New Bulgarian University

Department of Cognitive Science and Psychology

Analogies and Understanding Intentions

Luiza Shahbazyan

(Extended resume)

An extended resume of a thesis submitted to

New Bulgarian University

in partial fulfilment

of the requirements for the degree of

DOCTOR OF PSYCHOLOGY

Supervisor: Associate Professor Lilia Gurova

Scientific adviser: Assistant Professor Georgi Petkov

Sofia, 2015

ABSTRACT

Mindreading is the hallmark of humans' social cognition. However, the processes of understanding others' mental states have not been yet systematically explored. By linking theoretical and empirical findings about analogy making and understanding intentions, the current work proposes a novel hypothesis for the mechanisms underlying understanding intentions in ambiguous situations. This hypothesis states that people are able to spontaneously use concrete episodes to infer the intentions of a target agent in a novel, but structurally similar situation. In support of this hypothesis, the results of three experiments (Experiments 1, 2 and 5) show that the participants are more likely to attribute a negative intention to the actor in an ambiguous situation if the latter is preceded by a negative, structurally similar episode ("analog"). However, structural similarity tends to interact with activated stereotypes as the change of the characters in the base story from negative to positive attenuated the effect of the negative analog (Experiment 3 and 4) and enhanced the effect of the positive analog (Experiment 3). Furthermore, across three experiments (Experiment 2, 4 and 5) it is demonstrated that, under certain conditions, participants are more likely to attribute a negative intention to the actor in an ambiguous situation if the latter is preceded by a positive analog. This phenomenon, which we termed inverted effect, is suggested to occur as a result of a failure to evaluate the analogy between the analog and the target story as useful due to their alienable differences. Taken together, the findings of the five studies provide moderate support for the proposal that analogy processing plays an important role in understanding intentions and suggest that analogical inference might be the looked for unitary mechanism which underlies the use of both general knowledge and concrete episodes in understanding intentions in ambiguous situations.

TABLE OF CONTENT

CHAPTER 1. Introduction: Analogies and Understanding Intentions

Humans are mindreaders. We live in a world of others' minds and interact with friends, parents, co-workers, acquaintances, and even complete strangers, attributing constantly mental states such as beliefs, desires and intentions. Any deficits in this ability create serious consequences explaining the social and communication impairments of individuals with autism (Frith, 2001; Baron-Cohen, 1995), schizophrenia (Corcoran & Frith, 2003) and chronic aggression (Dodge, 2006; Orobio de Castro, Veerman, Koops, Bosch, & Monshouwer, 2002).

The current work will focus on understanding how people ascribe content to the intentions behinds others' actions. An intention can be defined as the "plan of action the organism chooses and commits itself to in pursuit of a goal" (Tomasello, Carpenter, Call, Behne, & Moll, 2005, p. 276).

It has been established in social psychology for some time that people understand the intentions of others by making attributions and these attributions are not properties of the actual behaviour, but rather come from the perceiver's own mind (Heider, 1958). We possess the capacity to infer the *content* of the mental states of others although they are not directly observable. Consider the following citation taken from an online forum:

"I was walking down the road and I dropped a paper to the ground. Someone who was passing by me picked it up for me. But then weirdly, when he gave it back to me he **winked** at me at the same time. I just said thanks and moved on. IS this a normal part of American culture?"

One way to infer the intention of the person who winked is to pay attention to his body language and facial expressions, so that we can directly perceive what he meant. Although such information can provide hints, there are significant individual (Cohn, Schmidt, Gross, & Ekman, 2002) as well as cultural differences (Jack, Garrod, Yu, Caldara & Schyns, 2012) in

the way people express themselves, which may cloud our judgments. Another approach is to try to understand others by pretending we are in the same situation and simulating what they would do. However, this could be very difficult for a person of other culture for who is unthinkable to behave in such a way. Still another approach is to use generalized knowledge such as schemas or stereotypes about Americans in order to infer what the intention of the stranger we met in New York are. But stereotypes contain a wealth of information – Americans are friendly, individualistic, or promiscuous. Some of them might be relevant to the situation, and some might not. So, there must be some way in which we choose some intentional interpretations over others.

The three approaches just described refer to the major theoretical schools that have addressed the question of how people understand the intentions of others: direct perception theories (e.g. Gallagher 2008; Newtson, 1973; Newtson & Engquist 1976), Simulation theories (ST; e.g. Harris, 1992; Goldman, 2006; Gordon, 2008), and Theory Theories (TT; e.g. Gopnik & Wellman, 1994; Nichols & Stich, 2003; Perner, 1991), respectively. They make conflicting claims about the kind of information that people need in order to understand the intentions of others.

However, a fourth type of knowledge – relational, has received little attention in literature on understanding intentions. Relational representations can be defined as "a binding between a relation symbol and a set of ordered tuples of elements" (Halford, Willson & Phillips, 2010, p.497). There is evidence that action perception involves coding motor information bound by relational structures (Zacks & Tversky, 2001; Blythe, Todd, Peter, Miller, & Geoffrey, 1999) in line with the idea of Penn & Povellini (2007) that people re-represent perceptual information in terms of higher order relations. Furthermore, episodic memory involves relational representations that bind the experienced objects and their attributes in a meaningful way (Ofen & Shing, 2013). In a similar vein, stereotypes are more

than a bundle of attributes; they are hold together by relational structures including causal assumptions (Wittenbrink, Gist & Hilton, 1997; Halford et al., 2010). Taken together, these findings suggest that relational knowledge is intrinsic to the information that is supposedly used to generate inferences about others' intentions.

This raises the question to what extend similarity in the relational structures of the current (target situation) and some prior source knowledge (perceptual, semantic or episodic) determines the inferences that are going to be generated. The process of finding relational similarity between two or more representations is known as "analogy making" and there is robust empirical support for its role in humans' ability to generate inferences about unknown aspects of the world.

So, the aim of this thesis is to explore the role which analogy making plays in understanding others' intentions in ambiguous situations.To the best of our knowledge, the current work provides the first systematic investigation of the phenomenon that addresses its role in understanding intentions both theoretically and empirically.

CHAPTER 2. Major Approaches to Understanding Mental States

People have the striking ability to understand and predict others' actions. The problem is that mental states such as intentions are to a large extend unobservable, "lying somewhere between perceptual inputs and behavioural outputs", opening a gap between our minds and the minds of others (Hutto et al., 2011, p. 390). A number of approaches have been put forward to account how we bridge this gap and readily perceive unobservable states.

2.1 Direct perception theories

The proponents of the so-called direct perception theories (Gallagher 2008; Newtson, 1973; Newtson and Engquist 1976), argue that we directly perceive states such as intentions, desires and beliefs without any additional processing. Such information can be derived from

the movements of her body (Meltzoff & Brooks, 2001) or the expression of her face (Ames & Johar, 2009).

While under limited circumstances (see Becchio, Cavallo, Begliomini, Sartori, Feltrin, & Castiello, 2012, for a review) sensory data seem to unambiguously reveal the intentions behind the action, but direct perception is not able to account for the complex and ill-constraint inferences that we generate every day. Activation of the mirror neurons (Iacoboni, 2009) does not seem to confirm this approach because while it was demonstrated that the mirror system is active during action perception, it does not seem that it is primary responsible for the attribution of mental states. Additionally, experimental evidence (Newtson, 1973; Zachs & Tverski, 2001) supports the idea that the available kinetic information is re-represented in terms of higher order relational structures even though such a re-representation is not necessary for the immediate comprehension. However, it is stored, so that later can be used in understanding and planning, when the available information is scarce.

2.2 Theory Theory

The idea that mentalizing is dependent on prior knowledge is in the heart of the probably most popular approach to this phenomenon – Theory Theory (TT). According to Stich and Nichols (2003), understanding of mental states depends on "a rich set of mental representations containing substantial amounts of information (or, sometimes, misinformation) about mental states and their interactions with environmental stimuli, with behaviour, and with each other." (p.102).

Evidence for the role of generalized knowledge in the generation of inferences comes from two lines of research. First, findings from social and developmental research have shown that past experience of peer victimization (Yeung & Leadbeater, 2007) and abuse during childhood (Dodge, Bates, & Pettit, 1990) are associated with Hostility Attribution Bias (HAB), i.e. a tendency to interpret even accidental behaviours as hostile ones. Second,

ctivation of schemas through priming using emotionally or semantically similar material was further used to demonstrate the causal role of generalized knowledge structures as social schemas and stereotypes on intentional attribution (Orobio de Castro, Slot, Bosch, Koops and Veerman, 2003; Duncan, 1976; DeVine, 1989).

One shortcoming of generalized knowledge must be learned from repetitive experience of concrete episodes (Tulving, 1972), but there are cases in which this process is too slow. This issue has been addressed by the next theoretical perspective that will be reviewed.

2.3 Simulation theory

A third solution to the problem of others' minds has been proposed by Simulation Theory (ST). Its basic tenet is that perceivers use their own cognitive system to pretend that they are in the situation of the other person, so that they can simulate what the other person would do. The output of the simulation is assigned to the other person (Shanton & Goldman, 2010; Gordon, 2006). We still need to have knowledge about the possible intentions of the other person (Doherty, 2009), but this knowledge does not need to be generalized.

If people use generalized theories to infer the intentions of other's behaviour, one would not expect any influence of information such as the mental state of the perceiver or his contextualized knowledge about the target situation. Nevertheless, such influence has been extensively documented (Keysar, 1994, Birch & Bloom, 2007, Lewicki, 19850

However, this approach limits mindreading only to situations, which are identical or at least very similar to our previous experiences, which is a significant and not a very realistic constraint.

2.4 Hybrids between TT and ST

In the recent years, a number of hybrid theories emerged proposing that some aspects of mentalizing are served by simulations, while others are served by theorizing (see Bach, 2011 for a review).

However, these theories face several challenges. First, by proposing that people use both theorizing and simulation, the hybrid models inherit some of the limitations of each of the individual mechanisms that have been already outlined above, including ignoring the role of relational knowledge to derive inferences. Second, hybrid models should specify the conditions under which each of the mechanism is employed, which is problematic and leads to conflicting claims or claims that do not lead to testable predictions (Apperly, 2008).

Third, the hybrid models imply that people selectively use either episodic or semantic memory despite evidence that the two systems interact (Barsalou, 2009; Moscovitch, 2008). The hybrid models do not address the possibility for interaction.

We argue that a way to forward the current debate beyond the three major paradigms and their hybrid models is to reconsider the mechanisms that underlie inference generation. Towards this aim, first, we identified a significant gap of knowledge concerning the role of relational processing in generating inferences about others' intentions as well as relevant, though scarce evidence that supports the necessity to bridge that gap. In the next chapters, the research on analogy making will be introduced and linked to the research on understanding intentions to highlight further their relevance.

CHAPTER 3. Relational Knowledge and Analogical Inferences

The ability to acquire and manipulate relational knowledge has been considered as an answer to the question "What makes humans so smart?" (Gentner, 2003; Penn & Povinelli, 2012; Penn, et al., 2008). Why is that? First of all, we are able to re-present the sensory stream

in terms of non-obvious relations such as cause, prevent, execute, promise and also want, feel and intend. Relations allow us to express ideas such as causations, chains of implications, counterfactuals and others, which will not be possible otherwise (Gentner, 2003). Furthermore, they enable us to think analytically beyond what is given in the situation (Halford et al., 2010). But probably the most important characteristic of relational knowledge is the role it plays in analogical inference – inferring unknown properties of the current situation based on relational similarity to prior episode or schema (Markman & Laux, 2008; Gentner & Smith, 2013; Holyoak, 2005).

The procccess of mapping is at the core of analogical processing and it involves the alignment of the common structure between two situations – usually some base situation we know about and a target situation that is less familiar. The corresponding objects and their attributes need not to be superficially similar, but they need to hold similar roles in the matching systems of relations, e.g. to be structurally consistent. If the mapping cannot be accomplished, the base situation is discarded as irrelevant to the understanding of the target situation. For instance, when comparing the solar system to the atom structure, people discard the identity of the matching objects (atoms and planets) and focus on the common structure in the two systems (Gentner, 1983).

Once two situations are mapped, a corresponding part of knowledge is transferred from the base to the target – the so-called analogical inference. Not all knowledge is transferred but only that information that is part of corresponding structures, preferable higher-order relational structures, to ensure that deeply embedded structural knowledge is transferred rather than a random bundle of facts. For instance, children may think that the Earth is flat like a pancake. However, although they may transfer the shape of the pancake to infer an observable property of the Earth, it is highly unlikely that they will also think that the Earth is sweet and goes well with maple syrup for breakfast (Markman & Laux, 2008).

After the structural alignment between two analogs has been found and the inferences projected, both the analogy and its inferences are evaluated (Gentner & Smith, 2013).

Analogical processing is particularly useful in dealing with complex, novel or ill-structured situations, because it allows constraining the information that is considered relevant and inferring causal chains or chains of implications (Gick & Holyoak, 1980). Due to these characteristics, it has been valuable tool for scientific discovery and exploration (Nersessian, 1988). Analogical inferences are also frequent in the domain of problem solving where a solution to a problem is inferred from the base and applied in the target situation (Gick & Holyoak, 1980; 1983; Holyoak & Koh, 1987).

There a number of experiments that demonstrated that people may use analogy spontaneously to infer unknown aspects of a situation or a solution to a problem based on previous example (Kokinov and Yoveva (1996; Schunn & Dunbar, 1996). This spontaneous use is also demonstrated when people try to comprehend ambiguous sentences (Day and Gentner, 2007; Popov & Hristova, 2014).

Evidence for the use of analogical inferences in making sense of everyday situations was also obtained in the realm of social cognition in the area of moral dillemas (Dehghani, Gentner, Forbus, Ekhtiari, & Sachdeva, 2009) and social judgment (Mussweiler & Gentner, 2007), but the research in this field is scarce and the intersections between analogy and social cognition research programs are rare.

CHAPER 4. Linking Analogy Making and Understanding Mental States

Analogy making research mainly focuses on non-social domains though there is no reason to restrain its application to them. For instance, Penn and Povinelli (2007; 2009; 2011) have repeatedly made the specific connection between the sophisticated mentalizing abilities of humans and our ability to represent and manipulate relational structures in analogical

processing. Similar proposals have been put forward by Bach (2009) and Baldwin & Baird (2001). However, the question of the role of relational knowledge in understanding intentions, in particular, has not been systematically surveyed, so that testable predictions about inference generation in adults can be formulated. In order to do so, we will survey evidence that there are parallels between analogy making and understanding intentions.

Defining mental states as propositional attitudes underscores its similarity to role-based relational reasoning. More specifically, propositional attitudes are statements that contain an agent and thought related by a particular attitude (Morton, 2009), such as I (agent) intend (attitude) to visit Sofia (thought). People can flexibly change the content of the propositional statement and the attitude towards the proposition. As such, it seems that understanding intentions is very much like role-based relational reasoning - differentiating relational roles from the entities that fill those roles, which is observed in analogy making (Holyoak, 2012).

The second parallel between analogy making and understanding intentions is that both processes involve generating inferences based on the perceived similarity between a target situation and preexisting knowledge. Similarly to analogy making, understanding intentions seems to depend on activation of relevant past knowledge, either concrete or generalized depending on the theoretical school. One reason to expect relational information to play a role is to consider the issue of the quality of the inferences that will be generated in a given situation (Jameson & Gentner, 2009). When trying to come up with ideas about the intentions that underlie a particular action, some intentions may provide better explanation than others and this depends on systematicity – the degree to which this concept or idea is part of a system of interrelated relations (Gentner, 1983; Jameson & Gentner, 2009).

Besides similarity in the processes involved in analogy making and understanding intentions, the two domains also seem to share common brain circuits (Burgess, Dumontheil,

& Gilbert, 2007; Kroger, Sabb, Fales, Bookheimer, Cohen & Holyoak, 2002; Ramnani and Owen, 2004; but see Van Overwalle, 2011). There is also patient data that specifically relates analogical reasoning and mentalizing (Krawczyk, Kandalaft, Dibehbani, Allen, McClelland, Tamminga & Chapman, 2014), where significant correlations between social cognitive disability and analogical reasoning disability have been found, especially for schizophrenia patients.

The best explanation of these similarities between understanding intentions and analogy making is that the two processes share a common underlying mechanism. As argued above, understanding mental states by definition involves manipulation of relational knowledge such as attitudes that relate agents and the propositional content of these attitudes. There is robust evidence from studies of other cognitive tasks that using relational knowledge contributes to the quality of the generated inferences and allows transfer of prior knowledge even from dissimilar situations given that they share a common structure with the novel episode. These aspects are undeniably important in mindreading. However, each of the three major accounts of understanding others' minds has difficulties explaining the mechanisms that allow flexible use of concrete and abstract knowledge in such a way.

We suggest that the mechanisms of analogy making provides a unitary account of generating inferences about the mental states of others, which is a simpler more parsimonious explanation of understanding others' intentions. However, the ability of people to use structurally similar episodes ("analogs") to understand the intentions of others has not been systematically explored yet.

The aim of the 5 studies that follow is to explore the role which analogy making plays in understanding others' intentions in ambiguous situations. We devised five studies in which systematically vary the base information available to the participants. Based on the suggested role that analogy making plays in transfers from both concrete and generalized knowledge, we

hypothesized that participants will be able to transfer intentions from structurally similar, superficially dissimilar episodes. Furthermore, if structural systematicity is a necessary constraint during generating inferences, we hypothesize that participants will be more likely to transfer intentional information from episode that shares relational information, rather than from episode with which it shares common objects or attributes such as mood valence.

CHAPTER 5. Experimental Studies

5.1 Experiment 1

In designing this study, our goal was to create a simplified laboratory analog of the kind of challenge that people face when trying to infer others' intentions behind an action that can be interpreted as both well and ill-intended and there is not enough information to disambiguate it. The aim of experiment 1 is to test the hypothesis that single episodes which are structurally similar to an ambiguous target situation will influence the attribution of intention to the ambiguous action in the target. The prediction is in line with the idea that if the target situation is ambiguous, relevant relational prior concrete knowledge will facilitate the representational process by 'filling-in' missing information. This prediction uniquely follows from the view that relational knowledge and analogical processing play a role in intentional understanding.

To test the predicted role of relational knowledge and disentangle its effect from the effect of alternative mechanisms such as superficial similarity and mood congruency, we presented seven groups of participants each with different base story and one and the same target story, which contains ambiguous action. Two of the base stories (the structurally similar stories or "analogs") contained either positive or negative intention. The latter was embedded in a relational structure, which was also present in the target situation. The two superficially similar base stories contained either positive or negative intention but it was embedded in a

relational structure that was dissimilar to the target situation. However, they shared identical objects and attributes with the target. Two of the stories (no similarity) contained either positive (pleasant surprise) or negative (death) event in order to communicate the respective mood, but without sharing neither common objects, nor common relations with the target, and without containing information about any positive or negative intention. The seventh story, a control one, was mood neutral and does not contain any common items with the target nor information about intentions.

This yields six experimental conditions that explored the effect of two factors: BASE CONTENT (positive vs. negative) and BASE SIMILARITY (structural, superficial and no similarity) as well as a seventh condition, which contained neutral information and serves as a control condition

The ambiguous target story was followed by three different interpretations of the ambiguous action that disambiguate it. The ratings of the quality of each TARGET INTENTION provided were the main dependent measures of the experiment.

We hypothesized that participants will rate significantly higher the intention that is consistent to the one embedded in the structurally similar base episode in comparison to the intention that is not embedded in a structurally similar base episode. More specifically, we expect an interaction between BASE CONTENT and BASE SIMILARITY due to significant difference between the ratings of the TARGET INTENTIONS (positive or negative) consistent with the content of the structurally similar base episodes and the rest of the base episodes.

The design was between subject and participants were randomly assigned to each of the seven conditions. The experiment was disguised as a study on memory processes and participants were told that they will read a story and later recall it (this was the base story), but

in between they will be asked to rate stimuli for another study as a distractor task (this was the

target story).

In agreement with the hypothesis of the study, there was a significant interaction

between SIMILARITY and BASE CONTENT over the ratings of the NEGATIVE TARGET

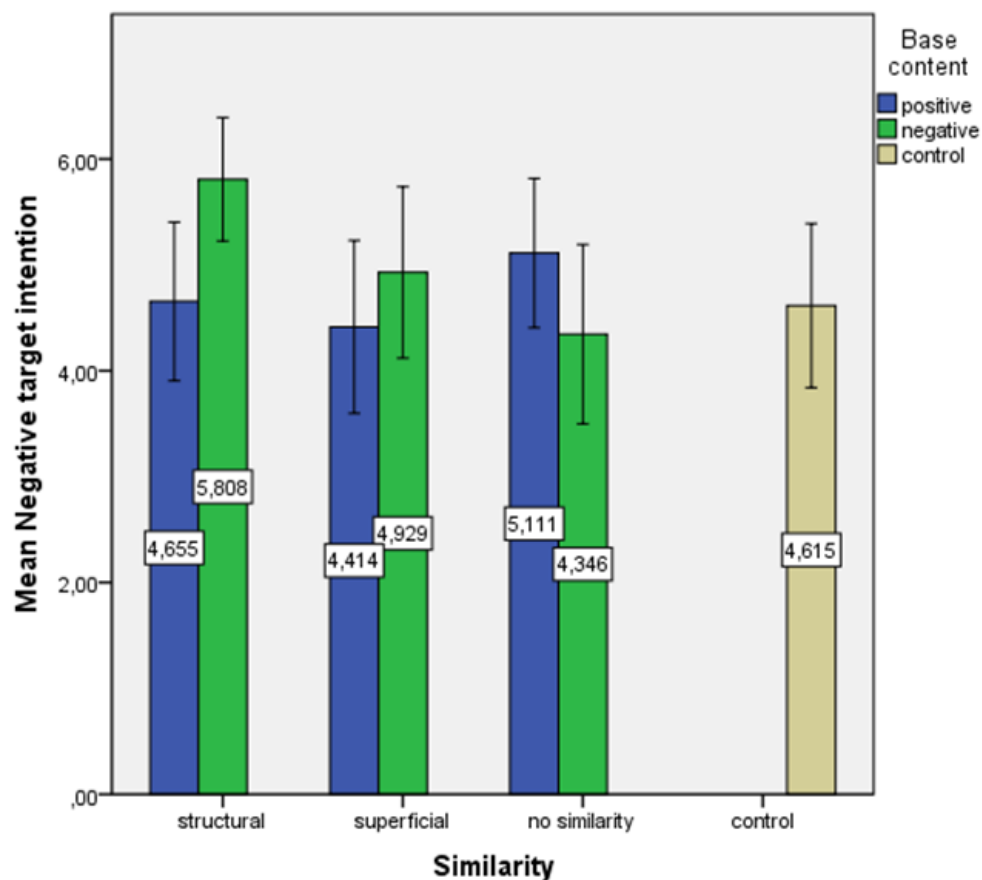INTENTION, $F_{(2,184)} = 3,407$, $p=0.035$. The interaction is plotted in the Figure 1.

Figure 1 Mean ratings of the negative target intention across the seven conditions

There was no significant interaction between BASE SIMILARITY and BASE

CONTENT over the ratings of the POSITIVE TARGET INTENTION, $F_{(2,184)} = 1.533$,

$p=0.219$. According to the post-hoc test, there was a difference between the structural and the

no similarity group ($p = 0.007$), the structural and the control group ($p=0.027$) and a marginal

difference between the structural and the superficial similarity group ($p=0.094$), whereas the

superficial similarity group did not differ from the no similarity and the control group (see Figure 1), confirming that negative structurally similar episode led to significant increase of the ratings of the negative intention.

No further effects yield any significance.

Overall, this pattern of results confirms that presenting a single negative structurally similar base story can affect the intentional inferences that people make thus altering the perceived explanatory quality of some intentions over others.

However, there was a notable asymmetry in the influence of the base stories because no such effect was observed in the positive BASE CONTENT condition. One explanation is that participants encoded the negative stories better than the positive, but it has been ruled out by our data on memory recall – no differences in the memory scores were obtained between the positive and the negative conditions. Another explanation is that in the negative BASE CONTENT condition the wolves in the base story may have activated the stereotype of wolves and, as a consequence, activated some relational schemas relevant to how wolves have bad intentions towards the other animals, which might have aided the encoding of the common structure between the two stories, so that the alignment is completed and the analogy between the two stories is made (Gick & Holyoak, 1983). As a result, analogical inference is generated and the base consistent intention is rated higher. This hypothesis will be further explored in Experiments 3 and 4.

5.2 Experiment 2

In experiment 2 we aim to further explore the role of structural similarity by two amendments to the procedure. First, given that a number of studies have established correlation between aggression and biased negative attributions (Matthews & Norris, 2002; Dodge, 2006), a measure of aggression was added to the experimental tasks in order to control for any potential role of aggression in creating bias towards negative intentions that is not

induced by the experimental manipulation. Second, in attempt to find a more sensitive measure of the effects of the structural similarity, a forced choice response instead of scale ratings was provided, because it was found to assess higher level processing of the information rather than operation of learnt schemas (Smyth et al., 2006; DeSteno et al., 2002; but see Russell, 1993). Thus we expect that, if the lack of effect of the positive ending story in Experiment 1 is due to impaired analysis of the relational structure of the stories, the forced choice response format by invoking deeper processing may increase the likelihood that participants will encode the common relations and generate the inference to produce base-consistent preferences.

As Experiment 1 already established the role of structural similarity in generating inferences from a base episode to a target ambiguous situation, at least when the intention is negative, in Experiment 2 we presented participants only with three stories: positive structurally similar ("positive analog"), negative structurally similar ("negative analog" [1]) and control condition. The specific hypothesis of the study is that participants will choose the base consistent TARGET INTENTION more often than the base inconsistent TARGET INTENTION in comparison to a control condition. Furthermore, it was hypothesized that this effect will be more robust among participants with higher aggression.

To test the prediction of the study that presenting an analog will increase the preferences of the consistent TARGET INTENTION, the frequencies of the preferred intentional attribution (see Figure 2) were submitted to Chi square analysis. There was a significant relationship between BASE CONTENT and choosing a particular TARGET INTENTION, X^2 (2, n=75) =7.852, p=.020. Further comparison between the negative structurally similar conditions and the control group showed that seeing a negative

---

[1] In order to improve readability of the text, the structurally similar episodes will be referred as analogs, where appropriate.

structurally similar base is associated with more negative attributions, X^2 (1, n=47)=7.743, p=.005.

The results of Experiment 2 replicated the effect of negative structurally similar base on increasing the preference towards negative intention. No effect of the positive structurally similar story on increase of preference towards the positive base-consistent TARGET INTENTION was observed. Quite the opposite, there is a tendency participants to prefer the base-inconsistent negative intention after seen the positive, structurally similar story.

One possible explanation of these findings is that the identity of the characters in the base episode (wolves) produced the preference toward negative intention in both the positive and the negative similar conditions, by activating a negative stereotype and the result is not
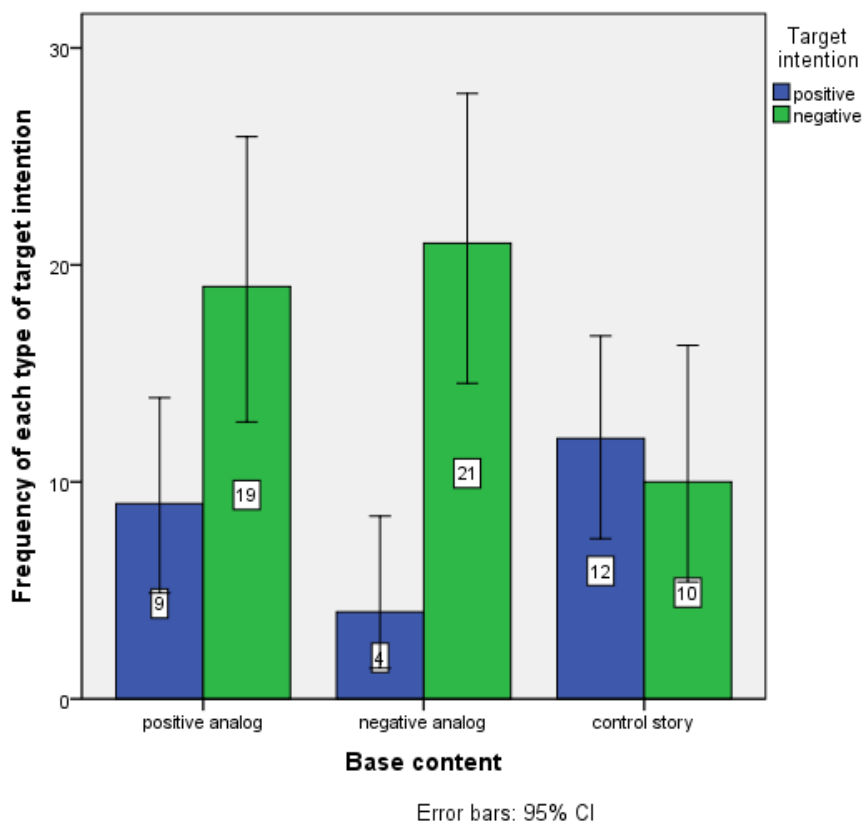
Figure 2 Frequency of preferred TARGET INTENTION as a function of BASE CONTENT (negative characters)

produced by any influence of the relational information, which will be explored in Experiment 3 and 4.

Another ecplanation may be that that participants made the analogy but for some reason evaluate it as inappropriate, which determined their preference for the alternative option – the negative intention.This hypothesis will be explored further in Experiment 5 by using experimental procedure to enhance analogical inferences to test if the inverted effect will remain.

Despite the documented relation between aggressiveness and preference towards attributing hostile intentions in ambiguous situation, our study failed to demonstrate such association.
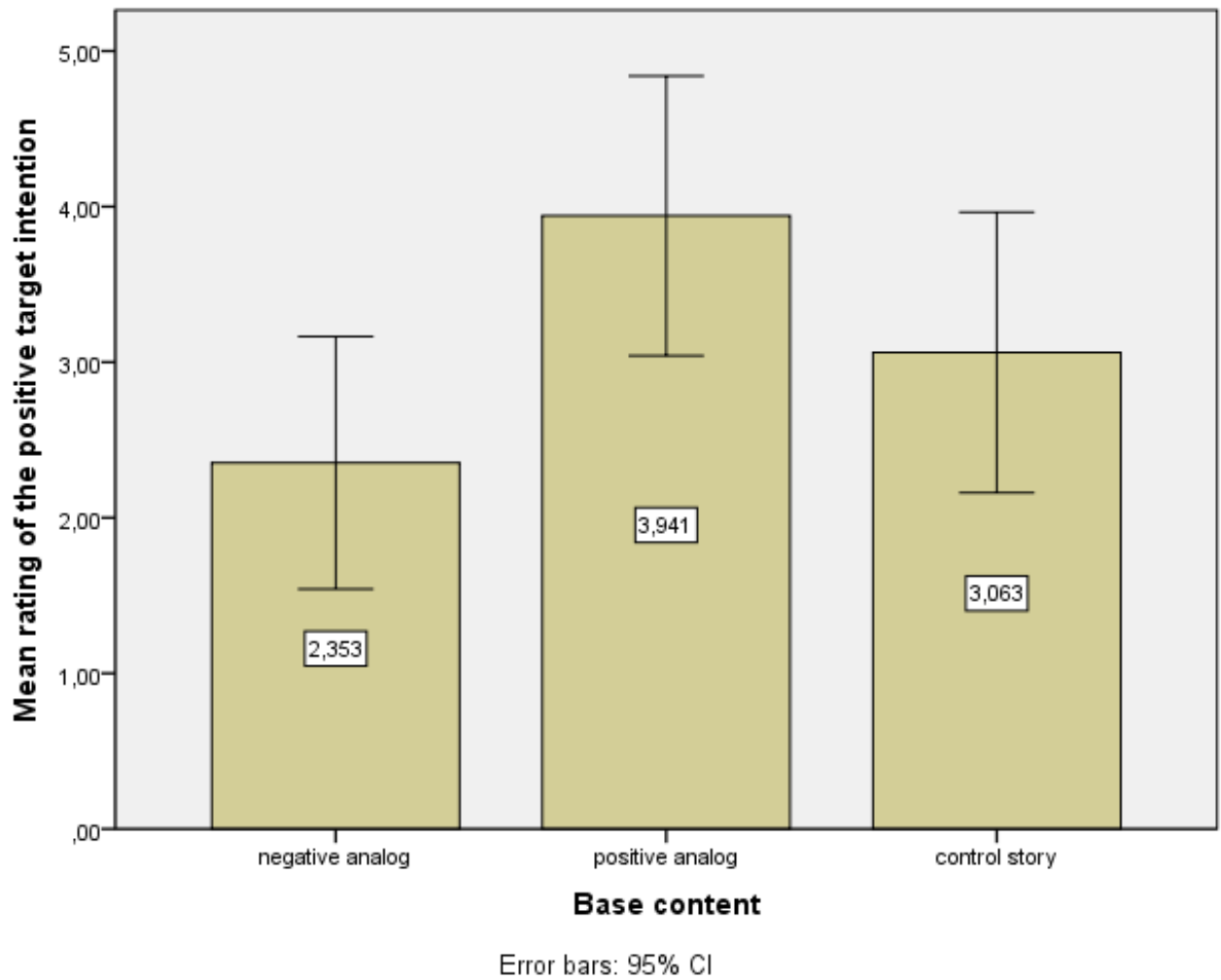
5.3 Experiment 3

Experiment 3 aims to answer the question to what extent the effect of the negative structurally similar base in Experiments 1 and 2 is aided by activation of a negative stereotype. To do this, a pilot study was conducted in which participants were presented with the base stories used in Experiment 2 but without their ending and were asked to complete the story with their own words. Critically, half of the participants read stories with wolves giving advice to bears, while half read stories with ghosts giving advice to bears. While it was equally likely to complete the stories with wolves in a positive or negative way (negative intention n=12 vs. positive intention n=10), ghosts yield predominantly positive completions (negative intention n=2 vs. positive intention n=22). Therefore, by modifying the stories we used we are able to answer to what extent the stereotypes activated by the characters in the base story impact generation of analogical inferences.

More specifically, unlike in Experiment 1 where there was no effect of positive structurally similar base, we hypothesize that when presented with stories with ghosts, people will rate the positive intention higher after they have seen the positive analog rather than the

negative analog and the control episode. Therefore, we expected effect of BASE CONTENT on the ratings of the positive TARGET ratings. To test the hypothesis of the study that there will be a main effect of BASE COONTENT on the ratings of the positive TARGET INTENTION, we performed MANOVA with one independent variable: BASE CONTENT (positive structurally similar, negative structurally similar and control episode) and three dependent variables: positive, negative and neural TARGET INTENTION [$F_{(6, 90)} = 1.403$, Wilks' Lambda = 0.836, p = 0.220]. The only significant effect was the effect of BASE CONTENT over the ratings of the positive TARGET INTENTION: $F_{(2, 50)} = 3.838$, p=0.029, consistent with the hypothesis of the study that positive analog will increase the ratings of the positive intention when the base episode activates consistent positive schema (See Figure 3). The post hoc analyses showed that participants rated the positive intention significantly higher when the base was positive structurally similar than when the base was negative structurally similar, 3.941 vs. 2.353, p=0.008.

Therefore, it seems that when the characters of the analog are associated with a positive stereotype, people tend to generate analogical inferences from the positive analog affecting the perceived explanatory quality of the positive TARGET INTENTION. This is the opposite pattern of results to Experiment 1 where participants tend to generate analogical inferences from the negative analog when the characters were associated with a negative schema. In this experiment there was no effect of the negative analog on the rating of the negative TARGET INTENTION. This confirms the prediction that activation of a positive or negative stereotype accounts for why the negative analog story exerts effect when the main characters are wolves, but the positive analog exerts effect when the main characters are ghosts.

Figure 3 Mean rating of the positive TARGET INTENTION as a function of BASE COTENT condition (positive characters).

5.4 Experiment 4

Experiment 4 aims to answer the question to what extent the preference towards negative intentions in the structurally positive condition in Experiment 2 is a result of activation of a negative stereotype. There are also a number of alternative explanations that suggest that the inverted effect will remain even without presence of a negative stereotype. For instance, participants may make the analogy between the positive analog and the target and make the mapping between the analog and the target. But due to some critical alienable differences between the base and the target such as the quality of the friends' advice, the

analogy may be rejected, which determined participants' preference for the alternative option – the negative intention.

To rule out the first explanation, we changed the identity of the characters in the stories replacing the wolves in with ghosts and used the same base stories as in Experiment 3. However, we used forced-choice procedure for obtaining the participants' preference to positive or negative intentions in the same manner as in Experiment 2.

The two conflicting explanations yielded two predictions. If the preference towards negative intention in Experiment 2 remains after changing the identity of the characters in the positive analog, this would rule out any explanation of the effect by the presence of wolves in both analogs. Furthermore, such results would support the interpretation that the forced choice format enhanced relational encoding but the analogy between the two stories was rejected, so that participants choose the opposite answer. It is worth noting that the study aims to test directly only the first hypothesis.

Therefore, we hypothesized that there will be a significant association between the BASE CONTENT and the preferred intention – participants will prefer the negative TARGET INTENTION more often after seeing the analogs (positive and negative) than after seeing the neutral, control story.

A Chi square test was run over the preferred TARGET INTENTION across the three conditions (positive analog, negative analog and control). There was a marginally significant preference towards negative intention after the positive analog, X^2 (2) =5.301, p=0.071 in comparison to the control condition (Figure 4).

This pattern of results ruled out any explanation of the inverted effect in terms of the negative characters that were common in the positive and negative analogs.

Furthermore, it supports a structural account of the findings, in which participants made the analogy inference but later rejected the analogy. But it is worth noting that the difference does not reach statistical significance.
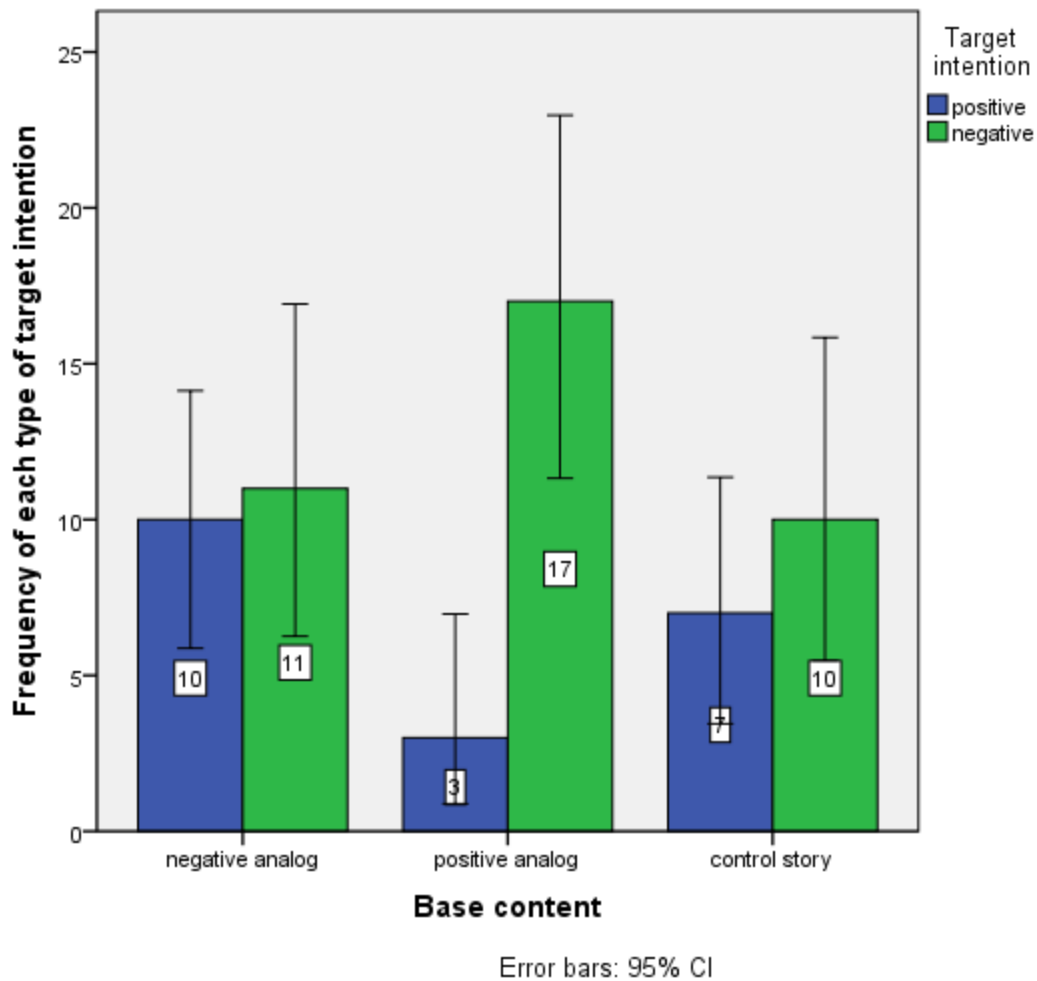
Figure 4 Frequency of preferred TARGET INTENTION (positive vs. negative) as a function of BASE CONTENT (positive characters)

5.5 Experiment 5

In Experiment 5 we aim to further scrutinize the proposed influence of deeper processing in producing the inverted effect in Experiments 2 and 4. More specifically, it was suggested that the inverted effect of the positive analog (increased preference towards the

negative intention) in Experiment 2 and 4 is due to participants' more thorough processing of the material because of the forced choice format. This supposedly facilitated the encoding of the common structure. So, by capitalizing on the findings that people are more likely to process deeper and to explicate the common relational structure if they are provided with two base examples rather than one (Gick & Holyoak, 1983; Gentner et al., 2009), it can be verified to what extend enhancing relational processing by a different method will produce similar pattern of results and confirm explanations in terms of deeper processing. The specific method used originally by Gick and Holyoak (1983) to enhance mapping was to ask participants to compare two analogical problems before asking them to solve a third one. As a control task they asked participants to write individual summaries of each of them, without instructions to compare their structure.

Furthermore, to rule out explanation of the "inverted effect" in terms of response format independent of processing, the experiment used scales rating.

 Based on the results of Experiment 3 where the positive analog base increases the ratings of the positive TARGET INTENTIONS but the supposed enhanced relational processing in Experiment 2 and 4 led to increased preference towards the negative intention, we predict that when participants are given positive relational schema, it will lead to increase in the ratings of the negative TARGET INTENTION in comparison to the condition where they are given positive analogs.

As far as the effect of the negative analog base is concerned, based on the results of Experiment 3 where the negative analog base did not exert any effect on the ratings of the negative TARGET INTENTION, we predict that the enhanced relational processing when participants are given negative relational schema will produce increase in the ratings of the negative TARGET INTENTION in comparison to the condition where they are given negative analogs to just summarize them without instruction to compare their structure.

The study has a between subject design with two independent variables: BASE CONTENT (positive and negative) and BASE TASK (summary vs. comparison).

BASE CONTENT, similarly to the previous experiments, refers to the intention (either positive or negative), that is embedded in the base stories.

BASE TASK refers to the tasks that participants are asked to accomplish – either write individual summaries of the two provided base stories, which has been demonstrated to produce results comparable to providing a single analog, or compare the two provided stories and describe the similarities between them, which has been demonstrated to enhance relational processing and increase inference transfer (Gick & Holyoak, 1983; Gentner et al., 2009).

Participants were randomly assigned to four conditions corresponding to the levels of the two factors: BASE CONTENT AND BASE TASK. The study has three dependent variables: positive, negative, and neutral TARGET INTENTIONS.

To test the hypothesis of the study that there will be an effect of BASE TASK on the ratings of the negative intention, we performed MANOVA with two independent variables: BASE CONTENT (positive and negative) and BASE TASK (summary vs. comparison) and three dependent variables: positive, negative and neural TARGET INTENTION [$F(3, 97) =$ 2.938, Wilks' Lambda = 0.917, p = 0.037 for BASE TASK, $F(3, 97) = 1.627$, Wilks' Lambda = 0.952, p = 0.188 for BASE CONTENT, $F(3, 97) = 1.731$, Wilks' Lambda = 0.949, p = 0.166 for the interaction].

There was effect of BASE TASK over the rating of negative TARGET INTENTION: $F(1, 99) = 5.774$, p=0.018, no effect of BASE CONTENT and no interaction between the two factors, consistent with the hypothesis of the study (see Figure 5), demonstrating that people are more likely to generate analogical inferences that affect the rating of the TARGET

INTENTION following manipulation that enhances the encoding of the common relational structure.

Furthermore, people tend to rate the negative base-inconsistent intention significantly higher after they have compared two positive structurally similar bases than after they have only summarized them. So, the results of the current experiment suggest that this "inverted effect" is most likely associated with altered evaluation of the intentional inferences as a result of the analogical processing.
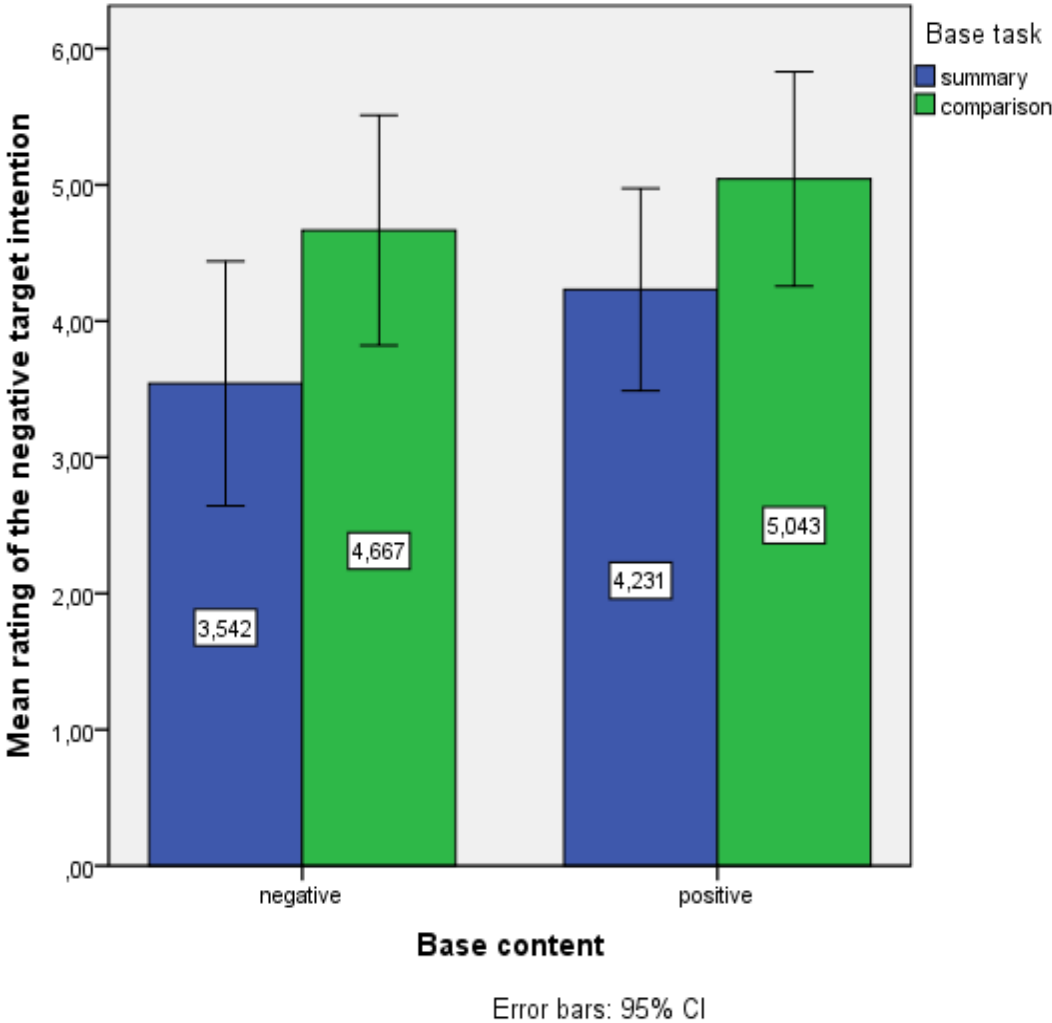

Figure 5 Mean rating of the negative TARGET INTENTION as a function of BASE CONTENT and BASE TASK

CHAPTER 6: General Discussion

6.1 Significance of the obtained results

Support for the hypothesis that analogical inferences play a role in understanding others' intentions. Taken together, the findings of the experiments provide moderate support for the role of relationally similar episodes in producing inferences about others' intentions in ambiguous situations. The starting point of this work was the proposal that understanding intentions may be understood as the output of operation of the mechanism of analogical processing. In a series of experiments it was demonstrated that analogical inferences, i.e. inferences that follow from an analogous prior situation, influence the interpretation of a character's intention in an ambiguous target situation, but their impact interacts with three other factors: depth of processing, evaluation, and activated stereotypes, which produced two

Enhanced understanding of the mechanisms of analogy making. Notwithstanding, our results contribute to current analogy research by showing that analogical inferences based on relational similarity play a role in understanding new situations, even without explicit instructions.

Furthermore, the results of the experiments provide support for the idea put forward by Kokinov and Petrov (2000) that once people encounter a given situation, this situation activates both concrete and abstract prior knowledge. Last, the presented results bear relevance to the role of the process of evaluation and element alignment in analogical processing. While their role has been recognized (Smith & Gentner, 2000), our results suggest a more substantial influence. It seems that a bad analogy is at least as influential as a good analogy. In further studies it will be interesting to explore this process further.

Linking analogy research and the traditional studies on social cognition. There is a general agreement among theorists in social cognition that understanding mental states in ambiguous situations requires filling-in of the missing information from prior knowledge. In

the same time the generation of inferences to unobserved objects, properties, or states, has been studied extensively in analogy making and the central role of relational knowledge has been confirmed in a number of domains. Linking the two fields allows drawing testable hypotheses from the knowledge accumulated in both fields. The current study is demonstration of the fruitfulness of this approach.

Providing a valuable alternative to the existing approaches. The results, which we received, cannot be accounted by ST and TT. For instance, the results are at odds with the proposal of TT that inferences about others' intentions are generated based on abstracted generalized knowledge, because we demonstrated that a single episode could influence understanding intentions in ambiguous situations. While in Experiment 5 the manipulation explicitly tested the effect of abstracted generalizations, these abstractions were very specific and referred to forming relational schemas, which are central for analogical processing, but are largely ignored by TT. Similarly, our results cannot be explained by ST, which proposes that people understand the intentions of others by mentally travelling to similar episodes of their own, because in our case the base situation that contained the intention to be transferred is very different from the target in terms of its objects and attributes. What was common between the two episodes is not the objects but the structure of which these objects are part of, which allows people to use what is known about one situation to extent their understanding about unknown aspects of another, novel situation.

Also, the documented interaction between concrete and abstract knowledge cannot be accounted by direct perception theory, ST or TT.

6.2 Implications

A more general implication of our results is that they pave the way to formulating testable predictions about how people understand intentions in addition to the already documented influence of analogical processing in other aspects of social cognition. This

provides a way forward from the debate between ST and TT toward theories that are more plausible given accumulated knowledge in areas such as memory and similarity processing.

Additionally, our results bear on the idea that analogical processing is a domain general mechanism in operation across all human cognitive functioning. The role of analogies has been demonstrated in a number of areas that have been considered uniquely human such as complex reasoning and scientific thinking. The presented findings extend this work by providing empirical evidence for the involvement of analogies in fast and effortless for humans, but impossible for non-humans processes such as mindreading.

6.3 Limitations

One notable limitation of the current studies is that the effects have been demonstrated with very similar stimuli. Although it allowed in-depth study of the factors that affect inferences, this limits the generalizability of the results.

Another limitation is that participants were studied in grouped setting during classes that may have reduced their motivation and created additional noise to their responses.

The presented studies use explicit measures of intentional inferences, but our subjective experience is that intentional attributions are fast and effortless. Further demonstration with more implicit measures such as skin conductance or reading times are needed to provide further support of these results.

6.4 Directions for further research

A number of testable hypotheses are suggested by the present findings.

First, changes in the target stories may alter the inversion effect demonstrated with the positive structurally similar stories.

Additionally, participants may be tested in more controlled environment.

Introducing a delay between the presentation of the base and the target stories may further confirm the results and extend their application to real-life understanding of intentions.

Last but not least, the effects need to be tested with more diverse set of stimuli such as schematic drawings of situations or even geometric figures.

Also, the general account outlined in the beginning suggests fruitful grounds for investigation. Direct comparison of the role of structural and superficial similarity in generation of inferences among clinical population may shed light on their deficits and provide further support of that account. Neuroimaging studies of the possibility to directly perceive intentions from sensory information may reveal the activation of areas that suggest re-representation of the behavioural data in terms of structural information.

6.5 Conclusions

The presented empirical findings support the hypothesis that relational knowledge plays an important role in the generation of inferences about the intentions of others. They show that even a single concrete prior episode, which is structurally similar to the current ambiguous situation, may influence the attribution of intention to the ambiguous action. In the same time, generalized schemas interact with (either facilitate or hinder) the analogical inferential process. Furthermore, deeper processing seems to enhance analogical inferences. However, these inferences are not unconditionally accepted, but seem to be evaluated according to the alignment of the elements. This evaluation seems to play a significant role as aligned inconsistencies does not lead to lack of predicted effect but rather to an inverted effect. These results cannot be accounted by any of the existing theoretical paradigms (ST and TT), neither by the proposed hybrid models for understanding mental states. The reported empirical findings, as well as the proposed explanatory hypothesis about the existence of a single domain-general mechanism of using relational knowledge to understand others' intentions, demonstrate the fruitfulness of the link which we have established between analogy research and the traditional studies on social cognition.

APPENDIX 1: Contributions of the Current Work

1. **Support for the hypothesis that analogical inferences play a role in understanding others' intentions.** The studies explored and provided moderate support for the hypothesis that analogical processing including inference generation plays a role in understanding intentions in ambiguous situations, which is the core of the current proposal, but also suggest that depth of processing, evaluation and activated stereotypes interact with the inferential process.

2. **Enhanced understanding of the mechanisms of analogy making.** Our results contribute to current analogy research by showing that analogical inferences based on relational similarity play a role in understanding new situations, even without explicit instructions, provide support for the idea that once people encounter a given situation, this situation activates both concrete and abstract prior knowledge and suggest that evaluation may have a substantial role in the process of using analogical inferences

3. **Linking analogy research and the traditional studies on social cognition.** The findings of the outlined experiments both propose methodology to study analogy processing in the context of intentional inferences and raise additional issues relevant to the way social cognition functions such as the role of evaluation and learning.

4. **Providing a valuable alternative to the existing approaches.** The suggested here mechanism for analogy-making provides a way to forward this debate between TT and ST by suggesting a third alternative that is also inferentialist but in the same time unificationist. The suggested mechanism does not need to constrain the influence of prior knowledge to either concrete episodes or generalized schemas as proposed by TT and ST, and shows as well that it is not necessary to assume the existence of two separate mechanisms in order to account for the controversial empirical data.

APPENDIX 2: Publications

Shahbazyan, L., Petkov, G. & Gurova, L. (2014). Relational priming enhances hostile attribution bias (HAB) in adults. *Folia Medica, 56(1),* 19.

Shahbazyan, L., Petkov, G. & Gurova, L. Analogical transfer of intentions. In P. Bello, M. Guarini, M. McShane, & B. Scassellati (Eds.), *Proceedings of the 36th Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society, 2907-2912.

Petkov, G., & Shahbazyan, L. (2013). Cognitive model of confirmation bias effects. In G. Lee (Ed.), *Advances in education research* (Vol. 32, pp. 61-66). Newark, Delaware: Information Engeneering Research Institute.

Shahbazyan, L., Kokinov, B. (2009). The Effect of Objecthood on Processing Efficiency. In: *Proceedings of the 31st Annual Conference on Cognitive Science*. Lawrence Erlbaum, Hillsdale, NJ. (673-678)