2003). Negative self-efficacy and goal effects revisited. *88*(1), 87.

Bargh, J. A. (2013). Our Unconscious Mind. *Scientific American, 310*(1), 30. doi:10.1038/scientificamerican0114-30

Bargh, J. A., Bond, R. N., Lombardi, W. J., & Tota, M. E. (1986). The Additive Nature of Chronic and Temporary Sources of Construct Accessibility. *Journal of Personality and Social Psychology, 50*(5), 869-878. doi:10.1037/0022-3514.50.5.869

Bargh, J. A., & Chartrand, T. L. (1999). The unbearable automaticity of being. *American Psychologist, 54*(7), 462-479. doi:10.1037/0003-066X.54.7.462

Bargh, J. A., & Morsella, E. (2008). The unconscious mind. *Perspectives on Psychological Science, 3*(1), 73-79.

Bargh, J. A., Schwader, K. L., Hailey, S. E., Dyer, R. L., & Boothby, E. J. (2012). Automaticity in social-cognitive processes. *Trends in Cognitive Sciences, 16*(12), 593-605. doi:10.1016/j.tics.2012.10.002

Bargh, J. A., & Thein, R. D. (1985). Individual Construct Accessibility, Person Memory, and the Recall-Judgment Link: The Case of Information Overload. *Journal of Personality and Social Psychology, 49*(5), 1129-1146. doi:10.1037/0022-3514.49.5.1129

Baron-Cohen, S. (1985). *Does the autistic child have a 'theory of mind'.*

Baron-Cohen, S. (1994). The Mindreading System: New directions for research.

Baron-Cohen, S. (1997). *Mindblindness: an essay on autism and theory of mind* (1 ed.). Cambridge, Mass.: MIT Press.

Baron-Cohen, S., Golan, O., Chakrabarti, B., & Belmonte, M. K. (2008). *Social cognition and autism spectrum conditions*: Oxford University Press.

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The "Reading the Mind in the Eyes" test revised version: A study with normal adults, and adults with Asperger syndrome or high-functioning autism. *Journal Of Child Psychology And Psychiatry And Allied Disciplines, 42*(2), 241-251.

Bartos, O. J. (1972). Foundations for a rational-empirical model of negotiation. In J. Berger, M. Zelditch, & B. Anderson (Eds.), *Sociological theories in progress* (Vol. 2, pp. 3-20). Boston: Houghton Mifflin.

Bateman, A., & Fonagy, P. (2012). *Handbook of mentalizing in mental health practice*: Washington, DC; London: American Psychiatric Publishing.

Bateman, A., & Fonagy, P. (2016). *Mentalization-based treatment for personality disorders: A practical guide*: Oxford University Press.

Bathke, A. C., & Harrar, S. W. (2016). Rank-based inference for multivariate data in factorial designs. In *Robust rank-based and nonparametric methods* (pp. 121-139). New York: Springer.

Batson, C. D., & Ahmad, N. (2001). Empathy-induced altruism in a prisoner's dilemma II: what if the target of empathy has defected? *European Journal of Social Psychology, 31*(1), 25-36. doi:10.1002/ejsp.26

Batson, C. D., & Moran, T. (1999). Empathy-induced altruism in a prisoner's dilemma. *European Journal of Social Psychology, 29*(7), 909-924. doi:10.1002/(SICI)1099-0992(199911)29:7&lt;909::AID-EJSP965&gt;3.0.CO;2-L

Battle, E. S. (1965). Motivational determinants of academic task persistence. *Journal of Personality and Social Psychology, 2*(2), 209.

Bazerman, M. H. (2013). *Judgment in managerial decision making* (8 ed.). Hoboken, N.J.: Hoboken, N.J.: Wiley.

Bazerman, M. H., & Chugh, D. (2006). Decisions without blinders. *Harvard business review, 84*(1), 88.

Bazerman, M. H., Curhan, J. R., Moore, D. A., & Valley, K. L. (2000). Negotiation. *Annual review of psychology, 51*(1), 279-314. doi:10.1146/annurev.psych.51.1.279

Bazerman, M. H., Magliozzi, T., & Neale, M. A. (1985). The acquisition of an integrative response in a competitive market. *Organizational Behavior and Human Decision Processes, 35*(3), 294-313.

Bazerman, M. H., & Neale, M. A. (1986). Heuristics in negotiation: Limitations to effective dispute resolution. In H. Arkes & K. Hammond (Eds.), *Judgment and decision making: An interdisciplinary reader* (pp. 311-321). New York: Cambridge University Press.

Bazerman, M. H., & Neale, M. A. (1991). *Cognition and rationality in negotiation*. New York: Free Press.

Bazerman, M. H., & Neale, M. A. (1992). *Negotiating rationally*: New York; London: Free Press.

Beersma, B., & De Dreu, C. K. W. (2002). Integrative and Distributive Negotiation in Small Groups: Effects of Task Structure, Decision Rule, and Social Motive. *Organizational Behavior and Human Decision Processes, 87*(2), 227-252. doi:10.1006/obhd.2001.2964

Benjamin, A., & Bjork, R. (1996). Implicit memory and metacognition.

Bernath, A., & Kovacs, J. (2015). The need for mentalizing and cooperation in integrative negotiation/Mentalizaciora iranyulo igeny es egyuttmukodes integrativ alkutargyalasban. *Magyar Pszichológiai Szemle, 70*(1), 233-248.

Betancourt, H. (2004). Attribution-emotion processes in white's realistic empathy approach to conflict and negotiation. *Peace and Conflict, 10*(4), 369-380.

Białaszek, W., Bakun, P., McGoun, E., & Zielonka, P. (2016). Standing in your peer’s shoes hurts your feats: the self-others discrepancy in risk attitude and impulsivity. *Frontiers in Psychology, 7*, 197.

Bialek, M., & Pennycook, G. (2018). The cognitive reflection test is robust to multiple exposures. *Behavior research methods, 50*(5), 1953-1959.

Biehl, M., Pollock, F. A., & Kanai, R. (2021). A technical critique of some parts of the Free Energy Principle. *Entropy, 23*(3), 293.

Bion, W. (1988). A theory of thinking. In E. Spillius Bott (Ed.), *Melanie Klein Today* (Vol. 1, pp. 178-186).

Birch, S. A. J., & Bloom, P. (2007). The curse of knowledge in reasoning about false beliefs. *Psychological Science, 18*(5), 382.

Blair, R. J. R. (2005). Responding to the emotions of others: Dissociating forms of empathy through the study of typical and psychiatric populations. *Consciousness and cognition, 14*(4), 698-718. doi:10.1016/j.concog.2005.06.004

Bo, S., Sharp, C., Fonagy, P., & Kongerslev, M. (2017). Hypermentalizing, attachment, and epistemic trust in adolescent BPD: Clinical illustrations. *Personality Disorders: Theory, Research, and Treatment, 8*(2), 172.

Bond, C. F. (2008). Commentary a few can catch a liar, sometimes: Comments on Ekman and O'Sullivan (1991), as well as Ekman, O'Sullivan, and Frank (1999). *Applied Cognitive Psychology, 22*(9), 1298-1300.

Bond, C. F., & DePaulo, B. M. (2006). Accuracy of deception judgments. *Personality and Social Psychology Review, 10*(3), 214-234. doi:10.1207/s15327957pspr1003_2

Botvin, G. J., Botvin, E. M., Baker, E., Dusenbury, L., & Goldberg, C. J. (1992). The false consensus effect: predicting adolescents' tobacco use from normative expectations. *Psychological Reports, 70*(1), 171-178.

Brandstätter, V., Lengfelder, A., & Gollwitzer, P. M. (2001). Implementation intentions and efficient action initiation. *Journal of Personality and Social Psychology, 81*(5), 946.

Brett, J. M., & Thompson, L. (2016). Negotiation. *Organizational Behavior and Human Decision Processes, 136*, 68-79. doi:10.1016/j.obhdp.2016.06.003

Britton, R. (1995). Psychic reality and unconscious belief. *International Journal of Psycho-Analysis, 76*, 19-23.

Brosnan, S. F. (2008). How primates (including us!) respond to inequity. *Advances in health economics and health services research, 20*, 99.

Bruner, J. S. (1957). On perceptual readiness. *Psychological Review, 64*(2), 123-152. doi:10.1037/h0043805

Bullinger, A., Wyman, E., Melis, A., & Tomasello, M. (2011). Coordination of chimpanzees (Pan troglodytes) in a Stag Hunt Game. *International Journal of Primatology, 32*(6), 1296-1310. doi:10.1007/s10764-011-9546-3

Burchett, W., & Ellis, A. (2017). npvm: Nonparametric comparison of multivariate samples. Retrieved from https://CRAN.R-project.org/package=npmv.

Burnham, T., McCabe, K., & Smith, V. L. (2000). Friend-or-foe intentionality priming in an extensive form trust game. *Journal of Economic Behavior and Organization, 43*(1), 57-73. doi:10.1016/S0167-2681(00)00108-6

Butler, J. K. (1999). Trust expectations, information sharing, climate of trust, and negotiation effectiveness and efficiency. *Group & Organization Management, 24*(2), 217-238. doi:10.1177/1059601199242005

Byrne, R., & Whiten, A. (1989). Machiavellian intelligence: social expertise and the evolution of intellect in monkeys, apes, and humans (oxford science publications).

Camerer, C., Loewenstein, G., & Weber, M. (1989). The curse of knowledge in economic Settings: An experimental analysis. *The Journal of Political Economy, 97*(5), 1232. doi:10.1086/261651

Campitelli, G., & Labollita, M. (2010). Correlations of cognitive reflection with judgments and choices. *Judgment and Decision Making, 5*(3), 182-191.

Carnevale, P. J., & De Dreu, C. K. W. (2006). Motive: The negotiator's raison d'être. *Frontiers of social psychology: Negotiation theory and research*, 55-76.

Carnevale, P. J., & Isen, A. M. (1986). The influence of positive affect and visual access on the discovery of integrative solutions in bilateral negotiation. *Organizational Behavior and Human Decision Processes, 37*(1), 1-13.

Carnevale, P. J., & Pruitt, D. G. (1992). Negotiation and mediation. *Annual review of psychology, 43*(1), 531-582. doi:10.1146/annurev.ps.43.020192.002531

Carroll, J. S., Bazerman, M. H., & Maury, R. (1988). Negotiator cognitions: A descriptive approach to negotiators' understanding of their opponents. *Organizational Behavior and Human Decision Processes, 41*(3), 352-370. doi:10.1016/0749-5978(88)90034-9

Carter, J. (1982). *Keeping faith: memoirs of a president*. London: Bantam Books.

Chaiken, S., & Trope, Y. (1999). *Dual-process theories in social psychology*: Guilford Press.

Chakrabarti, B., & Baron-Cohen, S. (2006). Empathizing: neurocognitive developmental mechanisms and individual differences. *Progress in Brain Research, 156*, 403-417. doi:10.1016/S0079-6123(06)56022-4

Chambers, J. R., & De Dreu, C. K. W. (2013). Egocentrism drives misunderstanding in conflict and negotiation. *Journal of Experimental Social Psychology, 51*, 15-26. doi:10.1016/j.jesp.2013.11.001

Chen, M., & Bargh, J. A. (1997). Nonconscious behavioral confirmation processes: The self-fulfilling consequences of automatic stereotype activation. *Journal of Experimental Social Psychology, 33*(5), 541-560.

Chugh, D., & Bazerman, M. H. (2007). Bounded awareness: what you fail to see can hurt you. *Cognitive Studies in Economics and Social Sciences, 6*(1), 1-18. doi:10.1007/s11299-006-0020-4

Ciccarelli, M., Nigro, G., D'Olimpio, F., Griffiths, M. D., & Cosenza, M. (2021). Mentalizing failures, emotional dysregulation, and cognitive distortions among adolescent problem gamblers. *Journal of Gambling Studies, 37*(1), 283-298.

Clark, M. S., & Chrisman, K. (1994). Resource allocation in intimate relationships. In *Entitlement and the affectional bond* (pp. 65-88): Springer.

Cohen, T. R. (2010). Moral emotions and unethical bargaining: The differential effects of empathy and perspective taking in deterring deceitful negotiation. *Journal of Business Ethics, 94*(4), 569-579.

Cokely, E. T., & Kelley, C. M. (2009). Cognitive abilities and superior decision making under risk: A protocol analysis and process model evaluation. *Judgment and Decision Making*.

Coleman, P. T., & Joanne Lim, Y. Y. (2001). A systematic approach to evaluating the effects of collaborative negotiation training on individuals and groups. *Negotiation Journal, 17*(4), 363-392. doi:10.1111/j.1571-9979.2001.tb00246.x

Colombo, M., & Wright, C. (2018). First principles in the life sciences: the free-energy principle, organicism, and mechanism. *Synthese*, 1-26.

Csibra, G., & Gergely, G. (1998). The teleological origins of mentalistic action explanations: A developmental hypothesis. *Developmental Science, 1*(2), 255-259. doi:10.1111/1467-7687.00039

Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends in Cognitive Sciences, 13*(4), 148-153. doi:10.1016/j.tics.2009.01.005

Csibra, G., & Gergely, G. (2011). Natural pedagogy as evolutionary adaptation. *Philosophical Transactions of the Royal Society B, 366*(1567), 1149-1157. doi:10.1098/rstb.2010.0319

Csibra, G., Gergely, G., Biró, S., Koós, O., & Brockbank, M. (1999). Goal attribution without agency cues: the perception of 'pure reason’ in infancy. *Cognition, 72*(3), 237-267. doi:10.1016/S0010-0277(99)00039-6

Curhan, J. R., Elfenbein, H., & Xu, H. (2005). What do People Value when they Negotiate? Mapping the Domain of Subjective Value in Negotiation. *IDEAS Working Paper Series from RePEc*.

Curhan, J. R., Elfenbein, H. A., & Eisenkraft, N. (2010). The Objective Value of Subjective Value: A Multi-Round Negotiation Study. *Journal of Applied Social Psychology, 40*(3), 690-709. doi:10.1111/j.1559-1816.2010.00593.x

Curhan, J. R., Elfenbein, H. A., & Xu, H. (2006). What Do People Value When They Negotiate? Mapping the Domain of Subjective Value in Negotiation. *Journal of Personality and Social Psychology, 91*(3), 493-512. doi:10.1037/0022-3514.91.3.493

Curhan, J. R., Neale, M. A., & Ross, L. (2004). Dynamic valuation: Preference changes in the context of face-to-face negotiation. *Journal of Experimental Social Psychology, 40*(2), 142-151. doi:10.1016/j.jesp.2003.12.002

Dadzie, K. Q., Dadzie, C. A., & Williams, A. J. (2018). Trust and duration of buyer-seller relationship in emerging markets. *Journal of Business Industrial Marketing, 33*(1), 134-144.

Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology, 44*(1), 113-126. doi:10.1037/0022-3514.44.1.113

Dawes, R. M., McTavish, J., & Shaklee, H. (1977). Behavior, communication, and assumptions about other people's behavior in a commons dilemma situation. *Journal of Personality and Social Psychology, 35*(1), 1.

De Dreu, C. K. W., Beersma, B., Steinel, W., & Van Kleef, G. A. (2007). The psychology of negotiation. In A. Kruglanski & E. Higgins (Eds.), *Social psychology: Handbook of basic principles* (pp. 608-629). New York; London: Guilford Publications.

De Dreu, C. K. W., Beersma, B., Stroebe, K., & Euwema, M. C. (2006). Motivated information processing, strategic choice, and the quality of negotiated agreement. *Journal of Personality and Social Psychology, 90*(6), 927-943. doi:10.1037/0022-3514.90.6.927

De Dreu, C. K. W., & Carnevale, P. J. (2005a). Disparate methods and common findings in the study of negotiation. *International Negotiation, 10*(1), 193-204. doi:10.1163/1571806054741074

De Dreu, C. K. W., & Carnevale, P. J. (2005b). Laboratory experiments on negotiation and social conflict. *International Negotiation, 10*(1), 51-66. doi:10.1163/1571806054741065

De Dreu, C. K. W., Evers, A., Beersma, B., Kluwer, E. S., & Nauta, A. (2001). A theory-based measure of conflict management strategies in the workplace. *Journal of Organizational Behavior, 22*(6), 645-668.

De Dreu, C. K. W., Giebels, E., & Van de Vliert, E. (1998). Social motives and trust in integrative negotiation: The disruptive effects of punitive capability. *Journal of Applied Psychology, 83*(3), 408.

De Dreu, C. K. W., Koole, S. L., & Steinel, W. (2000). Unfixing the fixed pie: A motivated information-processing approach to integrative negotiation. *Journal of Personality and Social Psychology, 79*(6), 975.

De Dreu, C. K. W., Weingart, L. R., & Kwon, S. (2000). Influence of social motives on integrative negotiation: a meta-analytic review and test of two theories. *Journal of Personality and Social Psychology, 78*(5), 889-905.

De Roo, M., Wong, G., Rempel, G. R., Fraser, S. N., & parenting. (2019). Advancing optimal development in children: examining the construct validity of a Parent Reflective Functioning Questionnaire. *JMIR Pediatrics and Parenting, 2*(1), e11561.

de Weerd, H., Verbrugge, R., & Verheij, B. (2013a). Higher-order theory of mind in negotiations under incomplete information. In *PRIMA 2013: Principles and Practice of Multi-Agent Systems, Lecture Notes in Computer Science* (Vol. 8291, pp. 101-116).

De Weerd, H., Verbrugge, R., & Verheij, B. (2013b). How much does it help to know what she knows you know? An agent-based simulation study. *Artificial Intelligence, 199-200*, 67-92. doi:10.1016/j.artint.2013.05.004

DeCasien, A. R., Williams, S. A., & Higham, J. P. (2017). Primate brain size is predicted by diet but not sociality. *Nature ecology & evolution, 1*(5), 0112.

Decety, J., Cacioppo, J. T., Morsella, E., & Bargh, J. A. (2011). *Unconscious action tendencies: Sources of "un-integrated" action*: Oxford University Press.

Decety, J., Jackson, P. L., Sommerville, J. A., Chaminade, T., & Meltzoff, A. N. (2004). The neural bases of cooperation and competition: an fMRI investigation. *Neuroimage, 23*(2), 744-751. doi:10.1016/j.neuroimage.2004.05.025

Dennett, D. (1971). Intentional Systems. *The Journal of Philosophy, 68*(4), 87-106. doi:10.2307/2025382

Dennett, D. (1978). Beliefs about beliefs. *Behav Brain Sci, 1*(4), 568-570. doi:10.1017/S0140525X00076664

Depaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to Deception. *Psychological Bulletin, 129*(1), 74-118. doi:10.1037/0033-2909.129.1.74

Deutsch, M. (1973). *The resolution of conflict: constructive and destructive processes*. New Haven; London: Yale University Press.

Diaz-Caneja, E. (1928). Sur l'alternance binoculaire. *Annales d'Oculistique, 165*, 721–731.

Diekmann, K. A., Samuels, S. M., Ross, L., & Bazerman, M. H. (1997). Self-interest and fairness in problems of resource allocation: allocators versus recipients. *Journal of Personality and Social Psychology, 72*(5), 1061.

Dunbar, R. I. M. (1992). Neocortex size as a constraint on group size in primates. *Journal of Human Evolution, 22*(6), 469-493. doi:10.1016/0047-2484(92)90081-J

Dunbar, R. I. M. (1998). The social brain hypothesis. *Evolutionary Anthropology: Issues, News, and Reviews, 6*(5), 178-190. doi:10.1002/(SICI)1520-6505(1998)6:5<178::AID-EVAN5>3.0.CO

Dunbar, R. I. M. (2007). The social brain hypothesis and its relevance to social psychology. *Evolution and the social mind: Evolutionary psychology and social cognition*, 21-31.

Dunbar, R. I. M. (2011). The social brain meets neuroimaging. *Trends in Cognitive Sciences*. doi:10.1016/j.tics.2011.11.013

Dunbar, R. I. M. (2014). The Social Brain. *Current Directions in Psychological Science, 23*(2), 109-114. doi:10.1177/0963721413517118

Dunbar, R. I. M., & Shultz, S. (2007). Evolution in the social brain. *Science (New York, N.Y.), 317*(5843), 1344.

Dziobek, I., Fleck, S., Kalbe, E., Rogers, K., Hassenstab, J., Brand, M., . . . Convit, A. (2006). Introducing MASC: a movie for the assessment of social cognition. *Journal of Autism and Developmental Disorders, 36*(5), 623-636.

Eagly, A. H., Wood, W., & Diekman, A. B. (2000). Social role theory of sex differences and similarities: A current appraisal. *The developmental social psychology of gender, 12,* 174.

Eiser, J. R., & Bhavnani, K. K. (1974). The effect of situational meaning on the behaviour of subjects in the Prisoner's Dilemma Game. *European Journal of Social Psychology, 4*(1), 93-97.

Ekman, P. (2009). *Telling lies: Clues to deceit in the marketplace, politics, and marriage (revised edition)*: WW Norton & Company.

Elahee, M. N., Kirby, S. L., & Nasif, E. (2002). National culture, trust, and perceptions about ethical behavior in intra-and cross-cultural negotiations: An analysis of NAFTA countries. *Thunderbird International Business Review, 44*(6), 799-818.

Elfenbein, H., Foo, M., White, J., Tan, H., & Aik, V. (2007). Reading your Counterpart: The Benefit of Emotion Recognition Accuracy for Effectiveness in Negotiation. *J Nonverbal Behav, 31*(4), 205-223. doi:10.1007/s10919-007-0033-7

Ellis, A. R., Burchett, W. W., Harrar, S. W., & Bathke, A. C. (2017). Nonparametric inference for multivariate data: the R package npmv. *Journal of Statistical Software, 76*(4), 1-18.

ElShenawy, E. (2010). Does negotiation training improve negotiators' performance? *Journal of European Industrial Training*.

Epley, N., Caruso, E. M., & Bazerman, M. H. (2006). When perspective taking increases taking: reactive egoism in social interaction. *Journal of Personality and Social Psychology, 91*(5), 872.

Erez, A., & Isen, A. M. J. J. o. A. p. (2002). The influence of positive affect on the components of expectancy motivation. *87*(6), 1055.

Etcoff, N., Ekman, P., Magee, J., & Frank, M., G. (2000). Lie detection and language comprehension. *Nature, 405*(6783), 139. doi:10.1038/35012129

Evans, J. S. B. T., & Stanovich, K. E. (2013). Dual- process theories of higher cognition. *Perspectives on Psychological Science, 8*(3), 223-241. doi:10.1177/1745691612460685

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior research methods, 39*(2), 175-191.

Feather, N. T. (1966). Effects of prior success and failure on expectations of success and subsequent performance. *Journal of Personality and Social Psychology, 3*(3), 287.

Fisher, R., Ury, W., & Patton, B. (1991). *Getting to Yes: Negotiating Agreement Without Giving In* (2 ed.): New York, N.Y: Penguin Books.

Flinn, M. V., Geary, D. C., & Ward, C. V. (2005). Ecological dominance, social competition, and coalitionary arms races: Why humans evolved extraordinary intelligence. *Evolution and Human Behavior, 26*(1), 10-46. doi:10.1016/j.evolhumbehav.2004.08.005

Follet, M. P. (1925). Constructive Conflict. In P. Graham (Ed.), *Mary Parker Follet: Prophet of management* (pp. 67-87). Boston, MA: Harvard Business School Press.

Fonagy, P. (2006). The mentalization-focused approach to social development. In P. Fonagy & P. Luyten (Eds.), *Handbook of mentalization-based treatment* (pp. 53-100).

Fonagy, P., & Allison, E. (2014). The role of mentalizing and epistemic trust in the therapeutic relationship. *Psychotherapy, 51*(3), 372-380. doi:10.1037/a0036505

Fonagy, P., Gergely, G., Jurist, E., & Target, M. (2004). *Affect regulation, mentalization, and the development of the self*. London: Karnac Books.

Fonagy, P., Gergely, G., & Target, M. (2007). The parent-infant dyad and the construction of the subjective self. *Journal of Child Psychology and Psychiatry*(3), 288-328. doi:10.1111/j.1469-7610.2007.01727.x

Fonagy, P., & Luyten, P. (2009). A developmental, mentalization-based approach to the understanding and treatment of borderline personality disorder. *Dev Psychopathol, 21*(4), 1355-1381. doi:10.1017/S0954579409990198

Fonagy, P., & Luyten, P. (2012). Introduction and overview. In *Handbook of metalizing in mental health practice* (pp. 419-444). Washington, DC; London: American Psychiatric Publishing.

Fonagy, P., Luyten, P., & Allison, E. (2015). Epistemic petrification and the restoration of epistemic trust: A new conceptualization of borderline personality disorder and Its psychosocial treatment. *Journal of Personality Disorders, 29*(5), 575-609.

Fonagy, P., Luyten, P., Allison, E., & Campbell, C. (2017). What we have changed our minds about: Part 2. Borderline personality disorder, epistemic trust and the developmental significance of social communication. *Borderline personality disorder and emotion dysregulation, 4*(9). doi:10.1186/s40479-017-0062-8

Fonagy, P., Luyten, P., Moulton-Perkins, A., Lee, Y. W., Warren, F., Howard, S., . . . Lowyck, B. (2016). Development and validation of a self-report measure of mentalizing: The Reflective Functioning Questionnaire. *PLoS One , 11 (7) , Article e0158678. (2016).*

Fonagy, P., Luyten, P., & Strathearn, L. (2011). Borderline personality disorder, mentalization, and the neurobiology of attachment. *Infant Mental Health Journal , 32 (1) 47 - 69. (2011).*

Fonagy, P., & Target, M. (1996). Playing with reality: I. Theory of mind and the normal development of psychic reality. *The International journal of psycho-analysis, 77 (Pt 2)*, 217.

Fonagy, P., & Target, M. (1997). Attachment and reflective function: Their role in self-organization. *Develop. Psychopathol., 9*(4), 679-700.

Fonagy, P., & Target, M. (2000). Playing with reality: III. The persistence of dual psychic reality in borderline patients. *The International journal of psycho-analysis, 81 (Pt 5)*, 853.

Fonagy, P., & Target, M. (2007). Playing with reality: IV. A theory of external reality rooted in intersubjectivity. *The International journal of psycho-analysis, 88*(4), 917.

Fonagy, P., Target, M., Steele, H., & Steele, M. (1998). Reflective-functioning manual, version 5.0, for application to adult attachment interviews. *University College Lodnon Publication*, 161-162.

Forgas, J. P. (1998). On feeling good and getting your way: Mood effects on negotiator cognition and bargaining strategies. *Journal of Personality and Social Psychology, 74*(3), 565.

Förster, J., & Liberman, N. (2007). Knowledge activation. In A. Kruglanski & E. Higgins (Eds.), *Social psychology: Handbook of basic principles*. New York; London: Guilford Publications.

Frank, M. G., & Ekman, P. (1997). The Ability to Detect Deceit Generalizes Across Different Types of High-Stake Lies. *Journal of Personality and Social Psychology, 72*(6), 1429-1439. doi:10.1037/0022-3514.72.6.1429

Frederick, S. (2005). Cognitive Reflection and Decision Making. *Journal of Economic Perspectives, 19*(4), 25-42. doi:10.1257/089533005775196732

Freed, P. (2010). Research Digest. *Neuropsychoanalysis, 12:1*, 103-106. doi:10.1080/15294145.2010.10773634

Friston, K. (2005). A theory of cortical responses. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences, 360*(1456), 815.

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience, 11*(2), 127. doi:10.1038/nrn2787

Friston, K. (2012). Embodied inference and spatial cognition. *Cognitive processing, 13*(1), 171-177.

Friston, K., & Frith, C. (2014). A Duet for one. *Consciousness and cognition*. doi:10.1016/j.concog.2014.12.003

Friston, K., Harrison, L., & Penny, W. (2003). Dynamic causal modelling. *Neuroimage, 19*(4), 1273-1302. doi:10.1016/S1053-8119(03)00202-7

Friston, K., Schwartenbeck, P., Fitzgerald, T., Moutoussis, M., Behrens, T., & Dolan, R. J. (2013). The anatomy of choice: active inference and agency. *Frontiers in human neuroscience*(7), 598.

Friston, K., Schwartenbeck, P., Fitzgerald, T., Moutoussis, M., Behrens, T., & Dolan, R. J. (2014). The anatomy of choice: dopamine and decision-making. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences, 369*(1655). doi:10.1098/rstb.2013.0481

Frith, U., & Frith, C. D. (2003). Development and neurophysiology of mentalizing. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences, 358*(1431), 459.

Froman, L. A., & Cohen, M. (1970). Compromise and logroll: Comparing the efficiency of two bargaining processes. *Behavioral Science, 15*(2), 180-183.

Furnham, A., & Boo, H. C. (2011). A literature review of the anchoring effect. *Journal of Socio-Economics, 40*(1), 35-42. doi:10.1016/j.socec.2010.10.008

Galinsky, A. D., Maddux, W. W., Gilin, D., & White, J. B. (2008). Why it pays to get inside the head of your opponent: The differential effects of perspective taking and empathy in negotiations. *Psychological Science, 19*(4), 378. doi:10.1111/j.1467-9280.2008.02096.x

Galinsky, A. D., & Mussweiler, T. (2001). First offers as anchors: The role of perspective-taking and negotiator focus. *Journal of Personality and Social Psychology, 81*(4), 657-669. doi:10.1037/0022-3514.81.4.657

Gehlbach, H., Marietta, G., King, A. M., Karutz, C., Bailenson, J. N., & Dede, C. (2015). Many ways to walk a mile in another’s moccasins: Type of social perspective taking and its effect on negotiation outcomes. *Computers in Human Behavior, 52*, 523-532.

Gergely, G., & Csibra, G. (2003). Teleological reasoning in infancy: the naïve theory of rational action. *Trends in Cognitive Sciences, 7*(7), 287-292. doi:10.1016/S1364-6613(03)00128-1

Gergely, G., Nádasdy, Z., Csibra, G., & Bíró, S. (1995). Taking the intentional stance at 12 months of age. *Cognition, 56*(2), 165-193. doi:10.1016/0010-0277(95)00661-H

Giebels, E., De Dreu, C. K. W., & Van De Vliert, E. (2000). Interdependence in negotiation: effects of exit options and social motive on distributive and integrative negotiation. *European Journal of Social Psychology, 30*(2), 255-272. doi:10.1002/(SICI)1099-0992(200003/04)30:2<255::AID-EJSP991>3.0.CO2-7

Gilin, D., Maddux, W. W., Carpenter, J., & Galinsky, A. D. (2013). When to use your head and when to use your heart. *Personality and Social Psychology Bulletin, 39*(1), 3-16. doi:10.1177/0146167212465320

Gilovich, T. (1990). Differential construal and the false consensus effect. *Journal of Personality and Social Psychology, 59*(4), 623.

Gilovich, T. (2008). *How we know what isn't so*: Simon and Schuster.

Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review, 102*(1), 4-27. doi:10.1037/0033-295X.102.1.4

Grice, H. P. (1957). Meaning. *The philosophical review, 66*(3), 377-388.

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. Morgan (Eds.), *Syntax and Semantics* (Vol. 3). New York: Academic Press.

Griffin, D. W., & Ross, L. (1991). Subjective construal, social inference, and human misunderstanding. *Advances in Experimental Social Psychology, 24*(C), 319-359. doi:10.1016/S0065-2601(08)60333-0

Gunia, B. C., Brett, J. M., Nandkeolyar, A. K., & Kamdar, D. (2011). Paying a price: culture, trust, and negotiation consequences. *Journal of Applied Psychology, 96*(4), 774-789. doi:10.1037/a0021986

Gunia, B. C., Swaab, R. I., Sivanathan, N., & Galinsky, A. D. (2013). The remarkable robustness of the first-offer effect. *Personality and Social Psychology Bulletin, 39*(12), 1547-1558. doi:10.1177/0146167213499236

Güth, W., Schmittberger, R., & Schwarze, B. (1982). An experimental analysis of ultimatum bargaining. *Journal of Economic Behavior and Organization, 3*(4), 367-388. doi:10.1016/0167-2681(82)90011-7

Haigh, M. J. A. i. c. p. (2016). Has the standard cognitive reflection test become a victim of its own success? *, 12*(3), 145.

Hare, B., & Tomasello, M. (2004). Chimpanzees are more skilful in competitive than in cooperative cognitive tasks. *Animal Behaviour, 68*(3), 571-581.

Hartwig, M., & Bond, C. F. (2014). Lie detection from multiple cues: A meta-analysis. *Applied Cognitive Psychology, 28*(5), 661-676.

Hastorf, A., & Cantril, H. (1954). They saw a game. *Journal of Abnormal and Social Psychology, 49*(1), 129-134.

Hayes, A. F. (2020). PROCESS macro for SPSS, SAS and R (Version 3.5).

Heider, F., & Simmel, M. (1944). An experimental study of apparent behavior. *The American Journal of Psychology, 57*(2), 243-259. doi:10.2307/1416950

Herbst, U., & Schwarz, S. (2011). How Valid Is Negotiation Research Based on Student Sample Groups? New Insights into a Long-Standing Controversy. *Negotiation Journal, 27*(2), 147-170. doi:10.1111/j.1571-9979.2011.00300.x

Heyes, C. M., & Frith, C. D. (2014). The cultural evolution of mind reading. *Science (New York, N.Y.), 344*(6190), 1243091. doi:10.1126/science.1243091

Higgins, E. T. (1996). Knowledge activation: Accessibility, applicability, and salience. *E. T. Higgins & A. W. Kruglanski (Eds.), Social psychology: Handbook of basic principles*.

Higgins, E. T., Rholes, W. S., & Jones, C. R. (1977). Category accessibility and impression formation. *Journal of Experimental Social Psychology, 13*(2), 141-154. doi:10.1016/S0022-1031(77)80007-3

Hill, W. E. (1915). My wife and my mother in law. Retrieved from https://www.loc.gov/pictures/item/2010652001/

Hockey, G. R. J. (1970). Effect of loud noise on attentional selectivity. *Quarterly Journal of Experimental Psychology, 22*(1), 28-36. doi:10.1080/14640747008401898

Hohwy, J. (2012). Attention and conscious perception in the hypothesis testing brain. *Frontiers in Psychology, 3*, 96.

Hohwy, J. (2014). *The predictive mind*. Oxford: Oxford University Press.

Hohwy, J. (2020). Self-supervision, normativity and the free energy principle. *Synthese*, 1-25.

Hohwy, J., Roepstorff, A., & Friston, K. (2008). Predictive coding explains binocular rivalry: An epistemological review. *Cognition, 108*(3), 687-701.

Hohwy, J., & Rosenberg, R. (2005). Unusual experiences, reality testing and delusions of alien control. *Journal of Mind and Language, 20*(2), 141-162.

Howard, E. S., Gardner, W. L., & Thompson, L. (2007). The role of the self-concept and the social context in determining the behavior of power holders: Self-construal in intergroup versus dyadic dispute resolution negotiations. *Journal of Personality and Social Psychology, 93*(4), 614.

Howieson, J., & Priddis, L. (2012). Mentalising in mediation: Towards an understanding of the "mediation shift". *Australasian Dispute Resolution Journal, 23*(1), 25-60.

Howieson, J., & Priddis, L. (2015). A Mentalizing-Based Approach to Family Mediation: Harnessing Our Fundamental Capacity to Resolve Conflict and Building an Evidence-Based Practice for the Field. *Family Court Review, 53*(1), 79-95.

Huang, Y. L., Fonagy, P., Feigenbaum, J., Montague, P. R., Nolte, T., & Consortium, M. D. R. (2020). Multidirectional pathways between attachment, mentalizing, and posttraumatic stress symptomatology in the context of childhood trauma. *Psychopathology, 53*(1), 48-58.

Huber, V. L., & Neale, M. A. (1986). Effects of cognitive heuristics and goals on negotiator performance and subsequent goal setting. *Organizational Behavior and Human Decision Processes, 38*(3), 342-365. doi:10.1016/0749-5978(86)90005-1

Humphrey, N. K. (1976). The social function of intellect. In *Growing points in ethology* (pp. 303-317). Cambridge: Cambridge University Press.

Idson, L. C., Chugh, D., Bereby-meyer, Y., Moran, S., Grosskopf, B., & Bazerman, M. H. (2004). Overcoming focusing failures in competitive environments. *Journal of Behavioral Decision Making, 17*(3), 159-172. doi:10.1002/bdm.467

Jackson, H. E., Kaplow, L., Shavell, S., Viscusi, W. K., & Cope, D. (2011). *Analytical Methods for Lawyers*: Thomas Reuters/Foundation Press.

Jann, B. (2021). robreg: Stata module providing robust regression estimators. Retrieved from http://ideas.repec.org/c/boc/bocode/s458931.html

Janssen, O., & van de Vliert, E. (1996). Concern for the other's goals: Key to (de)escalation of conflict. *International Journal of Conflict Management, 7*(2), 99-120. doi:10.1108/eb022777

Johnston, W. A., Hawley, K. J. J. P. B., & Review. (1994). Perceptual inhibition of expected inputs: The key that opens closed minds. *1*(1), 56-72.

Kahneman, D. (1992). Reference points, anchors, norms, and mixed feelings. *Organizational Behavior and Human Decision Processes, 51*(2), 296-312. doi:10.1016/0749-5978(92)90015-Y

Kahneman, D. (2003). A perspective on judgment and choice: mapping bounded rationality. *American Psychologist, 58*(9), 697-720. doi:10.1037/0003-066X.58.9.697

Kahneman, D. (2011). *Thinking, fast and slow*. London: London: Allen Lane.

Kahneman, D., & Klein, G. (2009). Conditions for intuitive expertise: a failure to disagree. *American Psychologist, 64*(6), 515.

Kahneman, D., Knetsch, J., & Thaler, R. (1990). Experimental tests of the endowment effect and the Coase theorem. *The Journal of Political Economy, 98*(6), 1325. doi:10.1086/261737

Kahneman, D., & Tversky, A. (1984). Choices, values, and frames. *American Psychologist, 39*(4), 341-350. doi:10.1037/0003-066X.39.4.341

Kay, A. C., & Ross, L. (2003). The perceptual push: The interplay of implicit cues and explicit situational construals on behavioral intentions in the Prisoner’s Dilemma. *Journal of Experimental Social Psychology, 39*(6), 634-643.

Kay, A. C., Wheeler, S. C., Bargh, J. A., & Ross, L. (2004). Material priming: The influence of mundane physical objects on situational construal and competitive behavioral choice. *Organizational Behavior and Human Decision Processes, 95*(1), 83-96.

Keltner, D., & Robinson, R. J. (1993). Imagined ideological differences in conflict escalation and resolution. *International Journal of Conflict Management, 4*(3), 249-262. doi:10.1108/eb022728

Keltner, D., & Robinson, R. J. (1996). Extremism, Power, and the Imagined Basis of Social Conflict. *Current Directions in Psychological Science, 5*(4), 101-105. doi:10.1111/1467-8721.ep11452765

Keltner, D., & Robinson, R. J. (1997). Defending the Status Quo: Power and Bias in Social Conflict. *Personality and Social Psychology Bulletin, 23*(10), 1066-1077. doi:10.1177/01461672972310007

Kemp, K. E., & Smith, W. P. (1994). Information exchange, toughness, and integrative bargaining: The roles of explicit cues and perspective-taking. *International Journal of Conflict Management*.

Keysers, C., & Gazzola, V. (2006). Towards a unifying neural theory of social cognition. *Progress in Brain Research, 156*, 379-401. doi:10.1016/S0079-6123(06)56021-2

Kiefel, M., & Bathke, A. C. (2018). Package 'nparMD'. Retrieved from https://cran.r-project.org/web/packages/nparMD/nparMD.pdf

Kimmel, M. J., Pruitt, D. G., Magenau, J. M., Konar-Goldband, E., & Carnevale, P. J. (1980). Effects of trust, aspiration, and gender on negotiation tactics. *Journal of Personality and Social Psychology, 38*(1), 9.

Klein, M. (1921). The development of a child. II. Early analysis. I. *The Psycho-Analysis of Children. Writings, 1*, 173.

Klein, M. (1930). The importance of symbol-formation in the development of the ego. *International Journal of Psycho-Analysis, 11*, 24-39.

Klein, M. (1935). A contribution to the psychogenesis of manic-depressive states. *International Journal of Psycho-Analysis, 16*, 145-174.

Kong, D. T., Dirks, K. T., & Ferrin, D. L. (2014). Interpersonal trust within negotiations: Metaanalytic evidence, critical contingencies, and directions for future research. *Academy of Management Journal, 57*(5), 1235-1255. doi:10.5465/amj.2012.0461

Kovács, Á. M., Téglás, E., & Endress, A. D. (2010). The social sense: Susceptibility to others' beliefs in human infants and adults. *Science, 330*(6012), 1830-1834.

Kramer, R. M., Newton, E., & Pommerenke, P. L. (1993). Self-enhancement biases and negotiator judgment: Effects of self-esteem and mood. *Organizational Behavior and Human Decision Processes, 56*(1), 110-133.

Ku, G., & Galinsky, A. D. (2006). The benefits and limits of perspective-taking for intergroup attitudes, expectancy confirmation, and negotiations. *7th Annual Meeting of the Society of Personality and Social Psychology*.

Larrick, R. P., & Blount, S. (1997). The claiming effect: Why players are more generous in social dilemmas than in ultimatum games. *Journal of Personality and Social Psychology, 72*(4), 810.

Lax, D., & Sebenius, J. K. (1986a). Interests: The measure of negotiation. *On the Process of Dispute Settlement, 2*(1), 73-92. doi:10.1007/BF00998936

Lax, D., & Sebenius, J. K. (1986b). *The Manager as Negotiator: Bargaining for Cooperation and Competitive gain*. New York, London: Free Press.

Lax, D., & Sebenius, J. K. (2006). *3-D Negotiation: Powerful tools to change the game in your most important deals*. Boston: Harvard Business Press.

Leary, K. (2004). Critical moments as relational moments: The centre for humanitarian dialogue and the conflict in Aceh, Indonesia. *Negotiation Journal, 20*(2), 311-338.

Leary, K. (2008). Critical moments as relational moments. In E. Jurist, A. Slade, & S. Bergner (Eds.), *Mind to mind: Infant research, neuroscience and psychoanalysis* (pp. 335-352). New York, NY: Other Press.

Lee, S., Adair, W. L., & Seo, S.-J. (2013). Cultural perspective taking in cross-cultural negotiation. *Group Decision and Negotiation, 22*(3), 389-405.

Lee, Y., Meins, E., & Larkin, F. (2020). Translation and preliminary validation of a Korean version of the parental reflective functioning questionnaire. *Infant Mental Health Journal*.

Levine, E. E., Barasch, A., Rand, D., Berman, J. Z., & Small, D. A. (2018). Signaling emotion and reason in cooperation. *Journal of Experimental Psychology: General, 147*(5), 702-719. doi:10.1037/xge0000399

Lewicki, R. (2014). Teaching negotiation: the state of the practice. In O. B. Ayoko, N. M. Ashkanasy, & K. A. Jehn (Eds.), *Handbook of conflict management research*: Edward Elgar Publishing.

Lewicki, R., Bunker, B. B., & Research. (1996). Developing and maintaining trust in work relationships. In R. Kramer & T. Tyler (Eds.), *Trust in Organizations: Frontiers of Theory and Research*. Thousand Oaks: Sage Publications.

Lewicki, R., Stevenson, M. A., & Bunker, B. B. (1997). *The three components of interpersonal trust: Instrument development and differences across relationships*. Paper presented at the Academy of Management Meeting.

Lewin, K., Dembo, T., Festinger, L., & Sears, P. S. (1944). Level of aspiration. In J. M. Hunt (Ed.), *Personality and the behavior disorder*

(Vol. 1). New York: Ronald Press.

Li, E. T., Carracher, E., & Bird, T. (2020). Linking childhood emotional abuse and adult depressive symptoms: the role of mentalizing incapacity. *Child Abuse & Neglect, 99*, 104253.

Liberman, V., Samuels, S., & Ross, L. (2004). The name of the game: Predictive power of reputations versus situational labels in determining prisoner's dilemma game moves. *Personality and Social Psychology Bulletin, 30*(9), 1175-1185.

Lieberman, M. D. (2005). Principles, processes, and puzzles of social cognition: An introduction for the special issue on social cognitive neuroscience. *Neuroimage, 28*(4), 745-756. doi:10.1016/j.neuroimage.2005.07.028

Lieberman, M. D. (2007). Social cognitive neuroscience: a review of core processes. *Annual review of psychology, 58*, 259.

Lieberman, M. D. (2013). *Social: Why our Brains are Wired to Connect*: Oxford University Press.

Lieberman, M. D., Gaunt, R., Gilbert, D. T., & Trope, Y. (2002). Reflexion and reflection: A social cognitive neuroscience approach to attributional inference. *Advances in Experimental Social Psychology, 34*, 199-249. doi:10.1016/S0065-2601(02)80006-5

Lindsey, C., & Sheather, S. (2010). Power transformation via multivariate Box–Cox. *The Stata Journal, 10*(1), 69-81.

Lissek, S., Peters, S., Fuchs, N., Witthaus, H., Nicolas, V., Tegenthoff, M., . . . Brüne, M. (2008). Cooperation and Deception Recruit Different Subsets of the Theory-of-Mind Network (Theory-of-Mind). *PloS one, 3*(4), e2023. doi:10.1371/journal.pone.0002023

Locke, E. A., Shaw, K. N., Saari, L. M., & Latham, G. P. J. P. b. (1981). Goal setting and task performance: 1969–1980. *90*(1), 125.

Lord, C. G., Ross, L., & Lepper, M. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology, 37*(11), 2098.

Loschelder, D. D., Swaab, R. I., Trötschel, R., & Galinsky, A. D. (2014). The first-mover disadvantage: The folly of revealing compatible preferences. *Psychological Science, 25*(4), 954-962.

Luyten, P., & Fonagy, P. (2015). The neurobiology of mentalizing. *Personality Disorders: Theory, Research, and Treatment, 6*(4), 366-379. doi:10.1037/per0000117

Luyten, P., Fonagy, P., Lowyck, B., & Vermote, R. (2012). Assessment of mentalization. In [check] (Ed.), *Handbook of mentalizing in mental health practice* (pp. 43-65). London: [check].

Luyten, P., Mayes, L. C., Nijssens, L., & Fonagy, P. (2017). The parental reflective functioning questionnaire: Development and preliminary validation. *PloS one, 12*(5), e0176218.

Lysaker, P., Gumley, A., Luedtke, B., Buck, K., Ringer, J., Olesek, K., . . . Dimaggio, G. (2013). Social cognition and metacognition in schizophrenia: evidence of their independence and linkage with outcomes. *Acta Psychiatrica Scandinavica, 127*(3), 239-247.

Maddux, W. W., Mullen, E., & Galinsky, A. D. (2008). Chameleons bake bigger pies and take bigger pieces: Strategic behavioral mimicry facilitates negotiation outcomes. *Journal of Experimental Social Psychology, 44*(2), 461-468. doi:10.1016/j.jesp.2007.02.003

Magee, J. C., Galinsky, A. D., & Gruenfeld, D. H. (2007). Power, propensity to negotiate, and moving first in competitive interactions. *Personality and Social Psychology Bulletin, 33*(2), 200-212.

Malcorps, S., Vliegen, N., Nijssens, L., Tang, E., Casalin, S., Slade, A., & Luyten, P. (2021). Assessing reflective functioning in prospective adoptive parents. *PloS one, 16*(1), e0245852.

Malmendier, U., Moretti, E., & Peters, F. S. (2018). Winning by losing: Evidence on the long-run effects of mergers. *The Review of Financial Studies, 31*(8), 3212-3264. doi:10.1093/rfs/hhy009

Maoz, I., Ward, A., Katz, M., & Ross, L. (2002). Reactive devaluation of an 'Israeli' vs. 'Palestinian' peace proposal. *The Journal of Conflict Resolution, 46*(4), 515-546.

Marchi, S., Targi, N., Liston, P. M., & Parlangeli, O. (2020). The possible role of empathy and emotions in virtual negotiation. *Ergonomics, 63*(3), 263-273.

Marietta, G. E., Brooks, J., Hahn, E. P., Xu, M., Dede, C., & Gehlbach, H. (2013). Encouraging Compromise Through Perspective Taking in Virtual Simulations. *Conference: Presentation*.

Martin, R. D. (1981). Relative brain size and basal metabolic rate in terrestrial vertebrates. *Nature, 293*(5827), 57-60.

Marwell, G., & Schmitt, D. R. (1972). Cooperation in a three-person Prisoner's Dilemma. *Journal of Personality and Social Psychology, 21*(3), 376.

Matte-Blanco, I. (2003). *Thinking, feeling, and being*: Routledge.

Matte-Blanco, I., & Rayner, E. (2018). *The unconscious as infinite sets: An essay in bi-logic*: Routledge.

Mayer, R. C., & Davis, J. H. (1999). The effect of the performance appraisal system on trust for management: A field quasi-experiment. *Journal of Applied Psychology, 84*(1), 123-136. doi:10.1037/0021-9010.84.1.123

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of management review, 20*(3), 709-734.

Mayes, L. C. (2000). A developmental perspective on the regulation of arousal states. *Seminars in Perinatology, 24*(4), 267-279. doi:10.1053/sper.2000.9121

Mayes, L. C. (2006). Arousal regulation, emotional flexibility, medial amygdala function, and the impact of early experience. *Annals of the New York Academy of Sciences, 1094*(1), 178-192. doi:10.1196/annals.1376.018

Meins, E., Fernyhough, C., Wainwright, R., Clark-carter, D., Das Gupta, M., Fradley, E., & Tuckey, M. (2003). Pathways to understanding mind: construct validity and predictive validity of maternal mind-mindedness. *Child Development, 74*(4), 1194-1211. doi:10.1111/1467-8624.00601

Menkel-Meadow, C. (1983). Toward another view of legal negotiation: The structure of problem solving. *UCLA Law Review, 31*, 754.

Meyer, A., Zhou, E., & Shane, F. (2018). The non-effects of repeated exposure to the Cognitive Reflection Test. *Judgment and Decision Making, 13*(3), 246.

Miller, G. A., Galanter, E., & Pribram, K. H. (1960). *Plans and the Structure of Behavior*. New York: Henry Holt.

Mislin, A. A., Campagna, R. L., & Bottom, W. P. (2011). After the deal: Talk, trust building and the implementation of negotiated agreements. *Organizational Behavior and Human Decision Processes, 115*(1), 55-68.

Mitchell, J. (2013). *Siblings: Sex and violence*: John Wiley & Sons.

Mnookin, R. H. (2000). *Beyond Winning: Negotiating to Create Value in Deals and Disputes*. London: Belknap Press of Harvard University Press.

Mnookin, R. H. (2010). *Bargaining with the devil: When to negotiate, when to fight*. New York, NY: Simon and Schuster.

Moffitt, M. (2005). Pleadings in the Age of Settlement. *Indiana Law Journal, 80*, 727.

Möhring, K., & Schmidt-Catran, A. (2013). MLT: Stata module to provide multilevel tools. Retrieved from https://ideas.repec.org/c/boc/bocode/s457577.html

Moll, H., & Tomasello, M. (2007). Cooperation and human cognition: the Vygotskian intelligence hypothesis. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences, 362*(1480), 639.

Moran, S., & Ritov, I. (2000). Logrolling without understanding why: Effects of anchoring, experience, and perspective taking. *Conference: Presentation.*

Moran, S., & Ritov, I. (2002). Initial perceptions in negotiations: Evaluation and response to 'logrolling'offers. *Journal of Behavioral Decision Making, 15*(2), 101-124.

Morandotti, N., Brondino, N., Merelli, A., Boldrini, A., De Vidovich, G. Z., Ricciardo, S., . . . Fonagy, P. (2018). The Italian version of the Reflective Functioning Questionnaire: Validity data for adults and its association with severity of borderline personality disorder. *PloS one, 13*(11), e0206433.

Morelli, S. A., & Lieberman, M. D. (2013). The role of automaticity and attention in neural processes underlying empathy for happiness, sadness, and anxiety. *Frontiers in human neuroscience, 7*, 160. doi:10.3389/fnhum.2013.00160

Morris, E. (Writer). (2004). The Fog of War: Eleven Lessons from the Life of Robert S. McNamara. In. USA: Sony Pictures Classics.

Moutoussis, M., Fearon, P., El-Deredy, W., Dolan, R. J., & Friston, K. J. (2014). Bayesian inferences about the self (and others): A review. *Consciousness and cognition, 25*, 67-76. doi:10.1016/j.concog.2014.01.009

Moutoussis, M., Trujillo-Barreto, N. J., El-Deredy, W., Dolan, R. J., & Friston, K. J. (2014). A formal model of interpersonal inference. *Frontiers in human neuroscience, 8*, 160. doi:10.3389/fnhum.2014.00160

Movius, H. (2008). The effectiveness of negotiation training. *Negotiation Journal, 24*(4), 509-531. doi:10.1111/j.1571-9979.2008.00201.x

Murnighan, J. K., Babcock, L., Thompson, L., & Pillutla, M. (1999). The information dilemma in negotiations: Effects of experience, incentives, and integrative potential. *International Journal of Conflict Management, 10*(4), 313-339.

Nadler, J., & Thompson, L. (2003). Learning negotiation skills: Four models of knowledge creation and transfer. *Management Science, 49*(4), 529-540.

Nalebuff, B. J., & Brandenburger, A. (1996). *Co-opetition*. London: Harper Collins.

Neale, M. A., & Bazerman, M. H. (1983). The role of perspective-taking ability in negotiating under different forms of arbitration. *Industrial and Labor Relations Review, 36*(3), 378-388. doi:10.2307/2523017

Neale, M. A., & Bazerman, M. H. (1992a). Negotiating rationally: The power and impact of the negotiator's frame. *The Executive, 6*(3), 42. doi:10.5465/AME.1992.4274183

Neale, M. A., & Bazerman, M. H. (1992b). Negotiator cognition and rationality: A behavioral decision theory perspective. *Organizational Behavior and Human Decision Processes, 51*(2), 157-175. doi:10.1016/0749-5978(92)90009-V

Neale, M. A., & Northcraft, G. B. (1986). Experts, amateurs, and refrigerators: Comparing expert and amateur negotiators in a novel task. *Organizational Behavior and Human Decision Processes, 38*(3), 305-317.

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84*(3), 231-259.

Northcraft, G. B., & Neale, M. A. (1987). Experts, amateurs, and real estate: An anchoring-and-adjustment perspective on property pricing decisions. *Organizational Behavior and Human Decision Processes, 39*(1), 84-97. doi:10.1016/0749-5978(87)90046-X

O'Connor, K. M., & Adams, A. A. (1999). What novices think about negotiation: A content analysis of scripts. *Negotiation Journal, 15*(2), 135-147. doi:10.1111/j.1571-9979.1999.tb00187.x

Ochsner, K. N., & Lieberman, M. D. (2001). The Emergence of Social Cognitive Neuroscience. *American Psychologist, 56*(9), 717-734.

Oechssler, J., Roider, A., & Schmitz, P. W. (2009). Cognitive abilities and behavioral biases. *Journal of Economic Behavior and Organization, 72*(1), 147-152.

Oettingen, G., & Mayer, D. (2002). The motivating function of thinking about the future: expectations versus fantasies. *Journal of Personality and Social Psychology, 83*(5), 1198.

Olekalns, M., Lau, F., & Smith, P. (2007). Resolving the empty core: Trust as a determinant of outcomes in three-party negotiations. *Group Decision and Negotiation, 16*(6), 527-538. doi:10.1007/s10726-007-9084-8

Olekalns, M., & Smith, P. L. (2005). Moments in time: Metacognition, trust, and outcomes in dyadic negotiations. *Personality and Social Psychology Bulletin, 31*(12), 1696-1707. doi:10.1177/0146167205278306

Olekalns, M., & Smith, P. L. (2009). Mutually dependent: Power, trust, affect and the use of deception in negotiation. *Journal of Business Ethics, 85*(3), 347-365.

Olekalns, M., & Smith, P. L. (2011). MindSets: Sensemaking and transition in negotiation. In W. Donahue, R. G. Rogan, & S. Kaufman (Eds.), *Framing Matters* (pp. 51-70). New York: Peter Lang.

Panksepp, J. (2004). *Affective neuroscience: The foundations of human and animal emotions*: United Kingdom: Oxford University Press.

Panksepp, J., & Biven, L. (2012). *The archaeology of mind: neuroevolutionary origins of human emotions (Norton series on interpersonal neurobiology)*: WW Norton & Company.

Patton, B. (2005). Negotiation. In M. Moffitt & R. Bordone (Eds.), *The Handbook of Dispute Resolution* (pp. 279-303). Boston, MA: Wiley & Sons.

Patton, B. (2009). The Deceptive Simplicity of Teaching Negotiation: Reflections on Thirty Years of the Negotiation Workshop. *Negotiation Journal, 25*(4), 481-498. doi:10.1111/j.1571-9979.2009.00240.x

Perner, J., & Wimmer, H. (1985). “John thinks that Mary thinks that…” attribution of second-order beliefs by 5- to 10-year-old children. *Journal of Experimental Child Psychology, 39*(3), 437-471. doi:10.1016/0022-0965(85)90051-7

Pillutla, M. M., & Chen, X.-P. (1999). Social norms and cooperation in social dilemmas: The effects of context and feedback. *Organizational Behavior and Human Decision Processes, 78*(2), 81-103.

Pinkley, R. L., Griffith, T. L., & Northcraft, G. B. (1995). "Fixed pie" a la mode: Information availability, information processing, and the negotiation of suboptimal agreements. *Organizational Behavior and Human Decision Processes, 62*(1), 101-112.

Pituch, K. A., & Stevens, J. P. (2015). *Applied multivariate statistics for the social sciences: Analyses with SAS and IBM’s SPSS* (6 ed.). New York: Routledge.

Porter, S., & Brinke, L. (2010). The truth about lies: What works in detecting high-stakes deception? *Legal and Criminological Psychology, 15*(1), 57-75. doi:10.1348/135532509X433151

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behav Brain Sci, 1*(4), 515-526. doi:10.1017/S0140525X00076512

Pronin, E., Gilovich, T., & Ross, L. (2004). Objectivity in the eye of the beholder: Divergent perceptions of bias in self versus others. *Psychological Review, 111*(3), 781-799.

Provis, C. (1996). Interests vs. positions: A critique of the distinction. *Negotiation Journal, 12*(4), 305-323. doi:10.1111/j.1571-9979.1996.tb00105.x

Pruitt, D. G. (1983a). Achieving integrative agreements In M. H. Bazerman & R. Lewicki (Eds.), *Negotiating in organizations* (pp. 35–50). Beverly Hills, CA: Sage.

Pruitt, D. G. (1983b). Strategic Choice in Negotiation. *American Behavioral Scientist, 27*(2), 167-194. doi:10.1177/000276483027002005

Pruitt, D. G., & Drews, J. L. (1969). The effect of time pressure, time elapsed, and the opponent's concession rate on behavior in negotiation. *Journal of Experimental Social Psychology, 5*(1), 43-60.

Pruitt, D. G., & Lewis, S. A. (1975). Development of integrative solutions in bilateral negotiation. *Journal of Personality and Social Psychology, 31*(4), 621-633. doi:10.1037/0022-3514.31.4.621

Pruitt, D. G., & Rubin, J. (1986). *Social conflict: Escalation, impasse, and resolution*. Reding, MA: Addision-Wesley.

R Core Team. (2013). R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing.

Rahim, M. A., Psenicka, C., Polychroniou, P., Zhao, J. H., Yu, C. S., Chan, K. A., . . . Ralunan, S. (2002). A model of emotional intelligence and conflict management strategies: A study in seven countries. *The International journal of organizational analysis*.

Raiffa, H. (1982). *The art and science of negotiation*. Cambridge, Mass ; London: Cambridge, Mass ; London : Belknap Press of Harvard University Press.

Raiffa, H. (2002). *Negotiation analysis: The science and art of collaborative decision making*: Cambridge, Mass.; London: Belknap Press of Harvard University Press.

Rameson, L. T., & Lieberman, M. D. (2009). Empathy: A social cognitive neuroscience approach. *Social and Personality Psychology Compass, 3*(1), 94-110. doi:10.1111/j.1751-9004.2008.00154.x

Ramirez-Fernandez, J., Ramirez-Marin, J. Y., & Munduate, L. (2018). I expected more from you: The influence of close relationships and perspective taking on negotiation offers. *Group Decision and Negotiation, 27*(1), 85-105.

Rand, D. G., Greene, J. D., & Nowak, M. A. (2012). Spontaneous giving and calculated greed. *Nature, 489*(7416), 427. doi:10.1038/nature11467

Rapoport, A. (1969). Editorial comment. *Jouirnal of Conflict Resolution, 13*, 511-520.

Rilling, J. K., Gutman, D. A., Zeh, T. R., Pagnoni, G., Berns, G. S., & Kilts, C. D. (2002). A Neural Basis for Social Cooperation. *Neuron, 35*(2), 395-405. doi:10.1016/S0896-6273(02)00755-9

Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annu. Rev. Neurosci., 27*, 169-192.

Robinson, R. J., Keltner, D., Ward, A., & Ross, L. (1995). Actual versus assumed differences in construal:" Naive realism" in intergroup perception and conflict. *Journal of Personality and Social Psychology, 68*(3), 404.

Roese, N. J., & Sherman, J. W. (2007). Expectancy. In *Social psychology: Handbook of basic principles*: Guilford Publications.

Rogers, T., Moore, D. A., & Norton, M. I. (2017). The belief in a favorable future. *Psychological Science, 28*(9), 1290-1301.

Rosati, A. G. (2017). Foraging cognition: reviving the ecological intelligence hypothesis. *Trends in Cognitive Sciences, 21*(9), 691-702.

Ross, L. (1995). Reactive devaluation in negotiaton and conflict resolution. In K. J. Arrow, R. H. Mnookin, L. Ross, A. Tversky, & R. Wilson (Eds.), *Barriers to conflict resolution*: WW Norton & Company.

Ross, L. (2014). Barriers to agreement in the asymmetric Israeli–Palestinian conflict. *Pathways toward terrorism and genocide, 7*(2-3), 120-136. doi:10.1080/17467586.2014.970565

Ross, L., & Nisbett, R. E. (1991). *The Person and the Situation: Perspectives of Social Psychology*. London: McGraw-Hill.

Ross, L., & Ward, A. (1995). Psychological barriers to dispute resolution. *Advances in Experimental Social Psychology, 27*(C), 255-304. doi:10.1016/S0065-2601(08)60407-4

Roth, P. (2001). The paranoid-schizoid position. In C. Bronstein (Ed.), *Kleinian theory: A contemporary perspective* (pp. 32-46).

Rubin, J. Z., & Brown, B. R. (2013). *The social psychology of bargaining and negotiation*. New York: Elsevier.

Salaminios, G., Morosan, L., Toffel, E., Tanzer, M., Eliez, S., Badoud, D., . . . Debbané, M. (2020). Associations between schizotypal personality features, mentalizing difficulties and thought problems in a sample of community adolescents. *Early Intervention in Psychiatry*.

Sánchez-Amaro, A., Duguid, S., Call, J., & Tomasello, M. (2019). Chimpanzees and children avoid mutual defection in a social dilemma. *Evolution and Human Behavior, 40*(1), 46-54.

Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E., & Cohen, J. D. (2003). The neural basis of economic decision-making in the ultimatum game. *Science, 300*(5626), 1755-1758.

Satpute, A. B., & Lieberman, M. D. (2006). Integrating automatic and controlled processes into neurocognitive models of social cognition. *Brain Research, 1079*(1), 86-97. doi:10.1016/j.brainres.2006.01.005

Schelling, T. C. (1980). *The Strategy of Conflict*: Cambridge, Mass.; London: Harvard University.

Schlegel, K. (2021). The effects of emotion recognition training on interpersonal effectiveness. *Basic and Applied Social Psychology, 43*(2), 141-153.

Schlegel, K., Mehu, M., van Peer, J. M., & Scherer, K. R. (2018). Sense and sensibility: The role of cognitive and emotional intelligence in negotiation. *Journal of Research in Personality, 74*, 6-15.

Schmidt, A. F., & Finan, C. (2018). Linear regression and the normality assumption. *Journal of Clinical Epidemiology, 98*, 146-151.

Schoorman, F. D., Mayer, R. C., & Davis, J. H. (2007). An integrative model of organizational trust: Past, present, and future. *The Academy of Management Review, 32*(2), 344-354.

Segal, H. (1957). Notes on symbol formation. *International Journal of Psycho-Analysis, 38*, 391-397.

Sharp, C., Ha, C., Carbone, C., Kim, S., Perry, K., Williams, L., & Fonagy, P. (2013). Hypermentalizing in adolescent inpatients: treatment effects and association with borderline traits. *Journal of Personality Disorders, 27*(1), 3-18.

Sharp, C., Pane, H., Ha, C., Venta, A., Patel, A. B., Sturek, J., & Fonagy, P. (2011). Theory of mind and emotion regulation difficulties in adolescents with borderline traits. *Journal of the American Academy of Child & Adolescent Psychiatry, 50*(6), 563-573. e561.

Simon, H. A. (1955). A behavioral model of rational choice. *The Quarterly Journal of Economics, 69*(1), 99-118.

Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review, 63*(2), 129.

Simons, D. J. (2010). Monkeying around with the gorillas in our midst: familiarity with an inattentional-blindness task does not improve the detection of unexpected events. *I-perception, 1*(1), 3-6.

Simons, D. J., & Chabris, C. F. (1999). Gorillas in our midst: Sustained inattentional blindness for dynamic events. *Perception, 28*(9), 1059-1074.

Sinaceur, M. (2010). Suspending judgment to create value: Suspicion and trust in negotiation. *Journal of Experimental Social Psychology, 46*(3), 543-550. doi:10.1016/j.jesp.2009.11.002

Soliman, C. G., Stimec, A., & Antheaume, N. (2014). The long-term Impact of negotiation training and teaching implications. *Conflict Resolution Quarterly, 32*(2), 129-153. doi:10.1002/crq.21110

Solms, M. (2013). The conscious id. *Neuropsychoanalysis, 15*(1), 5-19.

Solms, M. (2015). *The feeling brain: Selected papers on neuropsychoanalysis*: Karnac Books.

Sperber, D., Clément, F., Heintz, C., Mascaro, O., Mercier, H., Origgi, G., & Wilson, D. (2010). Epistemic vigilance. *Mind & Language, 25*(4), 359-393. doi:10.1111/j.1468-0017.2010.01394.x

Sperber, D., & Wilson, D. (1986). *Relevance: Cognition and Communication*. Oxford: Blackwell.

Sporer, S. L., & Schwandt, B. (2007). Moderators of nonverbal indicators of deception: A meta-analytic synthesis. *Psychology, Public Policy, and Law, 13*(1), 1-34. doi:10.1037/1076-8971.13.1.1

Spunt, R. P., & Lieberman, M. D. (2013). The busy social brain. *Psychological Science, 24*(1), 80-86. doi:10.1177/0956797612450884

Stagnaro, M., Pennycook, G., & Rand, D. G. (2018). Performance on the Cognitive Reflection Test is stable across time. *Judgment and Decision Making, 13*, 260-267.

Stanovich, K. E. (2011). *Rationality and the reflective mind*. Oxford: Oxford University Press.

Stanovich, K. E., West, R. F., & Toplak, M. E. (2016). *The Rationality Quotient: Toward a Test of Rational Thinking*. Cambridge, MA: MIT press.

Steele, H., & Steele, M. (2008). *Clinical applications of the adult attachment interview*: Guilford Press.

Steiner, J. (1992). The equilibrium between the paranoid-schizoid and the depressive positions. In *Clinical lectures on Klein and Bion* (pp. 46-58): London and New York: Tavistock/Routledge.

Stern, D. N. (2007). *The interpersonal world of the infant: a view from psychoanalysis and developmental psychology*. California: Psychoanalytic Electronic Publishing.

Street, S. E., Navarrete, A. F., Reader, S. M., & Laland, K. N. (2017). Coevolution of cultural intelligence, extended life history, sociality, and brain size in primates. *Proceedings of the National Academy of Sciences, 114*(30), 7908-7914.

Tajima, M., & Fraser, N. M. (2001). Logrolling procedure for multi-issue negotiation. *10*(3), 217-235.

Target, M., & Fonagy, P. (1996). Playing with reality: II. The development of psychic reality from a theoretical perspective. *The International journal of psycho-analysis, 77 (Pt 3)*, 459.

Taylor, S. E., & Brown, J. D. (1988). Illusion and well-being: a social psychological perspective on mental health. *Psychological Bulletin, 103*(2), 193.

Ten Brinke, L., Black, P. J., Porter, S., & Carney, D. R. (2015). Psychopathic personality traits predict competitive wins and cooperative losses in negotiation. *Personality and Individual Differences, 79*, 116-122. doi:10.1016/j.paid.2015.02.001

Ten Velden, F. S., Beersma, B., & De Dreu, C. K. W. (2010). It Takes One to Tango: The Effects of Dyads' Epistemic Motivation Composition in Negotiation. *Personality and Social Psychology Bulletin, 36*(11), 1454-1466. doi:10.1177/0146167210383698

Thompson, L. (1990a). An examination of naive and experienced negotiators. *Journal of Personality and Social Psychology, 59*(1), 82-90. doi:10.1037/0022-3514.59.1.82

Thompson, L. (1990b). Negotiation behavior and outcomes: empirical evidence and theoretical issues. *Psychological Bulletin, 108*(Nov 90), 515-532.

Thompson, L. (1991). Information exchange in negotiation. *Journal of Experimental Social Psychology, 27*(2), 161-179. doi:10.1016/0022-1031(91)90020-7

Thompson, L. (2005). *The Mind and Heart of the Negotiator* (3 International ed.). Upper Saddle River, N.J.: Pearson/Prentice Hall.

Thompson, L., & Deharpport, T. (1994). Social judgment, feedback, and interpersonal learning in negotiation. *Organizational Behavior and Human Decision Processes, 58*(3), 327-345. doi:10.1006/obhd.1994.1040

Thompson, L., & Deharpport, T. (1998). Relationships, goal incompatibility, and communal orientation in negotiations. *Basic and Applied Social Psychology, 20*(1), 33-44. doi:10.1207/s15324834basp2001_4

Thompson, L., & Hastie, R. (1990). Social perception in negotiation. *Organizational Behavior and Human Decision Processes, 47*(1), 98-123. doi:10.1016/0749-5978(90)90048-E

Thompson, L., & Hrebec, D. (1996). Lose–lose agreements in interdependent decision making. *Psychological Bulletin, 120*(3), 396-409. doi:10.1037/0033-2909.120.3.396

Thomson, K. S., & Oppenheimer, D. M. (2016). Investigating an alternate form of the cognitive reflection test. *Judgment and Decision Making, 11*(1), 99.

Tomasello, M. (2006). Why don't apes point? In N. J. Enfield & S. C. Levinson (Eds.), *Roots of Human Sociality: Culture, cognition and interaction* (Vol. 197, pp. 506-524). Oxford: Berg.

Tomasello, M. (2008). *Origins of human communication*. Cambridge, Mass.: MIT Press.

Tomasello, M. (2012). Why be nice? Better not think about it. *Trends in Cognitive Sciences, 16*(12), 580-581.

Tomasello, M. (2014). *A natural history of human thinking*. Cambridge, Mass.: Harvard University Press.

Tomasello, M. (2016). *A natural history of human morality*: Harvard University Press.

Tomasello, M. (2019). *Becoming human: A theory of ontogeny*. England: Harvard University.

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *The Behavioral and brain sciences, 28*(5), 675.

Tomasello, M., Melis, A. P., Tennie, C., Wyman, E., Herrmann, E., Gilby, I. C., . . . Tomasello, M. (2012). Two key steps in the evolution of human cooperation: The interdependence hypothesis. *Current anthropology, 53*(6).

Toplak, M., West, R. F., & Stanovich, K. E. (2011). The Cognitive Reflection Test as a predictor of performance on heuristics-and-biases tasks. *Mem Cogn, 39*(7), 1275-1289. doi:10.3758/s13421-011-0104-1

Toplak, M., West, R. F., & Stanovich, K. E. (2014). Assessing miserly information processing: An expansion of the Cognitive Reflection Test. *Thinking and Reasoning, 20*(2), 147-168.

Trötschel, R., Hüffmeier, J., Loschelder, D. D., Schwartz, K., & Gollwitzer, P. M. (2011). Perspective Taking as a Means to Overcome Motivational Barriers in Negotiations: When Putting Oneself Into the Opponent&#039;s Shoes Helps to Walk Toward Agreements. *Journal of Personality and Social Psychology, 101*(4), 771-790. doi:10.1037/a0023801

Tsoi, L., Dungan, J., Waytz, A., & Young, L. (2016). Distinct neural patterns of social cognition for cooperation versus competition. *Neuroimage, 137*, 86-96. doi:10.1016/j.neuroimage.2016.04.069

Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. *Science, 185*(4157), 1124-1131.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science (New York, N.Y.), 211*(4481), 453.

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *J Risk Uncertainty, 5*(4), 297-323. doi:10.1007/BF00122574

Valley, K., Neale, M., & Mannix, E. (1995). Relationships in negotiations: The role of reputation, the shadow of the future, and interpersonal knowledge on the process and outcome of negotiations. *Research in bargaining negotiation in organizations*, 65-93.

Van Boven, L., & Thompson, L. (2003). A look into the mind of the negotiator: Mental models in negotiation. In (Vol. 6, pp. 387-404).

Van Overwalle, F. (2011). A dissociation between social mentalizing and general reasoning. *Neuroimage, 54*(2), 1589-1599. doi:10.1016/j.neuroimage.2010.09.043

Van Overwalle, F., & Evandekerckhove, M. (2013). Implicit and explicit social mentalizing: Dual processes driven by a shared neural network. *Frontiers in human neuroscience, 7*. doi:10.3389/fnhum.2013.00560 van Schaik, C. P., & Burkart, J. M. (2011). Social learning and evolution: the cultural intelligence hypothesis. *Philosophical Transactions of the Royal Society of London B: Biological Sciences, 366*(1567), 1008-1016.

Vecchi, G. M., Wong, G. K., Wong, P. W., & Markey, M. A. (2019). Negotiating in the skies of Hong Kong: The efficacy of the Behavioral Influence Stairway Model (BISM) in suicidal crisis situations. *Aggression and violent behavior, 48*, 230-239.

Vorauer, J. D., & Miller, D. T. (1997). Failure to recognize the effect of implicit social influence on the presentation of self. *Journal of Personality and Social Psychology, 73*(2), 281.

Vroom, V. H. (1964). Work and motivation.

Vygotsky, L. S. (1978). *Mind in society: The Development of Higher Psychological Processes*. Cambridge MA: Harvard university press.

Waal, F. B. M. d. (1989). *Chimpanzee politics: power and sex among apes*. London: Johns Hopkins University Press.

Walton, R. E., & McKersie, R. B. (1965). *A behavioral theory of labor negotiations: An analysis of a social interaction system*. New York: McGraw-Hill.

Ward, A., Ross, L., Reed, E., Turiel, E., & Brown, T. (1997). Naive realism in everyday life: Implications for social conflict and misunderstanding. *Values and knowledge*, 103-135.

Warneken, F., & Tomasello, M. (2006). Altruistic helping in human infants and young chimpanzees. *Science (New York, N.Y.), 311*(5765), 1301. doi:10.1126/science.1121448

Weber, J. M., Kopelman, S., & Messick, D. M. (2004). A conceptual review of decision making in social dilemmas: Applying a logic of appropriateness. *Personality and Social Psychology Review, 8*(3), 281-307.

Weerd, H., Verbrugge, R., & Verheij, B. (2017). Negotiating with other minds: the role of recursive theory of mind in negotiation with incomplete information. *Auton Agent Multi-Agent Syst, 31*(2), 250-287. doi:10.1007/s10458-015-9317-1

Weingart, L. R., Brett, J. M., Olekalns, M., & Smith, P. L. (2007). Conflicting social motives in negotiating groups. *Journal of Personality and Social Psychology, 93*(6), 994.

Wertz, A. E., & German, T. C. (2007). Belief-Desire Reasoning in the Explanation of Behavior: Do Actions Speak Louder than Words? *Cognition, 105*(1), 184-194. doi:10.1016/j.cognition.2006.08.002

West, R. F. (2011). A Taxonomy of Rational Thinking Problems. In K. E. Stanovich (Ed.), *Rationality and the reflective mind*. Oxford: Oxford University Press.

Whittlesea, B. W., & Williams, L. D. (2001). The discrepancy-attribution hypothesis: II. Expectation, uncertainty, surprise, and feelings of familiarity. *Journal of Experimental Psychology: Learning, Memory, Cognition, 27*(1), 14.

Winnicott, D. (1960). The theory of the parent-infant relationship. *International Journal of Psycho-Analysis, 41*, 585-595.

Word, C. O., Zanna, M. P., & Cooper, J. (1974). The nonverbal mediation of self-fulfilling prophecies in interracial interaction. *Journal of Experimental Social Psychology, 10*(2), 109-120.

Yao, J., Zhang, Z. X., & Brett, J. M. (2017). Understanding trust development in negotiations: An interdependent approach. *Journal of Organizational Behavior, 38*(5), 712-729.

Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology, 18*(5), 459-482. doi:10.1002/cne.920180503

Yip, J. A., & Schweitzer, M. E. (2015). Trust promotes unethical behavior: Excessive trust, opportunistic exploitation, and strategic exploitation. *Current Opinion in Psychology, 6*, 216-220.

Yip, J. A., & Schweitzer, M. E. (2016). Mad and misleading: Incidental anger promotes deception. *Organizational Behavior and Human Decision Processes, 137*, 207-217.

Zeelenberg, M. (1999). Anticipated regret, expected feedback and behavioral decision making. *Journal of Behavioral Decision Making*, *12*(2), 93-106.

Zhang, J.-D., Liu, L. A., & Liu, W. (2015). Trust and deception in negotiation: Culturally divergent effects. *Management Organization Review*, *11*(1), 123-144.

Zhong, C.-B., Loewenstein, J., & Murnighan, J. K. (2007). Speaking the same language: The cooperative effects of labeling in the prisoner's dilemma. *Journal of Conflict Resolution*, *51*(3), 431-456.