GUESSING PERSONALITY FROM A BRIEF SAMPLE OF BEHAVIOUR

WENJIE WU, M.Ed.

Thesis submitted to the University of Nottingham (Malaysia Campus) for the degree of Doctor of Philosophy

October 2014

Abstract

Mentalising as a process for explaining and predicting behaviour relates to inferring mental states and traits of others. Previous research of mentalising has focused too heavily on *mental states* and insufficiently on *personality traits*. Given this context, the current thesis aimed to explore the phenomenon of forming first impressions of personality based on a brief sample of behavior.

In the current research, after being filmed in diverse naturalistic scenarios, targets filled in an "empathy quotient" (EQ) questionnaire and the NEO Five-Factor Inventory-3 (NEO-FFI-3) for respectively measuring empathic traits and the Big-Five personality dimensions. Perceivers were asked to guess the results of target self-reported EQ or the Big Five traits while observing the target in the context of minimal information presented in different types of way (e.g., videos, audios and photographs). Findings from Studies 1 to 8 converge in revealing that perceivers are surprisingly effective in accurately guessing targets who either had low or high EQ and targets who were extreme in one or more personality dimensions, but not so effective in identifying targets with average personality. These judgments were based on the behaviour of the target and not merely on an image of the target. Studies 1 and 2 revealed that perceivers were biased to assume the targets were rather similar to how empathising they perceived themselves, but perceivers' confidence did not predict their accuracy in judgments of target empathy. Study 6 demonstrated a relationship between perceivers' ratings of targets' expressivity and how perceivers judged target EQ. Additionally, a survey was created to examine people's commonsense views about first impressions of personality.

Results of all studies were discussed with reference to the processes by which people make first-impression personality judgments. The current research adds to the

literature of mentalising in speaking about the breadth, versatility and sensitivity of our mindreading abilities.

Acknowledgments

First and foremost, I am very grateful to my family who always support me to pursue my dreams. I especially thank my mother who has never stopped loving me; without her support and encouragements, I would not have made this far. This thesis is dedicated to my mother Ms Lan Chen.

I would like to thank my supervisors Professor Peter Mitchell and Dr. Elizabeth Sheppard for all their help, guide and support of my work. I would like to thank Dr. Chuma Owuamalam for his helpful comments on my annual progression reports.

Finally, many thanks to the students from the University of Nottingham Malaysia Campus and the University of Monash Sunway Campus who volunteered to take part in the research – Without their help, I would not be able to efficiently complete the data collection.

Table of Contents

List of Figures and Tables

CHAPTER ONE

Backgrounds: People' Ability to Make Psychological Inferences

1.1 History of Research on Mentalising

1.1.1 Philosophical Background: The Philosophy of Mind

The philosophy of mind centers on two important issues, the mind-body problem and the problem of other minds. The former concerns how mental phenomena are related to physical phenomena, while the latter asks how we can know about the mental states of another person (Stich & Nichols, 2003). Due to the nature of mind being non-physical and private, for centuries, philosophers had not found solutions to these problems. For example, for solving the mind-body problem, René Descartes proposed substance dualism, distinguishing a physical substance that exists in space and time from a mental substance that only has extension of time. If, as Descartes suggested, there is two-way causal interaction between the physical and the mental, how can this interaction occurs as one is in space while the other is not (Stich & Nichols, 2003)? And, if, as Descartes believed, we can merely experience and access to our own mental states from a first-person subjective position, then how can we come to understand the mental states of other people?

In contrast, abandoning many of the assumptions implicit in the Cartesian tradition, Sellars's ideas about the philosophy of mind focus on a naturalistic concept (Rosenberg, 2011). He believed that our concepts of mental states are in some important ways like theoretical concepts in sciences, allowing us to explain human behaviour (Sellars, 1956). Building on this idea, Lewis (1972) suggested that ordinary terms of psychological states imply a set of tacit laws and knowledge that function to explain and predict behaviour. Thenceforth, philosophers in the camp of folk psychology claim that people are inclined to endorse a broad body of default postulates about propositional attitudes like intentions, desires, and beliefs, which constitute the explanatory psychological system for everyday practices of interpreting and predicting behaviour.

1.1.2 Heider's Commonsense Psychology

In a similar vein, grounded in commonsense psychology, social psychologist Fritz Heider developed his person perception model, trying to make causal explanations about the processes by which people draw inferences about "the presumed events inside the other person's skin" (Heider, 1958, p.1) by observing the person's actions in the social world. By drawing analogy with object perception, where the inferential processes allow people to comprehend the properties of objects by their appearance and motion, Heider (1958) proposes that causal inferences also enable people to understand psychological characteristics (involving mental states and psychological dispositions) adduced from the streams of ongoing behaviour in everyday life. But he recognized that person perception is more complex than object perception in that people have motives that make them act purposefully. According to Heider, people possess a naïve psychology involving an array of unformulated principles and knowledge that allow a causal analysis of other people's intentional

behaviour. Although Heider provided an elaborate description for a conceptual framework of causal mental inferences, he provided no supporting experimental data. His research method was philosophical, which relies on conceptual analysis supplemented with linguistic examples and thought experiments (Malle, 2001).

1.1.3 Empirical Research on Mentalising

Though philosophers and psychologists have proposed intriguing theoretical explanations about the problem of how we engage in understanding mental phenomena, the empirical research on mentalising only began with a study by Premack and Woodruff (1978), asking 'Does chimpanzee have a theory of mind'. They originally invented the term "theory of mind" (ToM), referring to the ability to impute mental states, such as desires, intentions and beliefs to oneself and to others, and this ability allows one to explain other people's past behaviour and to predict what they will do in the future (e.g., Premack & Woodruff, 1978; Wimmer & Perner, 1983; Baron-Cohen, Leslie, & Frith, 1985; Wellman, Cross, & Watson, 2001; Flavell, 2004). The other labels such as "mindreading" and "mentalising" are also used to denote the same capacity.

To examine whether or not chimpanzee possesses a human-like capacity to attribute actions to mental states, such as intention, knowledge, belief, thinking and the like, Premack and Woodruff (1978) designed a series of experiments in their pioneering research. In these experiments, an adult chimpanzee was shown a series of videoed scenes of a human actor struggling with various problems, ranging from a relatively simple ones such as inaccessible bananas to the more complex cases, involving an actor unable to extricate himself from a locked cage, shivering because of a malfunctioning heater, and so on. Along with each videotape, the chimpanzee was given several photographs involving one that showed a solution to the problem,

such as a stick for inaccessible bananas and a key to release the locked up actor. Premack and Woodruff assumed that if the chimpanzee could consistently choose the correct solution for each scenario, then it would indicate that the chimpanzee had recognized that each videotape represented an unsolved problem and understood that the actor's behaviour would be guided by his intention of solving that problem. Selecting the correct solution was accordingly regarded as a sign of having a theory of mind.

However, according to Daniel Dennett (1978), the task Premack and Woodruff devised is not a good test for theory of mind. In his commentary, "Beliefs about beliefs", Dennett (1978) pointed out that performance in the problem-solving tasks cannot determine the basis for the chimpanzee's solution to the problem facing the actor. Did the chimpanzee make a prediction of the actor's behaviour (by choosing the solution to each problem) based on its beliefs about the beliefs and desires of the actor? Alternatively, was the chimpanzee simply offering a solution to a physical problem without giving any consideration to the actor's mental states? If so, then the chimpanzee did not need to have a theory of mind to pass the problem-solving tasks.

Dennett suggested an alternative task, enabling the concept of false belief to be tested. Beliefs are characterized by properties of mental representation (Flavell, Miller, & Miller, 1993; Wimmer & Perner, 1983). Saying that one has a state of belief means he/she is representing the physical world in accordance with his/her belief, the contents of which could truly reflect reality in some cases while in other cases (when the beliefs are false) could incorrectly represent or even conflict with the situations and events around the person. That is, the concept of false belief serves to connect the internal world with the external world (Dennett, 1978; Wellman et al.,

2001). More importantly, behaviour that is grounded in a false belief is predictive (Flavell et al., 1993): Whether a belief is false or not can to some degree be determined by observing an actor's overt behavior – If a person is earnestly searching for a particular object but in the wrong place, this could be a sign that he or she holds a false belief.

In one example suggested by Dennett (1978), based on the popular puppet show characters, Punch and Judy, he described a scenario in which Punch holds a false belief about the location of Judy: Judy had escaped the box while Punch's back was turned. If observing children expected Punch to incorrectly search for Judy in the box, this could be a sign that the children have an understanding of false belief. In other words, the concept of false belief provides a practical approach to studying ToM via examining prediction of belief-related actions.

Grounded upon these ideas, Wimmer and Perner (1983) developed the famous false belief paradigm, which has been warmly embraced as the "litmus test" of children's acquiring of sophisticated ToM. In the standard false belief task, the change-of-locations task (Wimmer & Perner, 1983; see Fig. 1.1), a story protagonist, Maxi, puts his chocolate in the green cupboard, and then leaves the scene. In his absence, the other character, his mother, moves the chocolate to the blue cupboard. A child participant is asked the focal false-belief question in terms of action prediction ("where will Maxi look for his chocolate when he returns?") or in terms of thoughts ("where does Maxi think his chocolate is?"). If the child can correctly answer the question by predicting Maxi will search for the chocolate in the original place where he left it or by saying Maxi thinks his chocolate is in the green cupboard, then the child is regarded as being able to understand false belief.

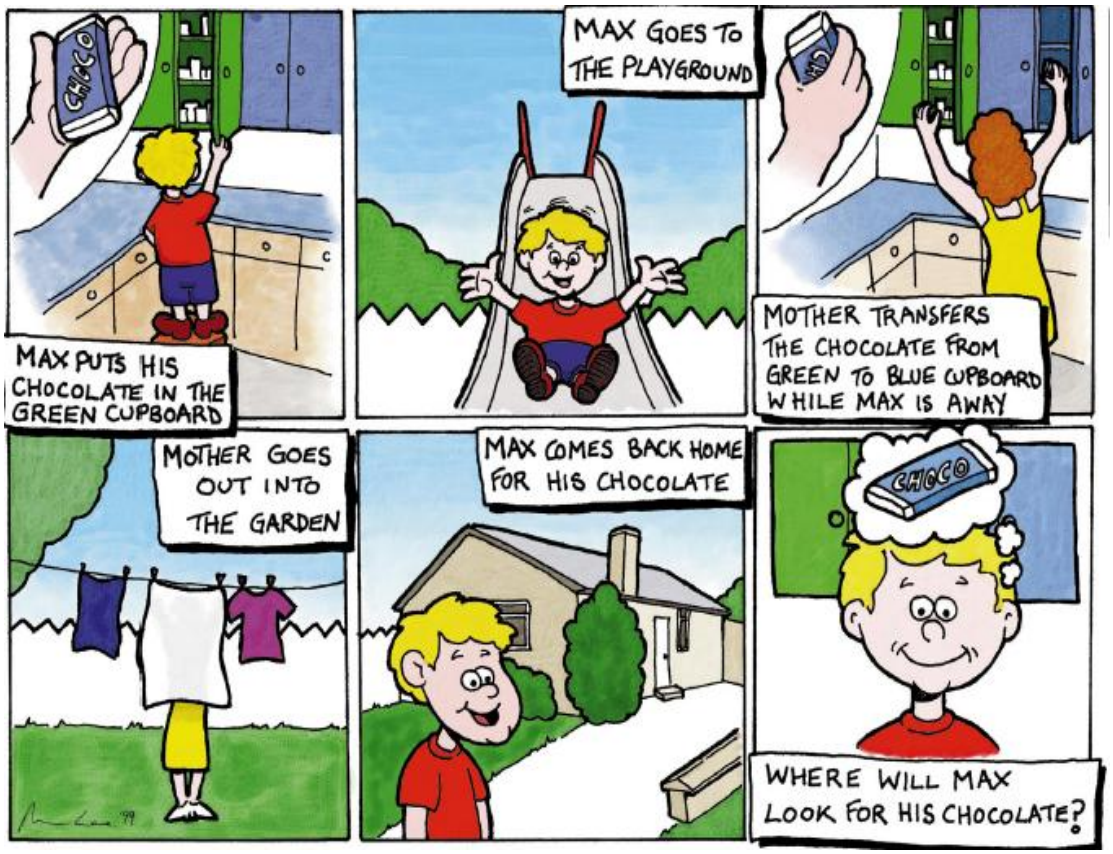


Fig. 1.1. The chocolate story (adopted from Perner & Lang, 1999).

The other widely used false-belief task, the unexpected-contents task (Gopnik & Astington, 1988), engenders a familiar container with an unexpected content (see Fig. 1.2). For example, in the "Smarties" task, a child is asked what she/he believes to be the contents of a box that looks as if it holds candy called "Smarties". After replying to the question (usually with "Smarties"), the child is showed that the box in fact contains pencils. And then the box is closed and the false-belief question is asked ("what do you think another person, who has not been shown the true contents of the box, will think is inside?"). The child passes the task if she/he answers "Smarties" instead of "pencils".

Fig. 1.2. An illustration of the unexpected-contents task (adopted from www.autismservice.org).

Using these false-belief tasks, researchers investigate when and how children begin to have a theory of mind. Numerous studies in the field of developmental psychology have revealed an improvement at the age of 4 years when children start to systemically pass the change-of-locations task and the unexpected-contents task (e.g., Wimmer & Perner, 1983; Flavell, 2000; Wellman et al., 2001; Sabbagh, Moses, & Shiverick, 2006). Some studies introduced modifications to the standard false belief tasks that led to improvements in 3-year-olds' performance (e.g., Chanlder, Fritz, Hala, 1989; Carlson, Moses, & Hix, 1998). Subsequently, Wellman et al. (2001) conducted a meta-analysis of 178 studies that used a false-belief task, and concluded

that although performance can be affected by the characteristics of any particular task, the general impression is still that children improve considerably around the time of their fourth birthday.

Does that mean infants and young children do not have a theory of mind? Recent work using spontaneous-response tasks (nonverbal false belief tasks) has shown that infants first understand that others hold false beliefs during the second year of life (e.g., Onishi & Baillargeon, 2005; Southgate, Senju, & Csibra, 2007; Surina, Caldi, & Sperber, 2007; Song, Onishi, Baillargeon, & Fisher, 2008; Buttelmann, Carpenter, & Tomasello, 2009; Baillargeon, Scott & He, 2010). In a violation-of-expectation task (Onishi & Baillargeon, 2005), for example, infants as young as 15 months looked for longer at the unexpected than at the expected action. In the case of an unexpected action, the actor searched for the target object as if she did not hold a false belief, even though this particular actor had no basis for holding a true belief. Hence, infants behaved as if they were surprised that the actor appeared to have a true belief, suggesting that they expected the actor to hold a false belief. These results indicate that when we use an implicit false-belief test even infants can demonstrate mentalising to some extent. We do not yet know if this mentalising takes the same form as that seen in a four-year-old child who gives the correct answer in an unexpected-contents task or a change-of-locations task. This matter might be clarified by the findings of future research.

1.2 Mental State Reasoning

1.2.1 Error and Bias in Mental State Reasoning

Although even infants exhibit some understanding of false belief in spontaneous-response tasks, studies using the standard false-belief tasks have

consistently suggested that children around the age of 3 years have difficulty in explicitly imputing false beliefs to other people. They systemically show errors and biases when asked to predict others' behaviour in relation to their false beliefs: In the change-of-locations task, they expect Maxi to look for the chocolate in the place where it is located; likewise, in the unexpected-contents task, they seem to think another person knows that the box holds pencils rather than Smarties when first shown the Smarties box. Despite that, some researchers have pointed out that young children's difficulties in inferring false beliefs may result from task artifacts related with requirements in language processing and demands on executive functioning skills, including working memory and inhibition (e.g., Apperly, Samson, & Humphreys, 2005; Birch & Bloom, 2004; Carlson & Moses, 2001; Bloom & German, 2000; Zelazo, Carter, Reznick, & Frye, 1997). Notwithstanding, the results of the meta-analysis by Wellman et al. (2001) have revealed that these factors cannot wholly explain age-related developmental changes in children's performance on the standard false belief tasks.

In the false belief tasks, child participants know the current state of reality ("where the chocolate is now" in the change-of-locations task; "what is really in the Smartie box" in the unexpected-contents task) that conflicts with the contents of Maxi or another person's beliefs. Young children thus seem not to understand that minds hold beliefs about the world and these beliefs have an essential impact on behaviour even if they sometimes can be false. Without this comprehension of mind, young children are ineffective in making predictions of others' behaviour related to false beliefs and are prone to be biased by their own knowledge about reality.

Given the fact that 4-year-old children pass a test of false belief, and therefore demonstrably have a sophisticated concept of false belief, would it be the case that

older individuals, especially adults, are extremely proficient in guessing what other people are thinking? Lines of research have reported that adults usually have trouble in making multiply embedded inferences about mental states of others (e.g., "Bob thinks that John knew that Mary wanted to go to the ship"; Kinderman, Dunbar, & Bentall, 1998; Rutherford, 2004), and they often make errors in false belief tasks when asked to perform a demanding concurrent task requiring working memory or other components of executive functioning (e.g., McKinnon & Moscovitch, 2007; German & Hehman, 2006), though they usually have no difficulty in inferring one person's belief, or one person's belief about another person's belief (Fletcher, Happé, Baker, Dolan, Frackowiak, et al., 1995; Stone, Baron-Cohen, & Knight, 1998). But some researchers view these findings more cautiously, and argue that this kind of ToM task might not adequately control for incidental demands on executive functioning, leading to underlying undervaluation of mentalising abilities (Bloom & German, 2000; Apperly et al., 2005; Apperly, Back, Samson, & France, 2008). Even so, when simply told about someone's false beliefs, adults perform much faster in judging the true state of an object instead of the false belief that a person holds about the object (Apperly et al., 2008). In particular, adults were presented two sentences corresponding to the information about a situation (e.g., "Really, the ball on the table is yellow"), and the information about someone's false belief about the situation (e.g., "He thinks that the ball on the table is red"), and then they were asked to judge the accuracy of a picture probe that either depicts reality or the person's false belief. The results show that adults made more errors and responded slower in the condition of false belief than the condition of reality.

In another experiment, Birch and Bloom (2007) manipulated participants' knowledge about the reality of the target object's (violin) place to where it was

moved and rearranged in the absence of the protagonist Vicki (see Fig. 1.3). The observing participant had to predict to which of four given places the protagonist will first look for the violin. The findings demonstrate that an adult's own knowledge biases his/her prediction of another's action. Participants who knew to where the violin was exactly displaced, and who had available a plausible explanation for Vicki to act in accord with their knowledge, were significantly less likely to predict that Vicki would act according to a false belief than were those who knew the reality of the place of the violin but did not have available a plausible explanation for Vicki to act in line with their knowledge (Birch & Bloom, 2007).

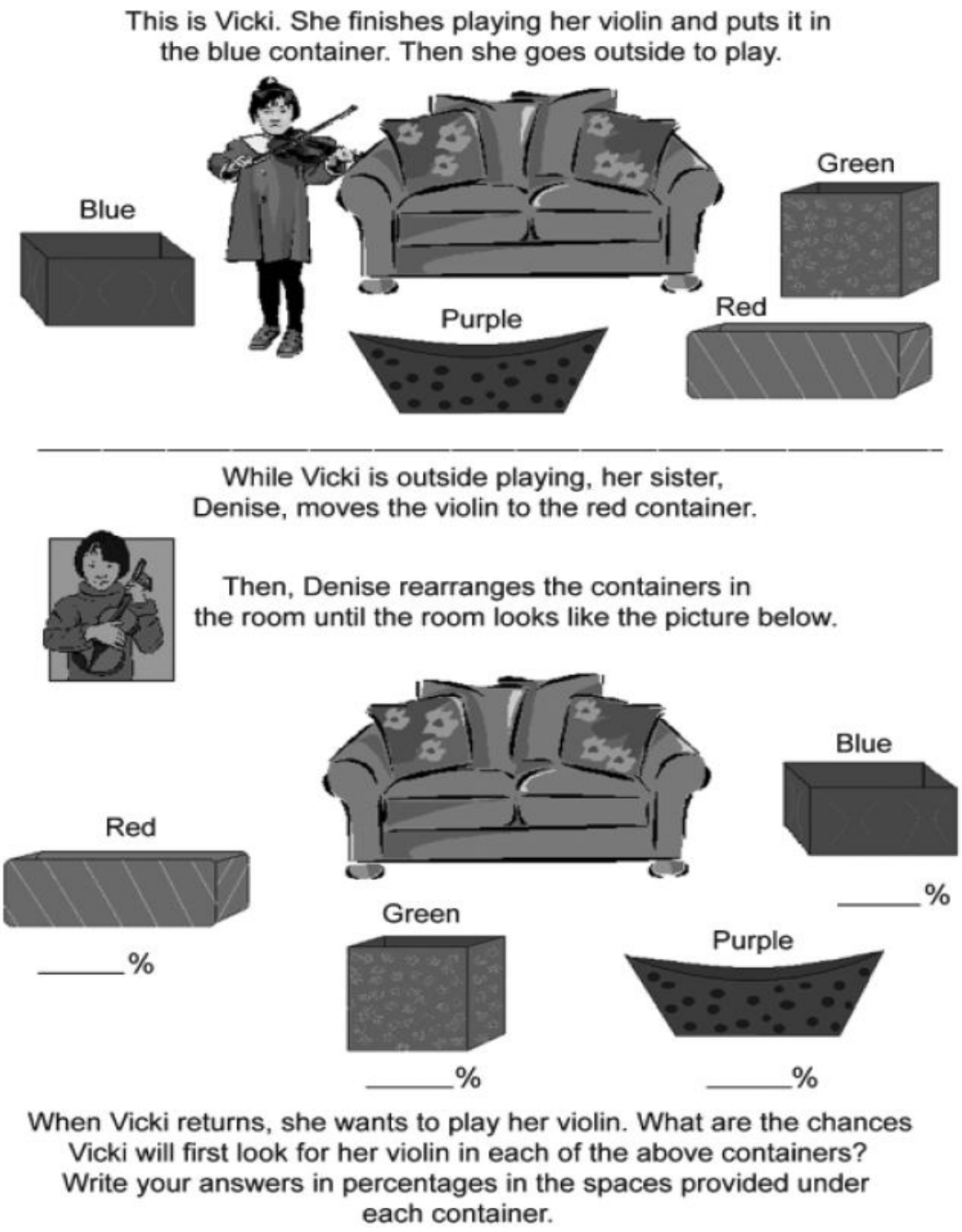


Fig. 1.3. The knowledge-plausible version of the task (adopted from Birch & Bloom, 2007).

Keysar and colleagues (Keysar, Barr, Balin, & Brauner, 2000; Keysar, Lin, & Barr, 2003) devised a communication game and found that adults frequently had difficulty separating their own privileged knowledge of a target from that of a competitor who held incomplete knowledge. Using a similar task, Apperly, Carroll, Samson et al. (2010) conducted three experiments to investigate the cognitive

processes contributing to adults' errors in ToM understanding. They suggest that adults are efficient at switching perspectives, but actually using what another person knows to interpret what they say is relatively inefficient, giving rise to egocentric errors during communication. Further evidence from a study using computer-based tests of false-desire reasoning (Apperly, Warren, Andrews, Grant & Todd, 2011), which records response times and error rates, reveals that as with young children, older children and adults found it more challenging to reason about false belief and negative desires than true beliefs and positive desires. The researchers therefore suggest developmental continuity in ToM (Apperly et al., 2011; Apperly, 2013, for a review of adults' ToM).

1.2.2 Is Mental State Reasoning Automatic?

It is beyond doubt that normal adults are equipped with mature mentalising abilities. Even so, according to the aforementioned studies, they still exhibit a pattern of reasoning bias similar to that found in children younger than 4 years when drawing inferences about false beliefs of others. Given this intriguing fact, it seems reasonable to ask, "Is mental state inference automatic?"

Apperly and colleagues (2006) conducted the first investigation to address this question. According to them, if false belief reasoning is an automatic process, then participants should draw false belief inferences even if they do not have any particular reason to do so. In each trial, participants watched the same video stimuli in which a male actor hid an object in one of two of the same opaque boxes and a female actor indicated where she thought it was hidden. During the presentation of the video stimuli, probe sentences were presented at unpredictable intervals to elicit false belief or reality judgments from participants. Participants needed to respond to the probe sentences corresponding to either a belief question or a reality question,

and their response times were recorded. In Condition 1, an incidental false belief task where participants could track relevant aspects of reality but with no particular reasons to track the woman's false beliefs, participants responded more slowly to unexpected questions regarding the female's false beliefs about the object's location than to questions concerning the object's real location. In Condition 2 explicit belief and reality tracking, participants were explicitly instructed to keep track of where the woman thought the object was located and where it really was located, whereas in Condition 3 explicit belief tracking, participants were only asked to monitor where the object was located but were not required to indicate the correct location of the object at the end of trials. The results of these two further conditions displayed no difference in response times to belief and reality questions when participants were instructed to track the woman's beliefs about the location. Taken together, the researchers concluded that adults do not automatically reason about false beliefs of another person.

In a subsequent study, Back and Apperly (2010) have extended the method to examine true beliefs that are sometimes thought to be imputed by "default" (e.g., Leslie & Thaiss, 1992). Sequences of pictures were shown in which the location of an object and a female character's belief about the location of the object often changed (see Fig. 1.4). During the picture sequences, participants had to respond to an unpredictable probe picture about where the woman thought the object was located or where the object was actually located. Following the same hypothesis in Apperly et al. (2006), in Experiment 1 using an incidental belief task, participants were explicitly instructed to monitor the real location of the object but they were not specially instructed to keep track of the woman's beliefs about the location of the object. If belief reasoning is automatic, participants should respond to the belief

questions as fast as to the reality questions even when there is no reason for making belief inferences. The results replicated and bolstered the previous findings by Apperly et al. (2006), providing new evidence indicating that even in the cases of true beliefs participants responded more slowly to belief probes in contrast with matched reality probes. Two further experiments ruled out the possibility of intrinsic differences between the belief and reality probes and confirmed that there was no difference in reaction times to belief and reality questions when participants had few reasons to infer beliefs spontaneously. All in all, the evidence suggests that adults do not spontaneously attribute beliefs to other people, whether their beliefs are true or false.

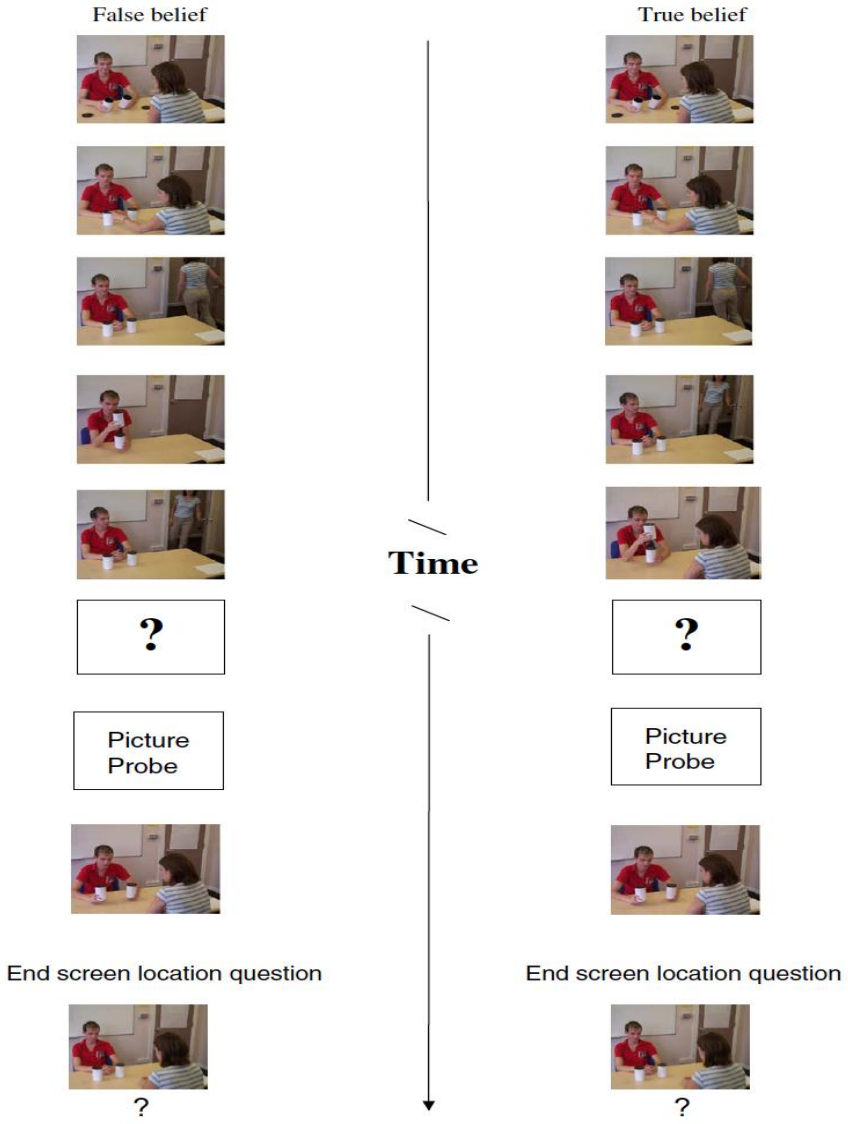


Fig. 1.4. Examples of a true and false belief experimental sequence (adopted from Back & Apperly, 2010).

1.3 People's Inferences of Trait

Comprehension of false belief has been widely thought to be the milestone of a mature adult-like mentalising capability which functions to explain and predict people's intentional actions in everyday life. This framework takes it for granted that we act according to our mental states involving intentions, desires and beliefs, and that we can infer such mental states from situational factors.

Take the change-of-locations task as an example. In order to predict Maxi's behaviour in searching for the chocolate, an observer needs to realize the scenario Maxi experienced (in which the location of the chocolate was transferred from one place to the other in his absence) and mentally picture the situation from Maxi's perspective. In other words, inferring false beliefs in this kind of task requires the observer to understand how the situation that Maxi experiences leads him to hold a false belief. In short, Maxi finds himself in a situation that gives rise to a state of false belief, and taking only this into consideration, we might predict that he will wrongly search in the place he last saw the chocolate.

Imagine, though, that we know that Maxi is an extremely intelligent boy who is good at guessing that his mother will move the chocolate to the fridge. Maxi might realize that the chocolate will melt in the cupboard and he might further realize that his mother will have enough sense to move the chocolate to a place where it will be preserved and remain edible. Hence, in taking account of Maxi's intelligence, we predict that he will not look in the place he last saw the chocolate but that he will look in the fridge on appreciating that he is sensible enough to realize that his mother will have moved the chocolate.

Another instance is taken from ordinary observation in the social world. Imagine that you were invited to a fantastic Christmas party held by your friend, and you were told that other friends of your friend were also invited, many of whom you had never met: Would you be happy to join the party? If only considering the situational factors (a fantastic party and a friendly social environment), we might predict that you would attend the party; however, your friends who know you are introverted and very shy might predict that you would not attend.

There is no doubt that the examples outlined above of reasoning about personality in making predictions of behavior require the ability to mentalise. Consequently, it seems that previous research has focused too heavily on *mental states* and insufficiently on *personality traits* when investigating how people mentalise.

There is evidence showing that it is very common to describe people and behaviour in terms of personality traits: In one study students were asked to provide confidential descriptions of their classmates, and traits dominated the description (65%) among the five categories (Park, 1986). Furthermore, the terms of personality traits not only serve to summarize behaviour, but also function as causal concepts. Classic attribution theories, concerning the processes by which people explain the causes of behaviour and mental events, suppose that people naturally give first consideration to an actor's personality traits when explaining and predicting his behaviour. For instance, Heider believed that psychological characteristics of the person (including mental states and psychological dispositions) should be on the central focus in scientific research of person perception. He claimed:

"The discussion (of interpersonal relations) will center on the person as the basic unit to be investigated… Of course, in dealing with the person as a member of a dyad, he cannot be described as a lone subject in an impersonal environment, but must be represented as standing in relation to and interacting with another person. Moreover, the fact that the interrelation is with another person and not an object means that the psychological world of the other person as seen by the subject must enter into the analysis" (Heider, 1958, p.1).

The subsequent attribution theorists, such as Jones and Davis (1965), duly credit the causal role of personality traits in people's explanations of others'

behaviour. They suggested that people use information about another person's behaviour and its effects to draw trait correspondence inferences. For example, if someone is willing to lend money for his friend who is in need (intentional behaviour), and also would like to help strangers, people probably attribute a corresponding trait to the person (He is helpful.).

Granted, the attribution theories have highlighted the kinds of information that promote inferring traits and dispositions about the person rather than the situation, but they have provided little empirical research for investigating fine-grained characterizations of the kind of information used to produce particular trait inferences (Uleman, 2005). Research into spontaneous trait inferences (STIs) offers a wealth of evidence for trait inferences, suggesting that we draw trait inferences to explain behaviour of others even if we are not cued to reason about trait information and such spontaneous trait inferences can even occur without awareness (Winter & Uleman, 1984; Uleman, Newman, & Moskowitz, 1996, for a review; Uleman, Hon, Roman, & Moskowitz, 1996; Todorov & Uleman, 2004; Rim, Uleman, & Trope, 2009; McCarthy & Skowronski, 2011).

In the very first STIs study, Winter and Uleman (1984) investigated whether people infer personality traits automatically while processing behavioral information. The study adopted Tulving's encoding-specificity paradigm, in which participants read sentences describing people performing trait-implied actions, and then they had to recall each sentence in the different cuing conditions, including a dispositional cue (e.g., generous), a strong non-dispositional semantic associate to an important word in the sentence, and no cue. Participants were asked to memorize each sentence under one of the three cuing conditions. They performed best in the condition of dispositional cues and were apparently unaware of having made trait inferences.

According to the encoding-specificity paradigm, these results suggest that people unintentionally make trait inferences during the process of encoding other people's actions. Researchers, thus, claimed "attributions may be made spontaneously, as part of the routine comprehension of social events" (Winter & Uleman, 1984, p.237).

These initial findings have been extensively replicated in considerable literature on STIs using various paradigms. In one example, Kressel and Uleman (2010) reported a relation recognition paradigm where participants were asked to determine whether there is causal connection between a group of word pairs in a nonsocial task and a social task. Nonsocial stimuli included 32 casually related word pairs (e.g., spark – fire) and 32 associated word pairs (e.g., shrimp – ocean), and social stimuli involved 32 trait-behaviours word pairs (e.g., dumb – fail) and 32 filler adjective-verb word pairs (e.g., gentle – touch). In both tasks, half the pairs were presented in each order. Participants completed the nonsocial and then the social relation recognition tasks on computer by responding to the instruction "determine whether the concepts described by each word pair are causally related" (see Fig. 1.5). The researchers hypothesized that if traits and actions are causally linked, participants should identify predictive sequences (e.g., clumsy → stumble) faster than diagnostic sequences (e.g., blush → shy). The results confirmed their hypothesis, revealing asymmetric reaction times for detecting causal relationships, with predictive words (trait → behaviour) being faster than diagnostic orders (behaviour → trait), and this spontaneous trait attribution was as strong as for nonsocial cause-effect inferences. Therefore, Kressel and Uleman suggest that traits and behaviours are mentally represented as causally correlated, and that isolated traits are inherently causes of actions.

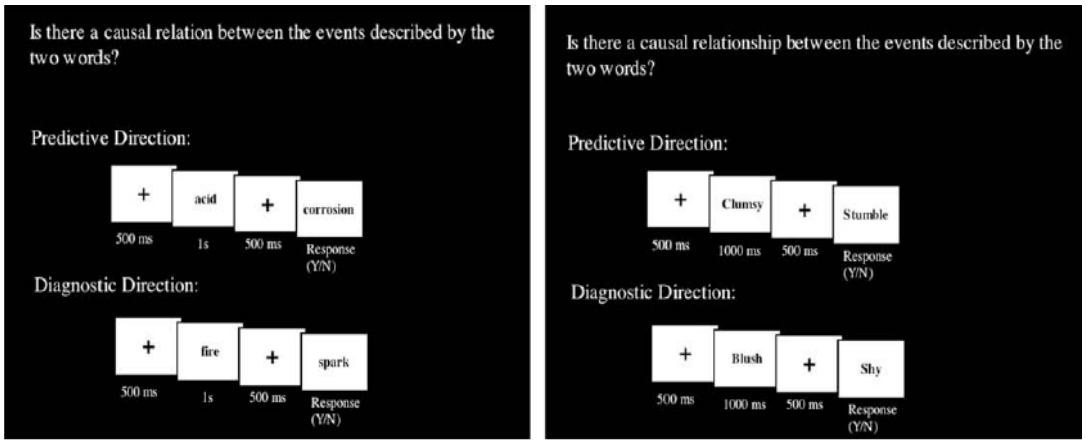


Fig. 1.5. Sample trials from nonsocial and social relation recognition tasks (adopted from Kressel & Uleman, 2010).

Spontaneous trait inferences can also influence how people predict what the other person will do in the future (McCarthy & Skowronski, 2011). In a series of studies, participants were first exposed to a group of photo-behaviour dyads in an initial exposure task, with each dyad pairing a trait-implicative behaviour described in a sentence with an actor (who is presumed to perform that action). This procedure aimed to elicit an implicit trait inference about the actor. In the following task, participants were required to match the actors with whom the behaviours were paired to new actions the actors were thought to be likely to perform. The results show that participants made predictions of actions that were consistent with the inferred traits, and this prediction occurred regardless of behaviour recall, and regardless of whether participants were explicitly instructed to make trait inferences or not, which altogether provides compelling evidence suggesting that an unintentional trait inference has an effect on people's prediction of other people's behaviour.

To summarize, from attribution theories it seems that it is important for people to be able to infer traits when trying to explain behaviour or when making predictions of behaviour. Moreover, research into spontaneous trait inferences has

provided credible evidence suggesting that inferring traits helps people to interpret and predict behaviour; and it might be an automatic process (e.g., McCarthy & Skowronski, 2011; Kreseel & Uleman, 2010; Winter & Uleman, 1984).

Why is it so important to infer traits? According to simulation theory of mentalising, we infer what is in the mind of another by the capacity to project ourselves imaginatively into the perspective of another person and then deploy our own decision-making capacity to simulate similar psychological states and processes in ourselves (e.g., Gordon, 1986; Gorman, 1989; Apperly, 2008, for a review). In this light, if we can make valid judgments on other people's psychological traits, we are more likely to stand a chance of adopting the other's perspective and then temporarily 'seeing the world through their eyes' (Peterson & Riggs, 1999). Doing so allows us to reason about their mental states effectively. To illustrate, here is a quote from the famous fictional detective Sherlock Holmes depicted in the short story The Musgrave Ritual. In order to trace what happened to Musgrave's missing servant Brunton who had stole the document of the Musgrave ritual that records the position of treasure, Sherlock Holmes utilized his amazing capacity for mental simulation:

"You know my methods in such cases, Watson. I put myself in the man's place and, having first gauged his intelligence, I try to imagine how I should myself have proceeded under the same circumstances. In this case the matter was simplified by Brunton's intelligence being quite first-rate, so that it was unnecessary to make any allowance for the personal equation, as the astronomers have dubbed it. He knows that something valuable was concealed. He had spotted the place. He found that the stone which covered it was just too heavy for a man to move unaided. What would he do next? He could not get help from outside, even if he had someone

whom he could trust, without the unbarring of doors and considerable risk of detection. It was better, if he could, to have his helpmate inside the house. But whom could he ask? This girl had been devoted to him. A man always finds it hard to realise that he may have finally lost a woman's love, however badly he may have treated her. He would try by a few attentions to make his peace with the girl Howells, and then would engage her as his accomplice. Together they would come at night to the cellar, and their united force would suffice to raise the stone. So far I could follow their actions as if I had actually seen them." (Arthur Conan Doyle, 1893).

In short, reasoning about personality traits plays an essentially important role in mentalising, because like mental state inferences it shares the same properties as intentionally-generalized inferences (McCarthy & Skowronski, 2011), serving to make sense of behaviour in daily life; because it can occur unintentionally as part of the routine comprehension of social events (Winter & Uleman, 1984), and because it focuses on the person instead of the situation, allowing us to be more or less a Sherlock Holmes in our complex social world – We put ourselves in the place of someone else, we make an adjustment for the "personal equation" (i.e., we take into consideration the target's personality traits) and in doing so we mentally stimulate what is in the mind of the target person.

1.4 Can People Infer Mental States Based on a Sample of Behaviour?

We already know that even typical adults show some egocentric biases, sometimes called "the curse of knowledge" (Birch & Bloom, 2007), when inferring false beliefs. This implies that even if we possess a fully-fledged mentalising capacity, it does not ensure that we are adept in employing our mentalising to reason about the multifaceted mental states without bias. In the real world, our practices of everyday mindreading are more complicated and much subtler than drawing

inferences of someone's false belief about the location of an object. We often only have information from a single observation to guide us in immediately drawing a causal inference. For example, on witnessing an athlete's face at the end of the game we might have conjectured whether his team had won even though we missed the live coverage of the contest. Consider another example in observing of our friends: We can probably tell whether that smile truly indicates happiness with their birthday presents or whether it is just a sign of courtesy.

To what extent is this everyday mindreading accurate when it is based on but a brief sample of behaviour? Apparently, various versions of false belief tasks cannot help us to answer this question. False belief tasks are usually simple and repetitive, and may lack the subtlety, sophistication and uncertainty of much everyday mindreading (Apperly, 2013).

Considering these factors, researchers have developed an ecologically valid approach "empathic accuracy" (indexed accuracy of everyday mindreading; Ickes, Buysse, Pham, Rivers, Erickson et al., 2000, for a review) to examine whether we can reason about the contents of mental states on observing segments of behaviour under naturalistic circumstances. In a study by Zaki and colleagues (2009; see Fig. 1.6), targets were asked to talk about the 4 most positive and 4 most negative autobiographical events (that they were willing to discuss in a laboratory context) while being videotaped. After that, they were asked to view the videos and rate each event for emotional valence and intensity using a 9-point Likert scale (from very negative to very positive) by responding to the question "how did you feel while talking?" Later, Perceivers were instructed to view the video clips of each target and evaluate each event for emotional valence and intensity using the same 9-point scale by responding to the question "how did this person feel while talking?" The results

demonstrate that perceivers' inferences of the targets' affect moderately correlated with the targets' self-ratings.

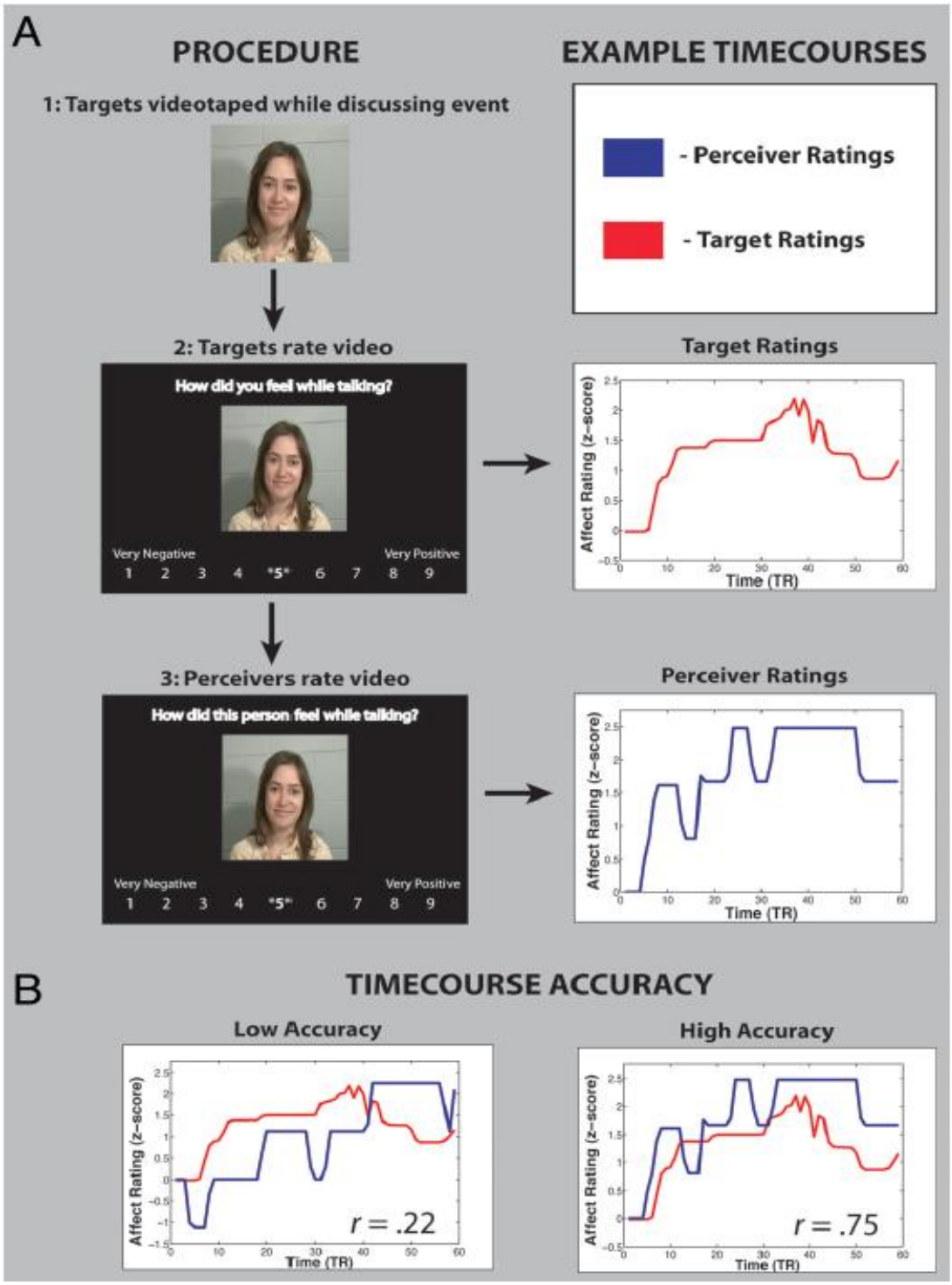


Fig. 1.6. Task design and sample behavioral data (adopted from Zaki, Weber, Bolger, & Ochsner, 2009).

Using a procedure with similar ecological validity, researchers have recently investigated the "retrodictive" aspect of mindreading in realistic situations instead of artificial settings or hypothetical scenarios. According to Gallese and Goldman (1998), an ability to read minds enables us to retrodict what another person had experienced or what they had thought and felt, that is, "making a 'backward' inference from the observed action to a hypothesized goal state" (Gallese & Goldman, 1998, p.497). Researchers have created a novel study for examining this kind of mindreading (Cassidy, Mitchell, Acquah et al., in press). Thirty normally developing adults and 19 adults with autism viewed 21 video clips lasting from 1.3 seconds to 6 seconds, each showing a target's reaction as he/she received one of three gifts (chocolate, monopoly money or a homemade novelty). Participants were asked to guess which gift the target had been offered out of the three options and to estimate the emotion of the target. The results show that normal adults could correctly guess who received chocolate, a homemade novelty or monopoly money at above chance levels, while autistic individuals performed above chance only in the scenario when the recipients had received monopoly money. Furthermore, typical adults who made accurate inferences about the gifts also tended to be successful in reasoning about the emotions expressed by the recipients in the three gift conditions, whereas the autistic individuals only successfully estimated the emotions when monopoly money had been received as a gift. These data suggest that in processing another person's facial expressions, typical adults can infer which gifts other people had received with reference to reasoning about their emotions. Moreover, even autistic individuals had success at least in one of the scenarios in that they could guess when the target had received monopoly money.

Pillai, Sheppard, and Mitchell (2012) reported a similar finding based on a study of whether people can guess what occurred to targets after observing their reactions to real-life scenarios. Forty video clips each lasting only a few seconds were used as stimuli, in each of which, a target person unexpectedly experienced one of four possible events performed by the researcher (joke, waiting, compliments and story). For example, in the story scenario, the researcher related a story about a series of misfortunes that she experienced earlier that day. Thirty-five participants were asked to guess which of the four scenarios they thought the target person was responding to while viewing each video clip, and their eye movements were recorded. The participants were able to successfully judge from a small sample of behavior which events had previously happened to the targets, with best performance in the scenario of waiting. As we might expect, the eye movement strategy of the participants varied according to scenarios experienced by the target but surprisingly, looking at the eye region of the target correlated with poorer identification of the scenarios. The researchers concluded that participants flexibly use different visual strategies for making retrodictive mentalising inferences about events happening in the world and that participants do not necessarily attend to the eyes most of all when mentalising.

In light of these data, it seems that people have a great talent for inferring a state based on a brief sample of behaviour. They can make inferences about the affect another person was experiencing while viewing fragments of behaviour; they can infer what gift one had unexpectedly received by observing their facial reactions spanning but a few seconds; they can guess what had happened to others from witnessing a brief sample of behaviour; and they also can infer the contents of mental and emotional states the other person was experiencing at some given moments

during unstructured dyadic interactions or merely in watching such interactions in videos (e.g., Ickes et al., 2000, for a review; Hall & Mast, 2007).

1.5 How Well Can People Infer Personality Traits on First Meeting?

Mentalising as a process for predicting and explaining behaviour has two broad components. One relates to the process of inferring the state of a target's mind and the other relates to inferences about the target's traits. The preceding paragraphs report that people have a great aptitude for determining mental states through a process known as retrodiction. Can people also draw somewhat accurate inferences about personality traits on first meeting someone? If STIs are basic to human processing, it seems that people should be able to infer traits accordingly based on scant behavioral (including speech) information. Indeed, research on spontaneous trait inferences using a false recognition paradigm seems to suggest that people can form an implicit initial impression on another's traits (Todorov & Uleman, 2002; 2003; 2004): They bind STIs to the person who performed a trait-implying behaviour (Todorov & Uleman, 2002) but not to the person who was only paired with the implied trait randomly (Todorov & Uleman, 2003), and such implicit impressions of trait persist after a week-long delay between the formation and the recognition test of that impression (Todorov & Uleman, 2004).

In one example, Todorvo and Uleman (2004) reported a series of studies using a false recognition paradigm. In each trial of each study, participants were presented two pictures of faces with two names (e.g., Judith & Kim) and a behavioral sentence describing an implied trait (e.g. "Judith picked out the best chocolate before the guesses arrived." → implying *selfish*) on a computer screen. In this case, Judith's face was the actor's face and Kim's face was the control face. In the following recognition test, they saw face-trait pairs and were instructed to indicate whether

they had seen the trait in the sentence presented with the face. If participants had drawn an implicit trait inference on the actor in the study trial, then they should tend to indicate that the implied trait appeared in the sentence when the actor's face was paired with the implied trait though in fact it was not true (the trait was not presented in the behavioral sentence). The results demonstrated that participants incorrectly recognized implied traits more when these traits were paired with actors' faces than with control faces. This effect was replicated for a large set of 120 faces, and after delay between study and recognition phase, when equal attention was paid to each face, and when orientation of the face at recognition was different from the orientation at encoding (Todorov & Uleman, 2004). In other words, there is compelling evidence that people form impressions of strangers spontaneously, without having a particular goal or even without being aware that they have made an inference.

However, this does not necessarily mean that people can form an accurate impression on personality of another person. Though it is fairly well established to explore whether and how people can make an implicit personality judgment of the other person, STIs research has not told us how well people can form a first impression of personality in the social world.

Social life presents frequent opportunities to form impressions of strangers; we interact with people every day, and directly observe the activities they perform in their lives and the ways in which they do them. More often, we may have formed an impression of another on the basis of minimal behavioral information, such as a fleeting face, a quick eye gaze, a brief nodding, and so forth. In contrast, inferring traits by reading a sentence or processing trait-implying words might lack external validity for the following reasons. First, people, either perceivers or targets, have

their own personalities and mindsets, which makes them behave in a more imperceptible and complicated way than a presumed person performing a single action in a decontextualized situation. Second, in the STIs paradigm, an inferred trait is always previously matched with a relevant action by the researchers. Yet, according to the concept of global traits (detailed discussion appears in Chapter 2), people have an array of traits, and different traits may manifest as similar actions, and the same behaviour may arise from different or even conflicting traits (e.g., Heyman & Gelman, 1998; Funder, 1991); hence, it seems unlikely that we can map one particular trait onto one particular behavior in everyday life.

In summary, this chapter explored people's ability to make psychological inferences. I began with a brief introduction of the history of research on mentalising, linking the philosophical concerns of mind with the psychological research of mentalising in both theoretical and empirical terms. As such, I described a mentalising capacity (also called ToM) for understanding false beliefs, which develops in the preschool years for the purpose of interpreting and predicting behaviour with reference to mental states. I then provided evidence for the errors and biases of mental state reasoning, and suggested that inferring mental states is not an automatic process. Meanwhile, I proposed that research into reasoning about mental states concentrates too much on the situation but ignores the person; and yet according to attribution theories, people naturally focus strongly on traits when making causal attributions of behaviour. I then discussed people's inferences of traits based on the studies of STIs, and concluded that like mental state inferences, inferring traits also functions as a causal psychological process for behavioral explanation and prediction. Moreover, it seems that inferring traits can occur

automatically. In addition, I emphasized the inherent association between mental state inferences and trait inferences that seems to be implied by simulation theory.

After that, I illustrated lines of research concerning people's ability to reason about mental states while observing a brief sample of behaviour, and ended this chapter by asking the question "How well can people infer personality traits on first meeting?" This question frames the empirical work in this thesis. As have argued, the STIs paradigm is incapable of providing an answer to this question; in Chapter 2, I will articulate an alternative accuracy-oriented approach that will be adopted in the practical studies reported in this thesis.

CHAPTER TWO

Methodological Considerations in Research on First Impressions of Personality

2.1 Introduction

Every day we encounter and meet people from all walks of life in a wide range of social contexts, with many of whom we are unacquainted. By noticing threads of clues in relation to their daily life, such as the bedrooms and the offices they arrange (Gosling, Ko, Mannarelli, & Morris, 2002), the profiles they publish on Facebook (Back, Stopfer, Vazire, Gaddis, Schmukle, et al., 2010), or the music they prefer (Rentfrow & Gosling, 2003), we intuitively form impressions of them in regard to their personalities. By glimpsing fleeting facial expressions (Todorov & Uleman, 2003), or watching the gait of a person (Thoresen, Vuong, & Atkinson, 2012), we swiftly form an impression of personality. As such, through connecting the visible with the invisible, we make sense of the implications of behaviour and perceive other people as individuals. This essential mentalising ability enables us to explain and predict behaviour and thus successfully navigate the complex social world.

How well can we infer another person's personality on first meeting? This problem is important for its theoretical and practical significance. For theoretical

reasons, there are at least three considerations. First of all, as psychologists have noted, the long-standing and controversial dichotomy of person-situation concerning how personality and situational factors contribute in explaining and predicting behaviour has been criticized (e.g. Kenrick & Funder, 1988; Malle, 2011; Funder, 2006; Shiner, 2009). Research on accuracy of personality judgments will be informative in the ongoing resolution of this debate by examining people's ability to infer personality based on a battery of observable behaviour samples in more than one situation (Funder, 2006). Second, empirical research of accuracy in personality judgments will help to build up testable models (Borkenau, Mauer, Rieman, Spinath, & Angleitner, 2004), such as the most notable realistic accuracy model (RAM, Funder, 1995) and weighted average model (Kenny, 1991), serving to decide the underlying factors that influence people' personality judgments and to investigate how these factors play a role in the judgmental accuracy (Borkenau et al., 2004). Last but not least, studying accuracy of personality judgments will help us to test and identify people's implicit folk theories, examining the extent to which people's intuitive theories about the relationship between personality and everyday behavior are accurate (Mehl, Gosling, & Pennebaker, 2006).

For practical reasons, it seems self-evident that personality judgments usually lead to important social consequences in our daily life, either for the judge or for the person who is judged (Funder, 2012). For example, in the social world, people often need to decide who will be only nodding acquaintances and who may be friends after engaging in a brief interaction with each other; in job interviews, employers usually have to quickly determine who will be the ideal candidates even after merely observing the ways in which the candidates introduce themselves. Once made, such judgments usually feed into a decision on whom to trust, befriend, hire, cooperate

with, date, and even marry. Such judgments could therefore vitally affect the quality of people's social lives and their success in the workplaces.

Accuracy is a fundamental concept in science that requires evaluation of validity, reliability, theoretical cogency and many other qualities of data and theory (Funder, 2012). In psychology, the concept of accuracy has been especially challenging in the research of personality judgments because there is an implication that people have stable and enduring traits that determine who they really are (Funder, 2012). But if the implication is wrong then we can hardly expect observers to be able to determine another's personality. This vexing problem has more or less directly guided researchers' scientific endeavors to seek answers to the quest to achieve accuracy of personality judgments in social psychology and personality psychology in a history that goes back more than 70 years (e.g., Funder, 1995; Ambady, Bernieri, & Richeson, 2000).

2.2 Background of Accuracy Research

As early as the 1930s, the pioneer personality psychologist Godorn Allport became very interested in the issue of the accuracy of everyday impressions and judgments of others. Early studies on accurate personality judgment concentrated on the agreement between self-ratings and evaluation of others, in a search of correlates of the "good judge" (Estes, 1938; Taft, 1955; see Ambady et al., 2000, for a review). In 1955, Taft reviewed what was known about people's accurate judgment of others; in the very same year, some researchers, such as Cronbach (1955) and Gage (Gage & Cronbach, 1955), casted doubt on the findings concerning judgmental ability because of severe methodological issues inherent in the existing experimental designs and data-analysis techniques. They argued, for example, self-other agreement (based on a variety of questionnaires) as an index for accuracy, used by all the studies of the time,

might reflect artifacts of base-rate accuracy and artifacts of shared stereotypes between perceivers and targets (Cronbach, 1955; Gage & Cronbach, 1955). Moreover, owing to a lack of reliable and valid assessment of personality, Mischel's (1968) critique of personality research led to questions about whether there is anything to be accurate about (Uleman & Saribay, 2012).

Although the problems that the critics raised were not insuperable, many researchers withdrew owing to the difficulty in establishing accuracy criteria to assess psychological constructs; and they eluded the challenge of solving the problems associated with the extant accuracy measurements (Ambady et al., 2000). Research into accuracy of personality judgments duly waned for several decades since the 1950s primarily for methodological rather than theoretical reasons (Funder, 1995; Jussim, 1991; Kenny, 1994; Ambady et al., 2000; Funder, 2010). Instead, researchers shifted their interest and attention to revealing people's limitations and fallibilities in the process of interpersonal judgment using an error paradigm (e.g., Tagiuri, 1958; Funder, 1995; Ambady et al., 2000), in which participants were asked to draw inferences about traits of hypothetical characters in hypothesized circumstances instead of forming impressions of personalities of real persons in the real life.

In Asch's logically process-oriented approach (Asch, 1946), for example, participants were instructed to form impressions of personalities of imagined characters portrayed by sentences involving trait-related adjectives, by which Asch intended to study the principles that govern the process of personality judgments. Studies like this often focused on investigating how and when people's social perception is biased and erroneous. In the above example, Asch found a primacy effect during the process of impression formation, that is, the earliest words in a list

tend to dominate the impressions people form of the depicted characters. As a result, researchers using the paradigm of cognitive processing were prone to suggest that our first impressions of other people are unreliable and that human inferences drawing on heuristic strategies are filled with shortcomings and errors (e.g., Nisbett & Ross, 1980).

However, "successful as it has been, the 'error paradigm' can tell no more than half of the story" (Funder, 1995, p.652). Some researchers rekindled their interest in the attempt to qualify the accuracy of social judgments in the 1980s (McArthur & Baron, 1983; Swann, 1984; Funder, 1987; Kenny & Albright, 1987). They have realised that the question of accuracy is different from the question of error (e.g. Funder, 1995; 2010): The former is based on critical realism concerning a real person making judgments of another in social life with practical consequences; the latter is based on researchers' presumably ideal model for examining how participants process the artificial stimulus (such as hypothetical target and trait-related words) in the laboratory, which might not necessarily reflect the true nature of human social judgment in the real world. With this agreement, researchers began to develop new methods for addressing accuracy issues raised earlier (Bernieri, Zuckerman, Koestner, & Rosenthal, 1994; Funder, 1995; Kenny, 1994; Snodgrass, 1985), along with developing theoretical frameworks to interpret what factors affect accuracy of personality judgment and how the judgmental accuracy can be achieved (e.g., Funder, 1995; Kenney, 1991). Since then, research on accuracy of personality judgment has reported plenty of novel and intriguing studies, and has become a lively area in the fields of social psychology and personality psychology (Uleman & Saribay, 2012).

2.3 Criteria of Accuracy

Three main criteria are commonly used to evaluate the accuracy of personality judgments. Self-other agreement, as the most often used benchmark for accuracy, refers to the correspondence between a target's self-ratings of some personality traits and a perceiver's judgments on the same traits of the target. It seems reasonable to expect that people know themselves better than anyone else knows them because the self has first-person privileged and direct access to his/her own states of mind. Besides, the self is the only person who experiences his or her life in a diverse range of situations over the life-span (e.g., Funder, 2010). Of course, not any accuracy criterion for personality judgment is perfect. If people sometimes are unwilling to reveal undesirable aspects of their personality or are prone to report socially desirable characteristics for self-enhancement, or even the self-reports of their own actions do not always agree with direct observations (e.g., Vazire & Mehl, 2008), then the accuracy of self-agreement will be weakened.

A criterion used relatively less often is consensus that involves having two or more judges making judgments about some traits of a particular target and then computing their degree of agreement with one another. For example, if more perceivers judge that a person is extraverted, then it is more likely that the person possesses the characteristics of a trait for extraversion, such as talkative and sociable. But not all researchers agree with consensus or reliability of judgments as an appropriate criterion of accuracy. Kenny (1991), for example, has argued that consensus is not equivalent to accuracy though it is closely related to accuracy. If all judges are subject to the same constant bias for a given target, or share a false consensus effect that occurs when we overestimate the extent to which others think,

feel and act as we do (Kilianski, 2008), or the judges cannot achieve consensus, then the consensus-based judgment will be inaccurate.

The so-called "gold standard" of accurate judgment is behavioral prediction. "If a judgment of personality can predict a behavior or a behaviorally-related life outcome, then it would seem likely that it is accurate in some sense" (Funder, 2012). Accuracy of behaviour ratings or predictions is of critical importance for many fields, for example, self-ratings and other-ratings as used to predict health behaviours, job performance, relationship outcomes, and academic performance (Ozer & Benet-Mart ńez, 2006; Roberts, Kuncel, Shiner, Caspi, & Goldberg, 2007). Nevertheless, systematic examinations of the predictive validity of personality ratings are still surprisingly rare in the literature due to the methodological challenges (Vazire & Mehl, 2008; Funder, 2012).

In short, there are pros and cons in the three criteria for accuracy of personality judgment. Given the fact that consensus does not actually measure what a target person' personality is and the criterion of behavioral prediction is faced with difficulty of operation, the current thesis uses self-other agreement as the standard of accuracy: Self-report ratings on personality questionnaires are used as the baseline to determine whether a perceiver's judgment agrees with the self-ratings. Although self-report may be imperfect for measuring personalities, it is the most common technique for examining the behaviours associated with psychological traits, and offers an efficient method to gather wide-ranging information about what people do in daily life (Funder, 2010). Moreover, when it comes to judging personality, it seems essentially important to achieve agreement between self-perception and other-perception: Without this agreement, we would be puzzled at what another person is

like, communicating with others would become utterly perplexing, and friendships would be unlikely.

2.4 Accuracy-oriented Approach to Research of Initial Personality Judgment

To explore people's ability to make an initial judgment of personality that occurs every day, an accuracy-oriented approach is more appropriate as opposed to a process-oriented method given its properties of ecological validity detailed in the following. First of all, according to Funder (1995), accuracy, has attendant consequences for social judgments, is realistic and testable. In line with the studies on STIs (see Chapter 1), we are accustomed to forming first impressions on someone else, and ordinary observation from our daily experiences seems sufficient to verify the fact that we are certainly concerned with whether our initial judgments of a person are correct or not. In addition, unlike the process-based approaches in which perceivers are asked to judge imagined persons in hypothetical situations, the accuracy-oriented methodology focuses on testing people's capacity for making trait inferences on real persons based on a sample of behaviour related to the events and phenomena that probably happen to every ordinary person in real life. In the following, I will articulate the concepts and paradigms involved in the accuracy-oriented approach to study first-impression personality judgments.

2.4.1 Global Personality Traits

According to Allport (1937), traits are psychological mechanisms that determine people's responses to stimuli; they motivate and organize people's behaviour. This definition implies that personality traits can be inferred from observable behaviour. Following Allport's position, Funder (2001; 2007) proposes

that personality traits are an individual's characteristic patterns of thought, emotion and behaviour that are relatively consistent over time and across situations, together with the hidden or visible psychological mechanisms underlying those patterns.

In the social world, almost everyone gets used to thinking about and portraying the people one meets and knows using terms like "friendly," "sociable," "pessimistic," "narcissistic," and the like. As claimed by Funder (1991), "Traits like these are global because each refers not just to one or a few specific behaviours, but patterns of behaviour presumed to transcend time and specific situations" (p.31). A body of evidence has shown that people's global personality affects their personality in specific contexts, their behaviour, and fluctuations in behaviour (Fleeson, 2001; Heller, Watson, & Ilies, 2004).

Aiming to investigate formation of first impressions on personality in daily life, this thesis adopts the concepts of global personality traits because global traits are an important part of everyday social discourse. They encompass a good deal of wisdom and common sense, and offer legitimate, if necessarily incomplete, explanations of behaviour (Funder, 1991). A person has more than one personality trait and behaves differently in different instances; meanwhile, two distinct psychological traits may manifest as the same behaviour and one trait may manifest as different behaviours. This means that though actions express relatively invariant personality traits, particular behaviours do not simply and necessarily express particular traits. In other words, a global trait gives rise to a complex pattern of behaviour from which the trait is inferred, and suggests the psychological mechanisms that are the source of the pattern (Funder, 1991).

2.4.2 Thin Slices of Behaviour

With the aim of understanding the meaningful, consequential, and for the most part social behaviours of daily life, research on interpersonal judgments should focus on direct observation of relevant activities that people perform in daily life situations. However, few studies on personality judgments have paid sufficient attention to direct observation of actual behaviour (Baumeister, Vohs, & Funder, 2007; Vazire & Mehl, 2008; Funder, 2010). Personality psychology has long relied extensively on questionnaire ratings and introspective self-reports; social psychology has moved in recent years to study reaction time (Baumeister, et al., 2007). However, questionnaire responses and laboratory-based behaviours may differ from how a person typically behaves.

The basic principle of behaviour is its continuous flow (Allport, 1937), and people's everyday behaviour usually reflects their multiple traits rather than a single particular trait. This means that disentangling the relationship between a given trait and a given behaviour is extremely difficult (Funder, 1991). A "thin slice" embodies this concept of behaviour. It is a brief excerpt of expressive behaviour drawn from the ongoing behavioral stream, with dynamic information occupying less than 5-minutes (Ambady, & Rosenthal, 1992; 1993; Ambady et al., 2000). According to Ambady et al. (2000):

"Thin slices can be sampled from any available channel of communication, including the face, the body, speech, the voice, transcripts, or combinations of the above. Thin slices remain much, if not most, of the information encoded via dynamic, fluid behaviour while reducing or sometimes eliminating: (a) the information encoded within the ongoing verbal stream; (b) the past history of targets; and (c) the

global, comprehensive context within which the behaviour is taking place" (p.203-204).

A wealth of research has examined judgments based on thin slices pertaining to a wide spectrum of psychological constructs, ranging from informing the affective states (e.g., Waxer, 1976), rating personalities (e.g., Borkenau et al., 2004), to evaluating teacher effectiveness (Ambady & Rosenhal, 1993). Most importantly, the thin-slice paradigm has proved to be ecologically valid as a measure of interpersonal sensitivity (see Ambady, LaPlante, & Johnson, 2001, for a review), providing an efficient means of assessing things in the head, such as others' intentions, affects, emotions, motivation, and personalities.

In tandem with other measures such as self-reports, judgments of thin slice can provide unique insight regarding the dynamics and processes underlying psychological inferences in the real world because thin slice judgments are intuitive and efficient (Ambady, 2010), and because thin slice methodology allows for manipulations in the amount of exposure time and temporal location of the slice presented to judges. This methodology also allows us to manipulate communication channels (including silent videotapes, audiotapes, and standard videotapes) presented to judges, and the types of behaviours that need to be judged from the slices (Ambady et al., 2001).

Given these merits, this thesis pursues a methodology involving 'thin slices of behaviour', defining behaviour as 'thin slices' that are extracted from ongoing observable behavioral streams that occur in naturalistic environments and that are depicted in terms of video, audio and still image. Such "thin slices" are ecologically valid because they derive from part of everyday behaviour in social interactions,

because they are products of global personalities, and because they provide minimal information for people to draw inferences about personalities.

2.4.3 Zero-acquaintance Paradigm

To evaluate the degree to which a personality judgment is accurate, researchers usually compare multiple sources of information about a person, including ratings of strangers and judgments made by the well-acquainted such as friends, spouse, and family. Some studies show that the tendency for observers to agree with a self-report by the target increases with greater acquaintance (e.g., Norman & Goldberg, 1966; Funder & Colvin, 1988; Paulhus & Bruce, 1992); other studies, however, provide no evidence to suggest that agreement increases with increasing acquaintance (e.g., Funder, Kolar, & Blackman, 1995; Kenny, Albright, Malloy, & Kashy, 1994). There is no intention in the current research to investigate how acquaintanceship affects accuracy in personality judgments; instead, the aim is to discover the accuracy of personality judgments and the process involved in making those judgments on first meeting someone else. Thus, the zero-acquaintance paradigm (Norman & Goldberg, 1966; Albright, Kenny, & Malloy, 1988) is adopted where the perceiver is asked to judge a target's psychological traits without any opportunity to interact with the target; that is, the perceiver is unacquainted with and has no prior knowledge about the given target.

Norman and Goldberg (1966) reported the first major study of zero acquaintance. They asked University of Michigan students to rate each other's personality traits on the very first day of class: The students were randomly assigned to a 6- to 9-person group in the absence of any opportunity to interact with one another and without any prior acquaintance. They were asked to independently rate each member of their group and themselves on personality traits. This face-to-face

zero-acquaintance procedure was criticized for its potential contamination of zero acquaintance – Participants may have interacted with one another while making personality judgments in the face-to-face environment. Thus, a video-based zero acquaintance procedure has been widely used in preference, in which perceivers are asked to make judgments on personalities of targets while viewing targets performing activities presented in the format of video.

2.4.4 Summary

To summarize, in order to explore how well we can draw an inference about personality of another person, this thesis adopts the accuracy-oriented approach, using 'thin slices' of behaviour in the context of zero-acquaintance.

In the opinion of Funder (1995; 2012), accuracy in personality judgments is a function of the availability, detection and utilization of relevant behavioral cues. The research in this thesis utilizes self-other agreement as the criterion of accuracy: A target's actual traits are assessed by his/her self-reports on relevant questionnaires, the results of which serve as the standard of accuracy; If perceivers' assessments of the corresponding traits are consistent with the self-report ratings, then accuracy is established. Instead of using artificially trait-laced behaviour, the perceiver is asked to form an intuitive impression of a target after being exposed to thin slices of behaviour that the target performed in realistic scenarios. The "thin slices" are extracted from target's behavioral streams in scenarios of the mundane life. The zero-acquaintance enables perceivers to share the same information about a common target, and on which they rely to form an impression of the target. The procedure of integrating the zero-acquaintance paradigm with the concept of "thin slice" is thus particularly appropriate for studying first-impression judgments of personality.

CHAPTER THREE

Can People Guess How Empathising Another Person Is After Watching a Short Video?

3.1 Introduction

3.1.1 The Role of Empathy in Mentalising

The word "empathy" was coined by Titchener as translation of the German word "Einfühlung", itself a term from philosophical aesthetics meaning "to project yourself into what you observe" (Titchener, 1909). At the end of the 19th century, the notion of empathy was introduced to the philosophy of mind as the primary means for gaining knowledge of other minds, and since then psychologists have taken it as an essential part of psychological events and processes to be studied by empirical methods (Stueber, 2013).

Generally, psychologists distinguish between situational empathy concerning a momentary mental state in a specific situation and dispositional empathy that is regarded as a stable psychological disposition (Zhou, Valiente, & Eisenberg, 2003; Rumble, Van Lange, & Parks, 2010; Stueber, 2013). The situational empathic state is

related to a phenomenon of "inner imitation" where one person mentally mirrors the mental activities or experiences of another person through observing the person's bodily activities or facial expressions in a certain context (Stueber, 2013). In this sense, empathy is thought to function in a way similar to mental simulation. Hence, some psychologists understand empathy as an everyday mindreading ability (that is labeled as "empathic accuracy") to infer the contents of mental and emotional states of other people in given moments (e.g., Ickes, Stinson, Bissonnette, & Garcia, 1990; Ickes et al., 2000; Hall & Mast, 2007; Zaki, Bolger & Ochsner, 2008).

In a typical paradigm of empathic accuracy (e.g., Ickes, 1993), two strangers are led into a waiting room and are asked to wait for the experiment to begin, and then they are left together in the experimenter's absence. During this interval, their verbal and nonverbal behaviours are inconspicuously videotaped. In the main study, each of them is instructed to view a separate videotape of the interaction and make a written, time-logged listing of their own specific thoughts and feelings during the interaction. After that, while watching the videotape a second time, they are required to infer the contents of their partner's thoughts and feelings when the tape is stopped for them at each of those points at which their interaction partner had reported a thought or feeling. Later, independent raters are instructed to code the extent to which the contents of mental and emotional states inferred by the perceiver are similar to the target's self-reports, which serves to establish the measurement of accuracy. Using this face-to-face procedure or a similar video-based task, researchers have found that perceivers can to some extent infer the contents of thoughts and feelings a person was experiencing in some moments in an unstructured dyadic interaction (e.g., Ickes et al., 1990; Ickes, 2003; Hall & Mast, 2007), and they can

also moderately infer how a target person might feel while watching the person talking about autobiographical events (Zaki, Bolger, & Ochsner, 2008; 2009).

As such, we are able to reason about transient mental and emotional states of other people (Ickes et al., 2000; Ickes, 2003; Baron-Cohen, 2012), and we are capable of experiencing an emotion triggered by the emotion of someone else (Baron-Cohen & Wheelwright, 2004; Zaki et al., 2009). According to Baron-Cohen (2012), the capacity for empathy is effective for anticipating and resolving a variety of interpersonal problems; without it, we would lack one of the most valuable resources in our world. If one lacked a capacity for empathy, as might be the case in autism and psychopathy, this could be associated with a severe difficulty in understanding minds of others, leading to difficulties in functioning in the social world (Baron-Cohen, 2012; Flury, Ickes & Schweinle, 2008).

3.1.2 The Present Study

Research on empathic accuracy has addressed the cognitive dimension of empathy that is implicated as the capability of inferring momentary psychological states (Stueber, 2013). However, the psychological architecture of human empathy embodies multidimensional factors in both cognitive and affective terms (e.g., Davis, 1980; Baron-Cohen & Wheelwright, 2004; Decety & Jackson, 2004), which is reflected in the concept of empathic disposition, concerning individual differences in empathy.

That is, people not only experience an empathic state induced by a variety of situations, but also possess empathic disposition that reflects relatively consistent characteristic patterns of behaviour and thought pertaining to empathy; those who have a strong empathizing disposition may experience more empathic states than those who have a weak empathizing disposition. When empathy is understood as a

comparatively stable psychological disposition, is this something that other people can sense? If so, to what extent can people make an accurate judgment on how empathizing another person is on the basis of thin slices of behaviour?

Accuracy research has suggested that people can make an accurate inference about some psychological dispositions on first meeting someone. For example, in lines of research by North, Todorov, and Osherson (2010; 2012), target persons' natural facial reactions to relatively mundane stimuli were recorded unobtrusively while they were reporting which ones they find more appealing. After watching the videos of the targets each lasting several seconds, perceivers had to infer the choices of the targets. The results show that perceivers could somewhat infer the preferences of the targets across four different stimuli categories (people (attractiveness), cartoons (humor), paintings (decorative appeal), and animals (cuteness)) from spontaneous facial expressions alone. Besides, research on personality judgments has suggested that at least in some cases perceivers show noteworthy levels of accuracy in forming a first impression of some dimensions of the Big Five personality traits. For instance, after watching a segment of video where a target read a standard weather forecast, perceivers could form an accurate first impression of the traits of extraversion and conscientiousness (Borkenau & Liebler, 1993); in watching college students having a get-acquainted conversation, perceivers could form an accurate first impression of different factors of the Big Five personality traits in different amounts of exposure time (Carney, Colvin, & Hall, 2007). Together with the findings in the research of empathic accuracy, these data consistently suggest that facial expressions and behavioral manners (including speech) play a role in conveying mental states and psychological dispositions. It seems that observers are

able to perceive and interpret this information to determine the contents of other minds and perceive the dispositions of the target.

Likewise, Empathic disposition, in relation to many domains of our everyday life, could be more or less revealed in empathy-related responding, such as facial expressions, bodily movements and vocal behaviour (Zhou, et al., 2003). Researchers have found that accuracy of emotion recognition is positively correlated with empathic concern following brief exposure to pictures of a target person's facial expression (Besel & Yuille, 2010); and recognition of facial expressions is significantly associated with self-reported emotional empathy (Martin, Berry, Dobranski, & Van Horne, 1996; Gery, Milijkovitch, Berthoz & Soussignan, 2009), trait emotional intelligence involving empathy (Austin, 2004; Petrides & Furnham, 2003), and social-cognitive mindreading tasks (Ferguson & Austin, 2010). This evidence suggests that empathy could leak out into a person's facial expressions and subtle behaviors; therefore an observer might stand some chance of being able to interpret those signs as being indicators that the person is empathic. Can people make use of these indicators to form an accurate first impression on how empathizing another person is after watching a short video?

To seek an answer to this question, the current study developed a novel procedure based on the accuracy-oriented paradigm articulated in Chapter 2. In particular, theoretically, based on the concept of realistic accuracy (Funder, 1995), this study is concerned with accuracy defined as the correspondence between perceivers' inferences and targets' self-report ratings of empathic traits. Operationally, a perceiver is asked to directly guess the target's score on the empathic trait measurement, thereby linking perceiver inferences to objective outcomes. In line with the concept of global traits, instead of using artificially trait-

relevant actions, this research utilizes a “thin slice” to define behaviour, which is extracted from ongoing behaviour happening in the real settings. After viewing segments of behaviour presented in a video, perceivers are asked to judge how empathizing the target is. The zero-acquaintance procedure is used to ensure that perceivers make a judgment on empathic trait on the basis of the presented thin slices of behaviour rather than their previous knowledge about the targets.

Considering its wide application and putative reliability and validity (Lawrence, Shaw, Baker, Baron-Cohen, & David, 2004), this study adopts the ‘empathy quotient’ (EQ) scale, developed by Baron-Cohen and Wheelwright (2004), to measure the trait for empathy. According to Baron-Cohen and Wheelwright (2004), empathy is the ability to understand what another person might be thinking or how a person is feeling, and to respond to the mental and emotional states of the person with an appropriate emotion. Following this understanding, they created the EQ questionnaire, providing a comprehensive measurement of the psychological structure of empathy covering both cognitive and affective factors. It comprises 40 items pertaining to a range of behaviours associated with empathizing, with a total score providing an overall rating of individual differences in empathy tendencies. All targets completed this EQ questionnaire, and their EQs served as the reference point for gauging whether perceivers can guess their scores on the EQ questionnaire.

3.2 Study 1

In the study by Pillai et al. (2012), targets were randomly assigned to experience one of four potential events performed by the researcher, in which the target might be induced to experience an empathic state. For example, in a story scenario, targets heard an empathizing story in which the researcher was relating a series of misfortunes she experienced earlier that day, such as missed the bus to

university, left mobile phone at home and the like. In hearing such a story, targets might be more likely to tune into the researcher's feelings and try to console her, thereby demonstrating an empathic state. Even so, some targets who are less empathizing might not feel sympathetic with the researcher: They came to participate in an experiment but were unexpectedly detained by the researcher's story; thus, they might feel annoyed or unhappy rather than sympathetic. That is, even in the same situation, different people could experience and respond in a different way depending on their capacity for empathy.

According to the study of Pillai et al. (2012), perceivers seemed to be able to draw inferences about targets' momentary empathic states in some contexts. After watching a brief video recording a target's natural reactions to a given scenario, perceivers could fairly accurately guess which scenario the target responded to. Would participants also be able to infer who is more empathic and who is less empathic? Using the same video stimuli, the main goal of Study 1 was to explore how well people could draw inferences about empathic traits of others while watching a 3- to 9-second video clip.

The other purpose of Study 1 was to explore how effectively people form a first impression of empathy of another based on minimal information. According to the simulation theory of mentalising, people sometimes understand others' mental states by mentally simulating others' minds or by projecting their own psychological states and processes into others, especially when accessible information about others is poor and limited (Stich & Nichols, 2003). In one example (Zaki et al., 2009), while watching a person narrating his/her autobiographic emotional events, perceivers experienced similarly emotional experience in themselves, and on this basis they reasoned about the affect of the target. Would people also anchor their empathic

traits to someone else when forming a first impression of empathizing? That is, would those who are highly empathic be inclined to judge others as highly empathic? And would those who are low in empathy tend to consider others as having low empathy?

3.2.1 Method

Summary

Participants (henceforth, 'perceivers') viewed short silent video clips (taken from Pillai et al, 2012) of targets' reactions either to a joke, listening to a person telling them about difficulties they had experienced earlier in the day, being subjected to a frustrating wait or receiving a compliment. The targets completed an EQ questionnaire (Baron-Cohen & Wheelwright, 2004) and each was placed into one of four categories (ranging from low to high EQ) according to their EQ score-range. Perceivers completed the EQ questionnaire as well and received feedback on their EQ category (they were handed a sheet in which their own EQ was circled on a four-point scale, ranging from low to high) before watching the videos and being asked to guess the EQ of each target. The purpose was to determine how well the perceivers could estimate the EQ of the target and whether there was a correlation between perceivers' own EQ and their judgments of the targets' EQs. The procedure was scrutinized and approved by the Faculty of Science ethics committee in the University of Nottingham Malaysia Campus, which was constituted and operated according to the guidelines prescribed by the British Psychological Society.

Participants

Sixty-one students (24 females & 37 males) aged 18 to 30 years (mean age = 21 years, *SD* = 2.40) were recruited from the University of Nottingham Malaysia Campus, including one male who later withdrew. After completing the task, the

perceivers were asked whether they had previously seen any of the targets in the videos. All perceivers denied having any knowledge of the targets.

Materials

A 13-inch MacBook Pro laptop (10.6.8 Mac OSX system) was used to present a sample of video clips using the software PsychoPy (1.70.00 OSX version) (Peirce, 2007). The 40 video clips (21 female & 19 male targets) were obtained from a previous study (Pillai et al., 2012). In these, ten targets in each of four scenarios were unobtrusively filmed; the researcher was not in the frame and not audible:

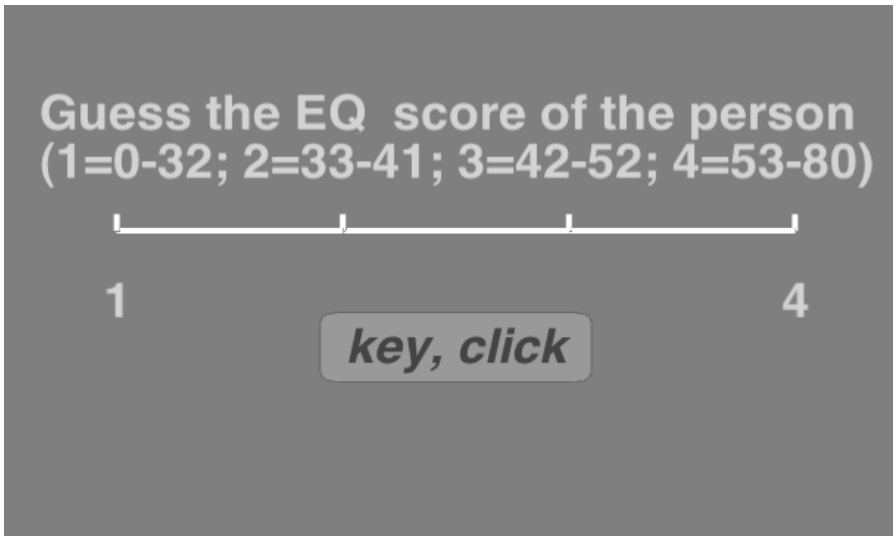
1. Joke Scenario: The researcher told a simple joke to the target.
2. Story Scenario: The researcher related a story about a series of misfortunes she had experienced earlier in the day.
3. Waiting Scenario: The researcher kept the target waiting for 5-8 minutes whilst she was doing personal tasks such as making a phone call and texting.
4. Compliment Scenario: The researcher paid the target a series of compliments.

After being filmed, the targets completed the EQ questionnaire, and their scores ranged from 11 to 58 (M = 41.38, SD = 11.32). A score in the range of 0-32 is low EQ and 8 targets were in this category, 33-52 is average and 25 targets were in this category, 53-63 is above average and 7 targets were in this category, and 64-80 is high but no targets were in this category (Baron-Cohen, 2012). In order to maintain four categories, we split the 'average' category into two categories ranging from 33 to 41 (12 targets) and 42 to 52 (13 targets), and combined the 'above average' and 'high' categories into one range from 53 to 80. We re-label these four categories as Scale 1 (8 targets), Scale 2 (12 targets), Scale 3 (13 targets) and Scale 4 (7 targets), where Scale 1 is lowest EQ and Scale 4 is highest EQ.

Procedure

Perceivers were tested individually and began by completing the EQ questionnaire, which took approximately 10 minutes. After completion of the EQ questionnaire, they were given the definition of empathy and EQ, and told that the questionnaire measures EQ and also told which of the four scales their score fell into, where Scale 1 is low empathizing ability and Scale 4 is high empathizing ability. Subsequently, all 40 video clips were presented on the laptop (600 × 400 pixels) in full colour without sound in a random order determined by the software PsychoPy. Due to the nature of the scenarios, the duration of the video clips varied somewhat, ranging from 3s to 9s (M = 6.03s, SD = 1.25). Following presentation of each video clip a response screen appeared immediately, displaying Scale 1 to Scale 4 (from low to high EQ) as response options (see Fig. 3.1). The perceivers registered their assessment of the target's EQ by using the mouse to click one of the four scales. Once the perceiver made the choice the screen moved immediately to next video clip. Responses were automatically recorded by the software for later retrieval. It took about 10 minutes to complete the video task.

Video presentation screen



Response screen

Fig. 3.1. An example of a trial in Study 1.

3.2.2 Results and Discussion

Preliminary Analysis – The EQ Scores of the Perceivers and the Targets

The EQ scores of the targets (M = 41.38, SD = 11.32, ranging from 11 to 58) were slightly higher than those of the perceivers (M = 37.15, SD = 9.52, ranging from 19 to 65), t (98) = 2.02, p = .05. Strangely, the EQ scores of the targets differed depending on the scenario they had been randomly assigned to according to a one-

way between-groups analysis of variance (ANOVA), F (3, 36) = 4.97, p = .006. Post hoc LSD tests showed that the targets assigned to the story scenario had higher average EQ than those assigned to the joke (p = .014) and the waiting (p = .001) scenarios, but there was no difference between those assigned to the joke and compliment scenarios; and those assigned to the compliment scenario were slightly higher than those assigned to the waiting scenario (p =. 045).

We examined the relationship between the perceivers' self-report EQs and their ratings of the targets' EQs. For each perceiver, the average rating of the 40 targets' EQ was calculated based on the four-point scale (M = 2.19, SD = .38). There was a significant correlation between the perceiver's own EQ (translated into the same four-point scale, M = 2.00, SD = .92) and the guesses they made about the targets, r (58) = .46, p < .001. In other words, an empathic perceiver was inclined to judge that targets were empathic, while a perceiver who lacked empathy tended to judge that targets lacked empathy. In short, it seems to some degree that perceivers' estimations of targets' EQ were anchored to the feedback they had received about their own actual EQ.

Main Analysis – Guessing the EQ of the Target

Adapting the procedure developed by Pillai et al (2012), signal detection (SDT) was used to investigate the accuracy of the perceivers' judgments on which EQ scale each target belonged. This method allows an assessment of accuracy that is independent of underlying base rates of target actual EQ scales and perceivers' response bias in a particular EQ scale. Correct judgments of the targets' EQs on each scale were counted as hits, incorrect judgments on each EQ scale were counted as false alarms, and the index of accuracy was computed as d-prime (d'). Table 3.1 shows the means of the hit rates, false alarm rates, d' in each category of the four-

point scale, along with the corresponding *t* values of one-sample t tests of each d', where the comparison value is zero. If a perceiver guessed at random when estimating the EQ of the target, this would yield a d' of zero.

Table 3.1. Means (and standard deviations) of hit rates (HR), false alarm rates (FAR), d-prime (d') in each EQ scale, and values of one-sample t tests associated with each d' in Study 1

	Scale 1	Scale 2	Scale 3	Scale 4
HR	.40 (.23)	.40 (.18)	.28 (.15)	.18 (.13)
FAR	.23 (.16)	.38 (.15)	.28 (.14)	.09 (.10)
d'	.51 (.57)	.06 (.47)	-.02 (.42)	.50 (.52)
t	6.90*	.96	-.37	7.35*

Note: *p < .001, two-tailed, *df* = 59.

As shown in Table 3.1, the measure of discrimination d' in Scales 1, 2, 3 and 4 was .51, .06, -.02 and .50 respectively, and perceivers were systematically correct in judging the EQs of targets who were in Scale 1 and in Scale 4 but not in Scale 2 and in Scale 3 according to the results of one-sample t tests. The average d' across the four scales (M = .26, SD = .27) was also computed as an indicator of perceivers' ability to guess the EQs of targets at an overall level; the result of a one-sample t test was significant (t (59) = 7.42, p < .001). A repeated measures one-way ANOVA revealed that the d' values among the four scales were significantly different, F (3, 117) = 20.28, p < .001. Post hoc LSD tests showed that d' values were higher in Scale 1 than in Scale 2 (p < .001) and Scale 3 (p < .001). Also, d' values were higher in Scale 4 than in Scale 2 (p < .001) and Scale 3 (p < .001); there was no difference between Scale 1 and Scale 4 and neither was there a difference between Scale 2 and Scale 3.

In summarizing the results of Study 1, perceivers demonstrated a general ability to form an initial impression of another person (indicated by the average d'),

and made systematically correct judgments when targets either had low or high EQ based on videos lasting only about 6s with sound muted. Indeed, perceivers were better at estimating the EQs of targets at the two extremes of the scale than targets who were in the middle two categories. Interestingly, perceivers seemed to assume that targets were somewhat similar to themselves: On average, perceivers' ratings of targets' EQs correlated with their own EQ (as conveyed to them in feedback).

3.3 Study 2

It is tempting to conclude from the results of Study 1 that people are adept in perceiving the trait of empathy in others, especially in cases of high and low empathy. They could do this despite the fact that they do not know the target and despite the fact that their estimation was on the basis of video clips that lasted merely 3 to 9 seconds presented without sound. However, there is an obstacle to such an interpretation. According to the preliminary analysis of the results in Study 1, the EQs of the targets varied according to scenarios, even though the targets were assigned to scenarios at random. How can we explain this unexpected result? One possibility is that targets varied in their EQ across scenarios purely by chance. Another possibility is that the different scenarios affected the way in which the targets filled in the EQ questionnaire. Notably, targets assigned to the story scenario tended to have relatively high EQ while those assigned to the waiting scenario tended to have relatively low EQ. Perhaps the story scenario caused targets to have a sense of empathy, as they consoled the researcher on her difficult day. In contrast, perhaps the waiting scenario caused the targets to feel annoyed and disagreeable in a way that inhibited empathizing. If so, then apparently the target's EQ score reflected not their trait but their state of empathy. We already know from Pillai et al (2012) that perceivers are able to make fairly accurate judgments of the mental state of the

target and now the question arises as to whether the result of Study 1 has merely replicated that finding.

In view of this lack of clarity on whether perceivers rated the target's state or trait, we conducted a further study where all targets experienced the same set of scenarios – Hence, all targets were subjected to the same state. If perceivers can only detect empathy as a state, then in Study 2 they would not be able to estimate EQ systematically. If, in contrast, perceivers can estimate empathy as a trait then they should be able to estimate EQ systematically in Study 2 as well as in Study 1.

3.3.1 Method

Summary

Targets completed the scenarios before they filled in the EQ to ensure that the way they filled in the questionnaire could not have affected their experience of the scenarios. Targets were classified according to the same four-point EQ scale devised for Study 1. A 4 × 3 mixed design was used. Each target experienced three scenarios and they were a brief conversation (the targets answered some questions about themselves), telling a joke (the target told a joke to the camera from a script) and performing a screen test (the target read out an advertisement to the camera about the University of Nottingham). Subsequently, the video clips were presented to perceivers as with Study 1; the perceivers had to guess the EQ of the target. The procedure was scrutinized and approved by the Faculty of Science ethics committee in the University of Nottingham Malaysia Campus, which was constituted and operated according to guidelines prescribed by the British Psychological Society.

Participants

The perceivers were 90 students (49 females and 41 males) aged 18 to 32 years (mean age 21 years, *SD* = 2.40), recruited from the University of Nottingham

Malaysia Campus. To ensure that the perceivers did not know the targets, they were shown photographs of the targets (taken from their videos) and asked whether they had previously seen any of them. Twenty additional participants who reported one or more acquaintances in the sample of the targets were excluded and replaced by another 20. Thus, we assumed that all 90 perceivers who proceeded to the testing phase were unacquainted with the targets. The perceivers were randomly assigned to view targets either in the Conversation, the Joke or the Screen Test Scenario. Details of targets appear below.

Materials

One hundred and forty-one video clips were developed as stimuli, with 47 clips in each condition where the targets were videoed during the conversation, reading a joke or doing the screen test. Therefore, 30 perceivers viewed 47 clips showing the targets in the Conversation Scenario, 30 perceivers viewed 47 clips showing the same targets in the Joke Scenario and 30 perceivers viewed 47 clips showing the same targets in the Screen Test Scenario. All the videos were presented either on the same 13-inch MacBook Pro laptop in Study 1 using the software PsychoPy 1.70.00 OSX version or on a 14-inch HP EliteBook 8460p laptop using PsychoPy 1.74.00 windows version.

Video Stimuli collection and editing

A Sony Handycam DCR-SR60 video camera was used to film targets. Videos were collected from 50 students (targets) from the University of Nottingham Malaysia Campus, including 3 whose data were later deleted because of mistakes in the process of recording. The remaining targets were 24 males and 23 females aged 18 to 32 years (mean age = 21 years, *SD* = 2.85), all of whom responded to a call to do a screen test advertising the University. On arrival, targets were issued with a

script for the joke and the screen test for them to study. All were individually videoed in a quiet laboratory with the camera mounted on the tripod placed approximately 1.2 meters away to record the target's face and the top part of their body. The researcher sat next to the target but out of view of the camera. Unknown to the target, the camera automatically began recording as soon as the target entered the room. Once inside the lab, after the target read some written information (including an information sheet, a script for the joke, a script for the screen test and a consent form), the researcher began with a brief conversation in which she asked a series of questions (and wrote down the responses) about the target's name, age, what course they were enrolled on, where they were from and so on. The conversation lasted approximately two minutes. The camera was then ostensibly switched to 'record mode' and the target was invited to read out the joke to the camera:

"Excuse me, but the seat you've taken is mine."

"Yours? Can you prove it?"

"Yes, I put a cup of ice cream on it."

After a pause of about one minute the target was then invited to read out a verbatim script of the screen test:

"At the University of Nottingham we are committed to providing a truly international education, inspiring our students, producing world-leading research and benefiting the communities around our campuses in the UK, China and Malaysia. Our purpose is to improve life for individuals and societies worldwide. By bold innovation and excellence in all that we do, we make both knowledge and discoveries matter."

After filming and a short break for a couple of minutes, the target was asked to fill in the EQ questionnaire (plus various other questionnaires that were not relevant to the aims of the current study). The EQ scores ranged from 19 to 61 (M = 37.96, SD = 10.19), and each target was classified according to the same four-point scale used in Study 1: Twelve were in Scale 1, 20 in Scale 2, 11 in Scale 3 and 4 in Scale 4.

The video of each target began when he/she entered the laboratory and ended when he/she finished the screen test. From this raw material, three separate video clips were made for each target (the Conversation, the Joke and the Screen Test) using a MacBook Pro laptop with the software Total Video Converter Pro 3.1.8, HandBrake 0.9.8 Mac OSX version and iMovie'09 8.0.6. In the Joke and the Screen Test scenarios, each video clip began when the target started the task and ended about two seconds after the target completed reading the script. The average duration of the video clips was 30.87s (SD = 2.56; ranging from 24s to 35s) for the Conversation, 8.94s (SD = 1.36; ranging from 7s to 12s) for the Joke and 29.36s for the Screen Test (SD = 4.48; ranging from 22s to 42s). Because the raw filming of the Conversation actually lasted around two minutes, we extracted 30-second clips from either the beginning (15 targets), the middle (16 targets) or the end (16 targets) of the conversation videos. In total, we created 47 Conversation, 47 Joke and 47 Screen Test video clips. Each of the three sets of 47 clips was presented to a different group of perceivers (30 in each group) in full colour and with sound.

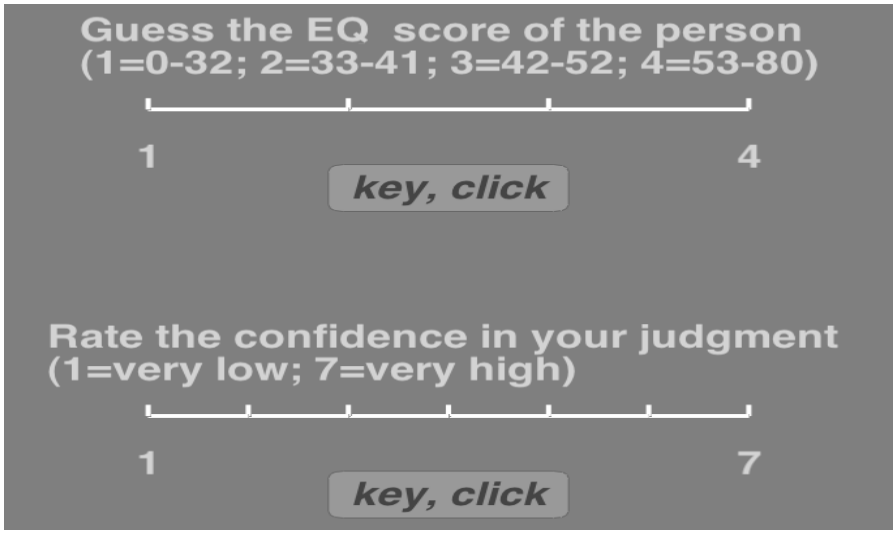
Procedure

The procedure was similar to Study 1 except for the following. First, as with Study 1, perceivers were fed back their own EQ before they judged the EQs of the targets by viewing the video clips. However, in Study 2 the researcher explained that

she did not have time to score the perceiver's EQ and would do it later. After the perceiver finished judging the EQs of the targets while watching the videos, the researcher asked the perceiver to guess his/her own EQ (on the 4-point scale) and to rate how confident he/she felt (on a 7-point scale from very low to very high). Second, in addition to guessing the EQ of the target in each video clip (on the same 4-point scale), perceivers also rated how confident they felt in making each judgment (on the same 7-point scale from very low to very high) (see Fig. 3.2). After presentation of each video clip, a new screen appeared showing the two rating scales, with the four-point EQ scale at the top and the 7-point confidence scale beneath. Perceivers registered their response by using the mouse to click the relevant point on each scale.

The 47 video clips were displayed in a random order to each perceiver (divided into three groups: Conversation, Joke, Screen Test) determined by the software PsychoPy on a laptop. Responses (a four-way forced choice guess of the target's EQ and a 7-way confidence rating) were automatically recorded by the software. Perceivers typically took about 15 minutes in the Joke Scenario and about 40 minutes in the Conversation and the Screen Test Scenario to view and rate the videos.

Video presentation screen ↓



Response screen

Fig. 3.2. An example of a trial in Study 2.

3.3.2 Results and Discussion

Preliminary Analysis – The EQ Scores of the Perceivers and the Targets

The average EQs of perceivers were 38.77 (*SD* = 10.63, ranging from 19 to 58) in the Conversation Scenario, 37.07 (*SD* = 8.31, ranging from 19 to 56) in the Joke Scenario, and 39.47 (*SD* = 8.40, ranging from 23 to 56) in the Screen Test Scenario. Preliminary analyses did not identify any differences between the average

EQs of the targets (M = 38.43, SD = 9.13, ranging from 19 to 58) and the perceivers and neither was there any evidence of difference between the three groups of perceivers in their average EQ scores.

Reassuringly, there was a correlation between the perceivers' actual EQ (converted to the four-point scale) and the guesses they made about their own EQ (on the same four-point scale), r (88) = .46, p < .001, suggesting that to some degree perceivers are aware of how empathizing they are. The actual EQs of the perceivers were not related with the averaged guesses that the perceivers made about the targets' EQs, r (88) = .09, p = .40. But, interestingly, the EQ that the perceivers guessed about themselves was related with the averaged guesses that the perceivers made about the targets' EQs, r (88) = .39, p < .001. The significant correlation between the EQs that perceivers guessed about themselves and guessed about the targets survived even when the actual EQs of the perceivers were partialled out, r (88) = .39, p < .001. As with Study 1, it seems that the EQ the perceivers believed they had impacted upon how they rated the targets. In Study 1, this belief would have been based on information of EQ fed back to the perceiver before he/she made judgments about the targets. In Study 2, in the absence of such feedback, the perceivers' belief about their own EQ was based on their own intuition, an intuition which surfaced at the end of the procedure when perceivers were invited to disclose what they thought their own EQ was.

Preliminary Analysis – Judgmental Confidence

In addition to estimating the EQ of targets and themselves, perceivers also registered how confident they were in making these estimations. The average confidence estimations of EQ judgments about the targets were 4.97 (SD =. 58), 5.09 (SD = .66) and 4.96 (SD = .71) in the Conversation, Joke and Screen Test Scenarios

respectively, and the average judgmental confidence of the perceivers' own EQ was 5.27 (*SD* = 1.11), 5.47 (*SD* = .82) and 5.07 (*SD* = 1.31) in the corresponding three scenarios. All six mean confidence values were higher than the middle point of the rating scale according to one-sample t tests, suggesting that perceivers had a rather positive feeling about being able to guess EQs. Not surprisingly, perceivers were significantly more confident of guessing their own EQ than guessing the EQ of the targets, t (89) = 2.47, p = .015. However, there was no evidence of a significant relationship between perceivers' judgmental confidence and their overall accuracy in judgments of the EQ (indicated by the average d' values). In other words, there was no evidence to suggest that confident perceivers were any better at guessing EQs than perceivers who lacked confidence.

Main Analysis – Guessing the EQ of the Target

The accuracy of the perceivers' ratings of the targets' EQs was analyzed using signal detection. As with Study 1, perceivers' accurate judgments of the targets' EQs on each scale were counted as hits, inaccurate judgments on each EQ scale were counted as false alarms, and the index of accuracy was computed as d'. Table 3.2 displays the means of the hit rates, false alarm rates, d' values in each category of the four-point scale in each condition, and the t values of the one-sample t tests of each d'. The average d' across scales and scenarios was .19 (*SD* = .30), which was significantly above chance according to a one-sample t test, t (89) = 6.01, $p < .001$. However, as Table 3.2 reveals, perceivers were not uniformly effective in guessing the EQs of the targets. As with Study 1, perceivers made systematically correct judgments (indicated by d' well above zero) in the case of low (Scale 1) and high (Scale 4) EQ, while in all but one case (Scale 2 in the Joke Scenario) there was no evidence of systematic judging for targets who were in the middle of the EQ range

(Scales 2 and 3). Fig. 3.3 offers a summary of these results: Perceivers' performance in each condition presented a U-shaped pattern in which d' values were high in the two extremes of EQ scales but low in the two middle EQ score ranges.

Table 3.2. Means (and standard deviations) of hit rates (HR), false alarm rates (FAR), d-prime (d') in each EQ scale, along with values of one-sample t tests associated with each d' in Study 2

	Conversation				Joke				Screen Test			
	S1	S2	S3	S4	S1	S2	S3	S4	S1	S2	S3	S4
HR	.14	.34	.44	.36	.22	.37	.38	.29	.16	.35	.36	.26
	(.13)	(.15)	(.18)	(.27)	(.20)	(.15)	(.19)	(.22)	(.14)	(.12)	(.17)	(.18)
FAR	.10	.34	.40	.16	.16	.39	.39	.11	.12	.36	.38	.17
	(.10)	(.12)	(.15)	(.14)	(.13)	(.17)	(.16)	(.09)	(.11)	(.11)	(.12)	(.11)
d'	.23	0	.08	.78	.31	-.17	-.07	.72	.18	-.01	-.08	.33
	(.44)	(.39)	(.49)	(.65)	(.42)	(.35)	(.47)	(.72)	(.42)	(.28)	(.45)	(.59)
t	2.92**	.02	.87	6.59***	3.97***	-2.75*	-.85	5.49***	2.32*	-.26	-.96	3.04**

Note: S1, 2, 3, 4 = Scale 1, Scale 2, Scale 3, and Scale 4. Three groups (n = 30 in each) of perceivers viewed targets in one of three scenarios (Conversation, Joke, Screen Test). *. p < .05, **.p < .01, & ***. p < .001; two-tailed.

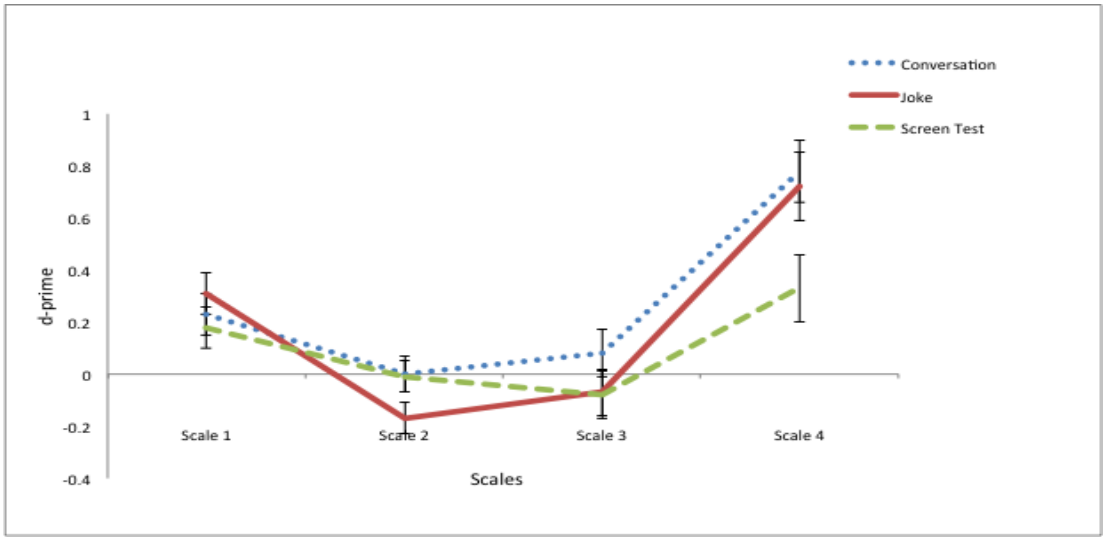


Fig. 3.3. Mean d-prime of each scale in each scenario (Conversation, Joke, & Screen Test) in Study 2. Error bars represent standard error of the mean.

To examine whether the perceivers differed across scales and scenarios in guessing EQ, a 4 × 3 mixed design ANOVA was computed, with the accuracy in the four scales as the within-subjects factor and the three scenarios as the between-subjects factor; the dependent variable was d'. There was a main effect associated with the scales, F (3, 261) = 43.39, p < .001, an interaction between the scales and the scenarios, F (6, 261) = 2.64, p = .017, but no main effect of scenario, F (2, 87) = 2.44, p = .09. Simple effects analyses revealed two things. First, there was a significant difference among the three scenarios only on Scale 4, F (2, 87) = 4.18, p = .018. As we can see in Table 3.2 and Fig. 3.3, d' values were lower in the Screen Test than in the other two scenarios in Scale 4, and this was confirmed by post hoc LSD tests: There were significant differences between the Screen Test and the Conversation Scenario (p = .009), and between the Screen Test and Joke Scenarios (p = .023), but not between the Conversation and Joke Scenarios (p = .74). Considering the fact that the Joke Scenario was much shorter than the scenarios of Conversation and Screen Test while the latter two scenarios were similar in length, it seems that the contents instead of the quantity of the scenario had an effect on perceivers' performance in guessing targets whose EQ fell into Scale 4.

Second, as with Study 1, there were significant differences among the four scales and this trend was apparent for each scenario: (Conversation: F (3, 87) = 17.73, p < .001; Joke: F (3, 87) = 22.79, p <. 001; Screen Test: F (3, 87) = 6.02, p = .001). Post hoc LSD analyses confirmed higher d' values in Scales 1 and 4 compared with Scales 2 and 3 in each of the three scenarios in the following. In the Conversation, there was a marginally greater d' in Scale 1 than in Scale 2 (p = .05) but no difference between the d' in Scale 1 and Scale 3, nor between Scale 2 and Scale 3; the d' in Scale 4 was significantly greater than in the other three scales (ps < .001). In

the Joke, the d' in Scale 1 was greater than in Scale 2 ($p < .001$) and Scale 3 ($p =. 003$); the d' in Scale 4 was significantly greater than in Scale 1 ($p = .007$), Scale 2 ($p < .001$) and Scale 3 ($p < .001$), but there was no difference between the d' in Scale 2 and Scale 3. In the Screen Test, the d' in Scale 1 was greater than in Scale 2 ($p = .037$) and Scale 3 ($p = .014$) but there was no difference between the d' in Scale 2 and Scale 3; the d' in Scale 4 was greater than in Scale 2 ($p = .005$) and Scale 3 ($p = .003$) but there was no difference between Scale 1 and Scale 4.

In summary, in Study 2, perceivers made systemically accurate judgments of EQ of the targets across different situations. As such, it seems that perceivers have the ability to identify others' empathic trait; they are especially good at identifying those who have extreme EQ from those who have average EQ. Generally, perceivers' performance was not influenced by the situations of the targets (Conversation, Joke, Screen Test) though they performed better in Scale 4 in the Conversation and Joke Scenarios than in the Screen Test Scenario.

3.4 General Discussion

3.4.1 How Well Can People Form a First Impression of Others' Empathic Traits after Watching a Short Video?

Overall, the two studies have revealed the following results. First, after watching a video clip lasting between three and thirty seconds, either with sound or without sound, perceivers demonstrated above-chance accuracy in making a first-impression judgment on how empathizing a target person was by identifying the target's EQ. Second, perceivers were especially accurate in identifying the targets who either had low or high EQ, but in most cases failed to recognize the targets with an average EQ (Scales 2 & 3). Study 2 further confirmed that accuracy of the

judgments of the EQ reflects perceivers' ability to infer stable empathic traits rather than a temporary empathic state. In addition, Study 2 revealed no difference in overall accuracy across the situations experienced by each target, but displayed worse performance in the scenario of Screen Test when perceivers guessed the targets with an EQ within Scale 4, though judgmental accuracy in Scale 4 in the Screen Test Scenario was still well above chance.

Previous studies on mentalising have shown that people are capable of identifying empathic *states* of another person: Based on thin slices of observable behaviour, they can infer the contents of thoughts, feelings and emotions that another person experienced (e.g., Ickes, et al., 1990; Ickes et al., 2000, for a review, Hall & Mast, 2007; Zaki et al., 2008; 2009), and retrodict what had happened to another person (Pillai, et al., 2012) or what gift had been received by the other person (Cassidy et al., in press). The findings of the present two studies have extended these findings by showing people's success in inferring empathy as a stable psychological trait measurable with the EQ. The research presented here empirically links judgmental accuracy of empathic traits with research into mentalising using a novel methodology with high ecological validity. Unlike laboratory-based behaviours, such as actions depicted in sentences, posed facial expressions, or deliberate performance, the behaviour samples used in Studies 1 and 2 were gathered from targets' natural behaviours that were observed under various naturalistic circumstances. Thus, we can assume that these behaviour samples were representative of targets' ordinary behaviour patterns in everyday life. That is, instead of using artificial trait-implying behaviour, the behaviour samples in the two studies reflect relatively mundane events that might happen to everybody in the real world. In addition, by asking perceivers to directly guess how empathizing they thought the target was, the

procedure allows perceivers to make their own judgments about the target's EQ, but also permits a direct comparison between the perceivers' judgments and the targets' actual EQs. This procedure closely resembles the way in which ordinary people make judgments about psychological traits of one another in real life.

Although previous studies on accuracy in personality judgments offered evidence for an overall level of accuracy in some personality dimensions, they have provided no examination for people's fine-grained abilities to identify different levels (e.g., low, average, & high) of the same trait. Instead of using correlation analysis, this research adopted a more sensitive and powerful SDT method to analyze the perceivers' judgmental accuracy, and found that (1) perceivers could not guess the targets who had average EQ falling into Scales 2 and 3, and (2) perceivers seemed well-adapted to detecting the targets whose EQ was low or high.

Why did perceivers fail to identify the targets within Scale 2 and Scale 3? Is this result caused by artifact in the way Scale 2 and Scale 3 were derived from but a single EQ category or does it genuinely reflect perceivers' limitations in recognizing the targets who have average empathic ability? As reported in the method section of Study 1, in order to maintain a four-point scale, we split the original average EQ scale into Scale 2 and Scale 3; is there a possibility that such a sub-categorization has no psychological value and that perceivers are thus unable to make a distinction (between Scale 2 and Scale 3) that does not really exist? If this explanation is correct, then perceivers should be able to detect the average EQ when combing Scale 2 and Scale 3 to a single average EQ scale. To examine this possibility, we combined the performance in Scales 2 and 3 in Studies 1 and 2 respectively (the mean combined d' M = .04, SD = .70 for Study 1; M = -.07, SD = .34 for Study 2). One-sample t tests did not show significant differences between the combined d' value and zero in

either of study (Study 1: t (59) = .42, p = .68; Study 2: t (89) = -1.78, p = .08). In the case of Study 2, p value approaches significance but the trend is in the wrong direction. In other words, the trend was for participants to perform below chance in Study 2 for the combined middle scale. Thus, the findings in Studies 1 and 2 indeed have revealed that perceivers were limited in guessing the targets who were average in empathizing.

Perceivers seemed to be good at identifying who lacks capacity for empathizing and who is more capacity for empathizing. Why would this pattern occur in forming a first impression of empathy? There might be two potential explanations. First, targets who had a strong or weak empathizing trait might correspondingly demonstrate more overt behaviour signs. For example, a low empathic target might show a less expressive face, less bodily movements, and weak vocalization. In contrast, a high empathic target might give a happy smiling face after hearing a joke, make more bodily gestures and generally be more demonstrative. Observing these signs, perceivers could "see" who has high empathy and who has low empathy. For example, in hearing the empathizing story described in Study 1, an empathic target might show a sympathetic face while a less empathic target might demonstrate negative facial expressions; based on such conspicuous behavioral cues from the targets, perceivers could stand a chance of distinguishing the targets who had unusual empathic traits from ordinary persons.

Meanwhile, it is worthy to note that the present studies involved emotion-eliciting scenarios (such as hearing an empathizing story or telling a joke) but also low emotion situations, such as a mundane conversation or reading a standard text for the screen test. Even in these less emotionally arousing situations, perceivers still were effective in recognizing those who were high or low in empathy. Nevertheless,

something about those with low and high empathy might reveal itself to the perceivers.

An alternative possibility is that even if those with low and high empathy do not give any more clues to observers about their empathic status, compared with those with average empathy, perceivers are nevertheless perhaps especially well adapted to detecting high and low empathy. Imagine you were traveling in an unfamiliar place where people speak their native languages and have their own customs and cultures. If you were good at recognizing who is empathic and who is not, then you would probably know who you should ask for help and who you should avoid, and thus you might have more chance to survive when facing difficulties. Researchers have argued that a capacity for empathy is associated with moral development (Hoffman, 2000; Batson, Lishner, Carpenter, Dulin, Harjusola-Wevv, Stocks et al., 2003); moreover, a capacity for empathy predicts people's prosocial behaviors, such as altruism (Hoffman, 1984; Eisenberg & Fabes, 1990; Batson, 1991), helping (Batson, O'Quin, Fultz, Vanderplas, & Isen, 1983) and cooperation (Rumble et al., 2010). Taking these factors into consideration, perhaps it is plausible to suppose that people might have evolved to be able to recognize those who have either strong or weak empathizing in the social life.

3.4.2 Assumed Similarity

According to simulation theory, people engage in reasoning about the minds of others by mentally simulating others' mental states and by projecting their own mental states into others. For example, people predict how other people feel in emotionally arousing situations usually based on their predictions of how they themselves would feel in the same situations (Boven & Loewenstein, 2003). Generally it seems that perceivers presume others have similar personalities to

themselves, as suggested by correlations between perceivers' self-report personalities and their ratings on the corresponding personalities of others (e.g., Cronbach, 1955). When information about another person is insufficient and limited, perceivers may utilize the information about themselves to "fill in the gaps" (Ready, Clark, Watson, & Wsterhouse, 2000) and project their own traits onto the other person.

The results of the two studies seemed to reveal an effect of assumed similarity when forming a first impression of empathy. In study 1, this projection effect was indicated as a correlation between perceivers' ratings of the targets' EQs and perceivers' actual EQs, whereas in Study 2, the EQs that perceivers judged about the targets were associated with the EQs that perceivers guessed about themselves. In other words, Study 2 shows that how perceivers think of themselves as an empathizing person affects how they perceive other people in terms of empathizing. If a perceiver believes he is high in empathic capacity, then he tends to judge other people to be high also; if a perceiver thinks he is low, then he is prone to evaluate other people as having low empathy as well.

3.4.3 Confidence and the Judgment of EQ

How confident do you think you have formed an accurate first impression of someone else? This question is important and likely adaptive in that judgmental confidence would influence the consequential effects of judgmental accuracy, such as in effective interpersonal functioning (Carlson, Furr, & Vazire, 2010). For example, if you were confident in your intuition of someone and it proved to be inaccurate, you might make the mistake of trusting the wrong person; if an employer were confident of his judgment of a candidate but that judgment was in fact incorrect, then he might have made the mistake of recruiting an inappropriate employee.

In a study by Carlson and colleagues (2010), after engaging in a 5-minute conversation, two unacquainted participants were asked to evaluate the Big Five personality traits of their partner and themselves, and then were asked to rate confidence in their first impressions of the partner using a 7-point scale. The results showed that perceivers who had more confidence in the accuracy of their first impressions of others were actually more accurate. However, Study 2 reported here shows that the overall accuracy of perceivers' first-impression judgments of the targets' EQs was not significantly correlated with perceivers' confidence in such judgments. The inconsistency between this study and Carlson et al. (2010) may result from a different calculation of accuracy in personality judgments: Carlson et al. indexed accuracy using a Pearson correlation while the current study used the more sensitive SDT method to indicate accuracy.

On the other hand, Ames, Kammrath, Suppes, and Bolger (2010) suggest a dissociation between accuracy and confidence in thin-slice impressions: In three studies of first impressions based on photos and videos, they examined the accuracy of first impressions of the Big Five personalities as well as corresponding reports of confidence, and found that perceivers showed a limited ability to intuit which of their first impressions were more accurate than others. These results are consistent with our findings in Study 2, both of which tend to support the conclusion drawn from the study by Realo et al. (2003), suggesting that self-reported mindreading ability is not associated with actual performance in the mindreading tasks. Those who believe that they are good at mindreading are generally neither significantly better than others in the recognition of emotions expressed in face or speech, nor superior in their estimation of personality traits of a stranger (Realo et al., 2003).

In addition, the results in Study 2 demonstrated that perceivers were inclined to trust their beliefs about first impressions of empathizing they made and also have fairly high confidence in the self-ratings of their own empathizing; indeed, they tended to be more confident of guessing their own EQ than guessing the targets' EQ. These findings seem to confirm some of our intuitive impressions of how people perceive others in the social world. First, people generally believe they know themselves better than others; second, once a judgment is made, either about the self or about others, people to some extent tend to trust the judgments even when these judgments are only based on thin slices of observable behaviour.

3.5 Conclusion

In conclusion, this chapter reported two studies based on a large sample of video stimuli, demonstrating that people can guess how empathizing another person is after viewing a short video spanning from several seconds to thirty seconds, and they are especially effective in guessing those who are either low or high in empathic capacity. In addition, how people perceive themselves as empathizing has an effect on their judgments of the empathy of another person. Finally, people generally are aware of their perception of themselves and others when it comes to making judgments about empathic traits, though they are more confident of self-perception than other-perception. Meanwhile, their judgmental confidence does not predict their judgmental accuracy of empathic traits.

CHAPTER FOUR

Guessing Empathy After Brief Exposure to Photographs or a Brief Sample of Sound

4.1 Introduction

In the social world, we often form an intuitive impression on other people in many ways. We might happen to encounter and communicate with a person; we might happen to catch thin slices of behaviour; we might also happen to spot a photograph or hear snippets of the voice of someone else as in a telephone conversation. From whatever channels we get information of other people, we would probably have formed an impression on them. Some of our intuitions based on such sparse information may be reliable while others may be inaccurate (Ames et al., 2010).

In Chapter 3, based on two studies investigating first impressions of empathic traits, it has been suggested that perceivers can to some degree guess the EQs of the targets after watching a sample of behaviour lasting only a few seconds, and they are especially good at identifying the targets who were low or high in empathic capacity. How did perceivers infer the empathic capacity of the targets? Perhaps it is necessary for perceivers to see an animation of the behaviour of the target. Another possibility is that perceivers can make accurate assessments even in the absence of seeing an

animation. It seems fair to assume that the target's behaviour reveals their empathic capacity but a still photograph of the target engaging in some kind of behaviour might be sufficient for perceivers to make inferences accurately.

If animation of the behaviour is necessary for making an accurate judgment of the EQ, then perceivers should not be able to accurately guess the EQs of the targets when the actions are presented in static form, such as the still photographs of a target performing a certain activity. If behaviour of the target (whether or not animated) plays an essential role in allowing the perceiver to form an accurate impression of empathy, then perceivers should be equally good at guessing the EQs of the targets whether the behaviour is presented in dynamic or static form. Conversely, if still images of the targets not engaged in some kind of behaviour provide sufficient information, then perceives should be able to draw inferences about empathic traits even when the image does not give any information about behaviour (as in a passport photograph). Study 3 and Study 4 were designed to explore these possibilities, and examine how well perceivers can guess the EQ after viewing pictures for several seconds.

Finally, how important is sound? Would perceivers be able to guess the empathic capacity of the target merely after hearing the target speaking for a few seconds? Study 5 addressed this question.

4.2 Study 3

In order to examine whether perceivers' capability of inferring empathic traits is based on the animation of the stimuli or the actions of the targets, perceivers were asked to guess the target's EQ while either viewing a short video or watching three sequential photographs taken from the same video. If the animation was critical and sufficient for making an accurate judgment of the EQ, then perceivers should have

difficulty in guessing the EQs of the targets while having access to only static photographs.

4.2.1 Method

Summary

For the sake of simplicity we only used the Joke Scenario in the present study because the results in Study 2 were very clear despite the fact that the video clips of target behaviour for the Joke Scenario were much shorter than for the other two scenarios. A 2 × 4 mixed design was adopted, with the two information channels (video & picture) as the between-subjects factor and the four EQ scales (based on targets' responses to the EQ questionnaire) as the within-subjects factor. After either viewing the video clip or the three sequential photographs taken from the video clip, the perceiver was required to guess the EQ of the target.

Participants

Sixty students (27 females and 33 males) aged 19 to 27 years (mean age 21 years, *SD* = 2.08) were recruited from the University of Nottingham Malaysia Campus. To ensure that the perceivers were unacquainted with the targets, they were shown photographs of the targets (taken from the videos) and asked if they knew any of them before proceeding to the task of empathic trait judgments. Sixteen additional participants who declared that they did were excluded. Perceivers were randomly divided into two groups of 30 to view either video clips or photographs.

Materials and procedure

The 47 video clips in the Joke Scenario taken from Study 2 (see Chapter 3) were used with sound muted. The set of photographs was derived from these same videos. For each target, three photographs were extracted from his/her video clip,

with each photograph corresponding to the beginning, the middle and the end of the joke video. Each photograph was trimmed in the software Drawing to standardize the size. Using the software Windows Movie Maker, each target's three photographs were combined into one single video in which the three pictures were presented sequentially for three seconds each. Thirty perceivers viewed 47 animated video clips (the video condition) and another 30 perceivers viewed 47 photograph video clips (the photograph condition). All the video stimuli were displayed in 800 × 650 pixels on the HP ElieBook 8460p laptop using the software PsychoPy (1.74.00 windows version). The procedure was the same as that in Study 1 (see Chapter 3) except that the perceivers did not receive any feedback of their own EQ.

4.2.2 Results and Discussion

Preliminary Analysis – The EQ Scores of the Perceivers and the Targets

The perceivers' average EQs were 38.17 in the video condition (*SD* = 10.66, ranging from 19 to 68) and 39.66 in the picture condition (*SD* = 11.03, ranging from 14 to 57). Preliminary analyses did not identify any differences between the average EQs of the targets (*M* = 38.43, *SD* = 9.13, ranging from 19 to 58) and the perceivers and neither was there any evidence of difference between the two groups of perceivers in their average EQ scores.

Main Analysis – Guessing the EQ of the Target

As with the previous studies reported in Chapter 3, perceivers' guesses of the targets' EQs were coded using signal detection. Table 4.1 displays the means of the hit rates, false alarm rates, d' values in each category of the four-point scale in each condition, and *t* values of one-sample t tests for each d'. The average d' across scales in the video condition was .22 (*SD* = .28) and the average in the picture condition

was .15 (*SD* = .22). Both values were significantly higher than zero, suggesting that overall perceivers could systematically estimate the EQs of the targets: Video: t (29) = 4.24, $p < .001$; Picture: t (29) = 3.71, $p = .001$. Table 4.1 shows a pattern of performance that is consistent with the previous two studies, where perceivers made systematically accurate judgments in the cases of low (Scale 1) and high (Scale 4) EQ. This pattern maintained for the picture as well as the video conditions. There was no evidence of perceivers systematically estimating the EQs of targets who were in the middle ranges (Scales 2 and 3) except in one case (scale 2 in the picture condition). These results were also quite clear from Fig.4.1, which depicts the distribution of accuracy in the conditions of video and photograph across the four scales, demonstrating U- shaped trends, like Fig.3.3 in Study 2 (see Chapter 3).

Table 4.1. Means (and standard deviations) of hit rates (HR), false alarm rates (FAR), d-prime (d') in each EQ scale, and *t* values of one-sample t tests associated with each d' in Study 3

	Video				Picture			
	Scale 1	Scale 2	Scale 3	Scale 4	Scale 1	Scale 2	Scale 3	Scale 4
HR	.34	.36	.30	.25	.34	.31	.29	.25
	(.19)	(.14)	(.16)	(.21)	(.19)	(.14)	(.12)	(.15)
FAR	.19	.39	.33	.10	.21	.37	.32	.13
	(.09)	(.10)	(.11)	(.07)	(.11)	(.11)	(.10)	(.09)
d'	.44	-.09	-.12	.63	.41	-.19	-.12	.52
	(.47)	(.41)	(.48)	(.58)	(.42)	(.43)	(.47)	(.43)
t	5.06**	-1.18	-.1.30	5.94**	5.32**	-2.44*	-1.42	6.59**

Note: Two groups (*n*=30 in each group) of perceivers each viewed targets in one of two conditions (Video, Picture); *.$p < .05$, & **. $p < .001$, two-tailed.

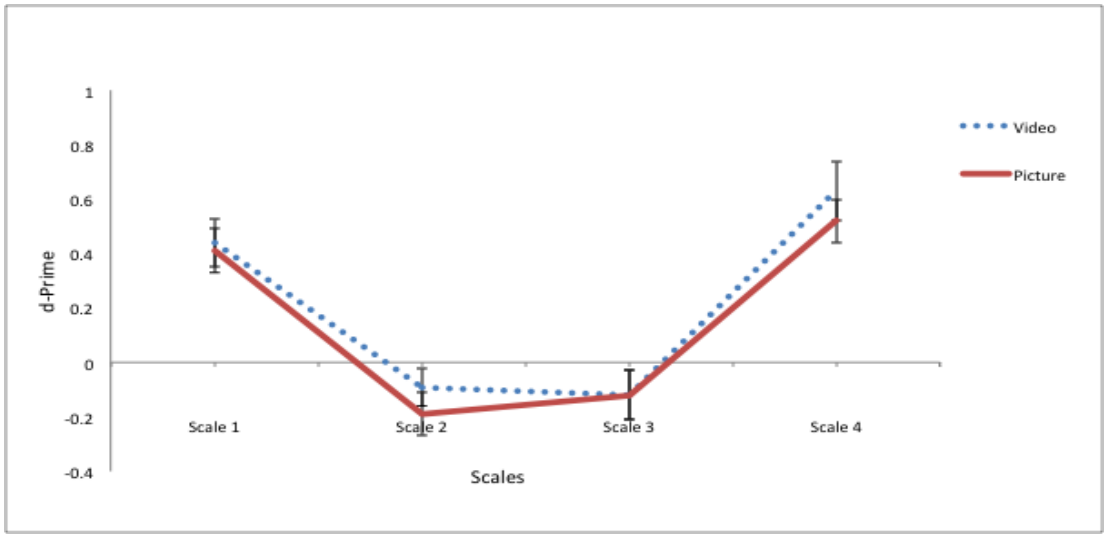


Fig. 4.1. Mean d-prime of each scale in each condition (Video – Picture) in Study 3. Error bars represent standard error of the mean.

To examine whether perceivers performed differently across EQ scales and conditions in guessing the EQs of the targets, a 2 × 4 mixed design ANOVA was computed, with the two conditions (Video – Picture) as the between-subjects factor and the four EQ scales as the within-subjects factor; the dependent variable was d'. There was a main effect associated with the scales, F (3, 261) = 37.84, $p < .001$, but neither a main effect of condition, F (1, 58) = 1.08, $p = .30$, nor an interaction between the scales and the conditions, F (3, 174) = .20, $p = .90$. Post hoc LSD tests of the main effect associated with the scales revealed higher d' values in Scale 1 than in Scale 2 ($p < .001$) and Scale 3 ($p < .001$). Also, d' values were higher in Scale 4 than in Scale 2 ($p < .001$) and Scale 3 ($p < .001$); there was no difference between Scale 1 and Scale 4 and neither was there a difference between Scale 2 and Scale 3.

In summary, Study 3 replicated the results of the earlier two studies: The perceivers generally performed well above-chance in guessing the EQs of the targets, and were especially effective in making judgments about the targets who had high or low EQ. These basic findings were supplemented with a surprising new result.

Namely, perceivers were also able to correctly guess the EQs of the targets while merely viewing three photographs of each target for several seconds. Indeed, there was no evidence suggesting that perceivers performed any better in the video than in the picture condition. In other words, observing limited information based on pictures seems sufficient for perceivers to accurately form a first impression on how empathizing another person is. Granted, the still photographs *suggested* behaviour in so far as they were captured from three moments of the targets' actively engaging in a task of telling a joke, which may be quite different from what we might see in a passport photograph. Perhaps the latter would not provide sufficient information for a perceiver to successfully estimate EQ, raising a question that needs to be addressed in future research. In contrast, the photographs taken from the Joke Scenario might have conveyed some revealing behaviours of the targets, such as facial expressions and bodily gestures. These findings thus raise the possibility that it is the behaviour rather than the animation that led to perceivers' accuracy in making judgments of EQ, whether the behaviour is presented in dynamic streams or in still photographs.

4.3 Study 4

The results of Study 3 were surprising in suggesting that perceivers were systematically able to identify targets with high and low EQ after looking for a few seconds at three still pictures taken from video clips. On what basis were perceivers able to make correct judgments? Is it that a still image of the target in any pose, including a neutral pose, providing sufficient information to identify high and low EQ? In other words, is it possible that perceivers made accurate judgments of EQ based only on the facial appearance of the target instead of behavioral cues suggested by the facial expressions of the target? Alternatively, is it that a photograph capturing the apex of the target's expression as he or she delivers the punch line of a joke

uniquely revealed those who had high and low EQ? If the former, then perceivers should perform well in identifying targets with high and low EQ whether the still picture was at a point when the target delivered the punch line or at an earlier point in the video when we might suppose the target was less expressive. If the latter, then perceivers should be able to identify cases of high and low EQ on condition that the still photograph was at a point when the punch line was delivered but not at any other point. The purpose of Study 4 was to clarify this matter.

4.3.1 Method

Summary

Two sets of photographs drawn from the joke video clips were created, corresponding to the first and the last moments of each video clip of each target when they read the joke. A 2 × 4 mixed design was adopted, with the two conditions (first and last photographs) as the between-subjects factor and the four EQ scales as the within-subjects factor. Perceivers were randomly assigned to view the first photograph or the last photograph, and then were asked to judge the EQ of the target.

Participants

Sixty students (27 females and 33 males) between 18 years old and 25 years old (mean age 21 years, *SD* = 1.37) were recruited from the University of Monash Sunway Campus. Perceivers were randomly assigned to two groups of 30 to view an array of photographs either in the first or the last photograph condition. After completing the task, the perceivers were asked whether they had previously met any of the targets. All the perceivers reported no prior acquaintance with any targets.

Materials and Procedure

Two groups of 47 photographs were taken from the videos in the Joke Scenario. For each target, two photographs were extracted, in each of which the target was either in the beginning of reading the joke or at the end of the joke (the punch line) (see Fig. 4.2). To match the endurance of the joke videos, the two photographs in each condition appeared for 9 seconds in total. All the picture stimuli were displayed in full colour on the 14-inch HP laptop using the software PsychoPy (1.74.00 windows version). The procedure was similar to that in Study 3.

The first photograph

The last photograph

Fig. 4.2. An example of the first and the last pictures in Study 4.

4.3.2 Results and Discussion

Preliminary Analysis – The EQ Scores of the Perceivers and the Targets

The perceivers assigned to the first photograph condition had a mean EQ of 38.17 (*SD* = 10.73, ranging from 19 to 54) while those assigned to the last photograph condition had a mean of 35.13 (*SD* = 8.90, ranging from 18 to 54). Preliminary analyses did not reveal any differences between the mean EQs of the

targets and the perceivers and neither was there any evidence of difference between the two groups of perceivers in their mean EQ scores.

Main Analysis – Guessing the EQ of the Target

The procedure of coding based on signal detection was the same as that used in the previous studies. Table 4.2 presents the means of the hit rates, false alarm rates, d' values in each category of the four-point scale in each condition, and *t* values of one-sample t tests for each d'. The average d' across scales in the first photograph condition was .06 (*SD* = .25) – not significantly above chance; in the last photographs condition the average d' was .17 (*SD* = .25) and this value was significantly above chance according to a one-sample t test (t (29) = 3.82, p = .001). As we can see from Table 4.2 and Fig. 4.3, perceivers were better at identifying the targets with low (Scale 1) or high EQ (Scale 4) than those with average EQ in the last photograph condition. In the first photograph condition, where perceivers viewed the photograph capturing the beginning moment of the target reading the joke, this pattern was diminished.

Table 4.2. Means (and standard deviations) of hit rates (HR), false alarm rates (FAR), d-prime (d') in each EQ scale, along with *t* values of one-sample t tests of each d' in Study 4

	First Photograph Condition				Last Photograph Condition			
	Scale 1	Scale 2	Scale 3	Scale 4	Scale 1	Scale 2	Scale 3	Scale 4
HR	.15	.40	.37	.17	.28	.33	.35	.24
	(.13)	(.15)	(.15)	(.11)	(.22)	(.14)	(.17)	(.15)
FAR	.14	.40	.36	.13	.15	.39	.35	.13
	(.10)	(.10)	(.12)	(.08)	(.11)	(.12)	(.12)	(.08)
d'	.09	-.03	-.01	.19	.44	-.19	-.04	.47
	(.35)	(.36)	(.38)	(.54)	(.40)	(.39)	(.42)	(.42)
t	1.44	-.39	-.18	1.97	6.02**	-2.64*	-.56	6.20**

Note: Two groups (*n*=30 in each group) of perceivers each viewed targets in one of two conditions; *.*p* = .01, **. *p* < .001; two-tailed.

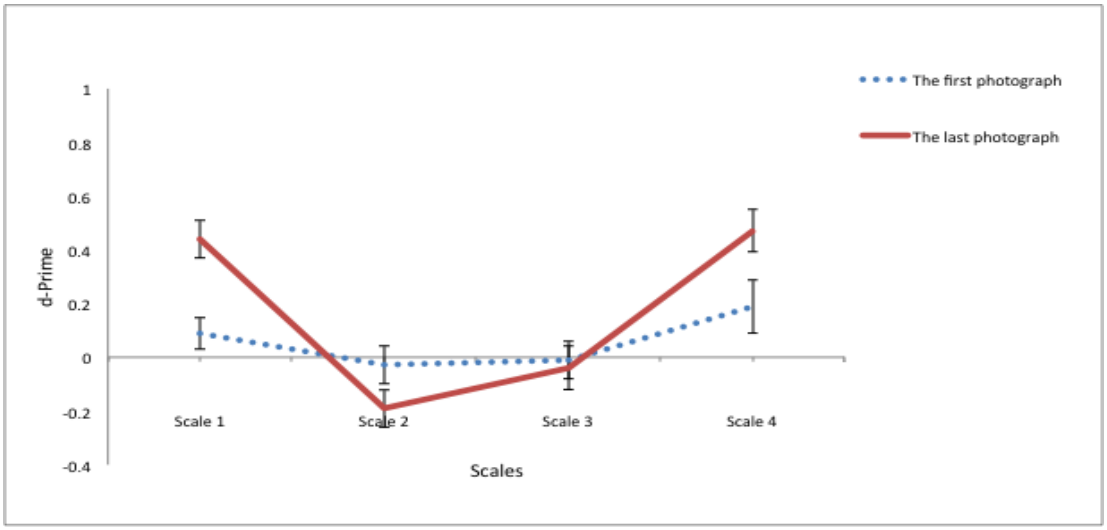


Fig. 4.3. Mean d-prime of each scale in each condition in Study 4. Error bars represent standard error of the mean.

A 2 × 4 mixed design ANOVA was carried out, with the two photograph conditions as the between-subjects factor and the four EQ scales as the within-subjects factor; the dependent variable was d'. There was a main effect associated with the scales, F (3, 174) = 19.64, p < .001, a main effect of condition, F (1, 58) =

13.35, p = .001, and an interaction between the scales and the conditions, F (3, 174) = 6.31, p < .001. Simple-effects analyses revealed the following. First, the significant effect associated with the four EQ scales was only found in the last photograph condition, F (3, 87) = 24.02, p < .001. Post hoc LSD analyses in the last photograph condition confirmed greater accuracy in Scales 1 and 4 as opposed to Scales 2 and 3 (ps < .001), and there was no evidence of difference in d' between Scales 1 and 4 (p = .60) nor was there any evidence of difference between Scale 2 and Scale 3 (p = .18). Second, as shown in Table 4.2 and Fig. 4.3, the d' values were much higher in the last photograph condition than in the first photograph condition in Scale 1 (t (58) = 3.59, p = .001) and Scale 4 (t (58) = 2.27, p = .027); there was no difference between the two conditions in Scale 2 and Scale 3.

In summary, the results in the last photograph condition replicated the same U-shaped pattern we found in the previous studies. However, the evidence for such a U-shaped pattern was not compelling for the first photograph condition. It seems therefore that information from the target's delivery of the punch line of the joke was sufficient for perceivers to infer high and low EQ; information of a neutral kind from a photograph where the target was merely reading text before he or she reached the punch line apparently was not revealing of the target's EQ status. In other words, when a static picture involves a task performed by a person and depicts the person's actions such as facial expressions and bodily movements, perceivers have an opportunity to be able to guess the empathic capacity of the person, whether it is high or low. In comparison, in observing a picture that is less related to any activity or demonstrates only neutral poses, such as a passport photograph, or the photograph capturing the very first moment of telling a joke, perceivers have little access to

behavioral cues and can not determine who is more empathising and who is less empathising.

Taken together, Study 3 and Study 4 demonstrate that perceivers can accurately make judgments of targets who were low or high in empathy after briefly observing behaviour, and performance is equally good whether the visual cues are still or animated. These results help us to rule out an alternative explanation of the successful performance of perceivers. It could have been that people construct their images and create their persona such that they were perceived as either strong or weak in empathizing. If so, perceivers should still be able to detect who has high EQ and who has low EQ after being exposed to still images that do not suggest behaviour. However, the data in Study 4 do not support this explanation. Instead, it seems that perceivers rely on information about the target's behaviour when estimating empathic capacity.

4.4 Study 5

The previous four studies demonstrated perceivers' capacity for inferring empathic traits on the basis of visual behaviour cues. Would perceivers also be able to identify who has high EQ and who has low EQ after listening to the target talking for a few seconds? Previous studies have indicated that people can sometimes predict others' daily behaviours after hearing fragments of sound unobtrusively recording their daily lives (Holleran, Mehl, & Levitt, 2009); people can also infer others' emotions while hearing them talking about their life experiences (Zaki, Bolger, & Ochsner, 2009). However, if the talking does not relate to an individual's personal life but is merely reading aloud a couple of lines, as in the case of the Joke Scenario, would perceivers be able to make an accurate judgment of the EQ of the target? The objective of this study was to tackle this question.

4.4.1 Method

In the attempt to identify the scope of evidence of perceivers' ability to estimate the EQ of targets, Study 5 presented a new condition in which perceivers could only hear the soundtrack of the Joke Scenario. Is it the case that being able to perceive EQ depends on having visual access to the target or is auditory evidence sufficient? Some researchers have insisted that the face, especially the eyes, is the principal source of psychological information (e.g., Baron-Cohen, Jolliffe, Mortimore, & Robertson, 1997). If they are right, then we should expect perceivers to be much more accurate in the video than in the auditory conditions that are described below.

Summary

The video stimuli (Joke Scenario without sound) were those used in Study 2. Sound tracks were extracted from these same videos for presentation in the audio condition. A 2 × 4 mixed design was adopted, with the two information channels (video & audio) as the between-subjects factor and the four EQ scales as the within-subjects factor. After either viewing the video clip or listening to the target telling a joke, the perceiver was required to guess the EQ of the target.

Participants

Sixty students (28 females & 32 males) aged 18 to 23 years (mean age 20 years, *SD* = 1.60) were recruited from the University of Nottingham Malaysia Campus. To ensure that the perceivers were unacquainted with the targets, they were shown photographs of the targets and asked if they knew any of them. Twelve who responded positively were excluded and replaced by a further 12 who did not know the targets, thus giving a working sample of 60. Perceivers were randomly divided into two groups of 30 to either view video clips or listen to targets telling a joke.

Materials and Procedure

The 47 video clips in the Joke Scenario of Study 2 were used as the set of visual stimuli. The auditory stimuli were separated from the same video clips using the MacBook Pro laptop with the software Total Video Converter Pro 3.1.8, thus yielding 47 samples of audio stimuli. Thirty perceivers viewed 47 video clips (Video Condition) and another 30 heard 47 audio tracks (Audio Condition). The video stimuli were displayed in the size of 800 × 650 pixels on the 14-inch HP ElieBook 8460p laptop using PsychoPy 1.74.00 windows version, and the audio stimuli were presented on the 13-inch MacBook Pro laptop using PsychoPy 1.70.00 OSX version. The procedure was similar to that in the previous studies.

4.4.2 Results and Discussion

Preliminary Analysis – The EQ Scores of the Perceivers and the Targets

The perceivers' average EQ was 38.63 in the video condition (*SD* = 8.30, ranging from 22 to 54) and 36.60 in the audio condition (*SD* = 9.11, ranging from 22 to 54). Preliminary analyses did not identify any differences between the average EQs of the targets and the perceivers and neither was there any evidence of difference between the two groups of perceivers in their average EQ scores.

Main Analysis – Guessing the EQ of the Target

As with the calculation of previous studies, perceivers' guesses of the targets' EQs were coded using signal detection. Table 4.3 displays the means of the hit rates, false alarm rates, d' in each category of the four-point scale in each condition, and values of one-sample t tests for each d'. The average d' across scales was .27 (*SD* = .23) in the video condition and was .14 (*SD* = .25) in the audio condition, both of which were significantly above chance according to one-sample t tests (Video, *t* (29)

= 6.40, $p < .001$; Audio, t (29) = 3.14, $p = .004$), suggesting that in either condition, perceivers can make an overall accurate judgment on the EQs of the targets. Furthermore, perceivers made systematically correct judgments in the case of low (Scale 1) and high (Scale 4) EQ in the video condition, whereas in the audio condition perceivers systemically estimated the EQs of targets only in the case of high EQ (Scale 4). There was no evidence in either group of perceivers systematically estimating the EQ of targets who were in the middle categories (Scales 2 and 3). As shown in Fig. 4.4, only d' values in the video condition yielded an apparent U-shaped curve while d' values in the audio condition presented a flat horizontal line close to chance level from Scales 1 to 3 and sharply increased in Scale 4.

Table 4.3. Means (and standard deviations) of hit rates (HR), false alarm rates (FAR), d-prime (d') in each EQ scale, along with *t* values of one-sample t tests of each d' in Study 5

	Video				Audio			
	Scale 1	Scale 2	Scale 3	Scale 4	Scale 1	Scale 2	Scale 3	Scale 4
HR	.33	.34	.33	.29	.22	.34	.32	.31
	(.24)	(.14)	(.18)	(.17)	(.13)	(.12)	(.14)	(.22)
FAR	.16	.37	.34	.12	.21	.35	.30	.15
	(.12)	(.11)	(.12)	(.09)	(.13)	(.13)	(.09)	(.11)
d'	.54	-.11	-.08	.71	.0	-.02	.01	.57
	(.46)	(.44)	(.50)	(.52)	(.48)	(.37)	(.44)	(.55)
t	6.44*	-1.34	-.94	7.52*	.01	-.30	.16	5.74*

Note: Two groups (n=30 in each group) of perceivers each viewed targets in one of two conditions (Video, Audio); *. $p < .001$, two-tailed.

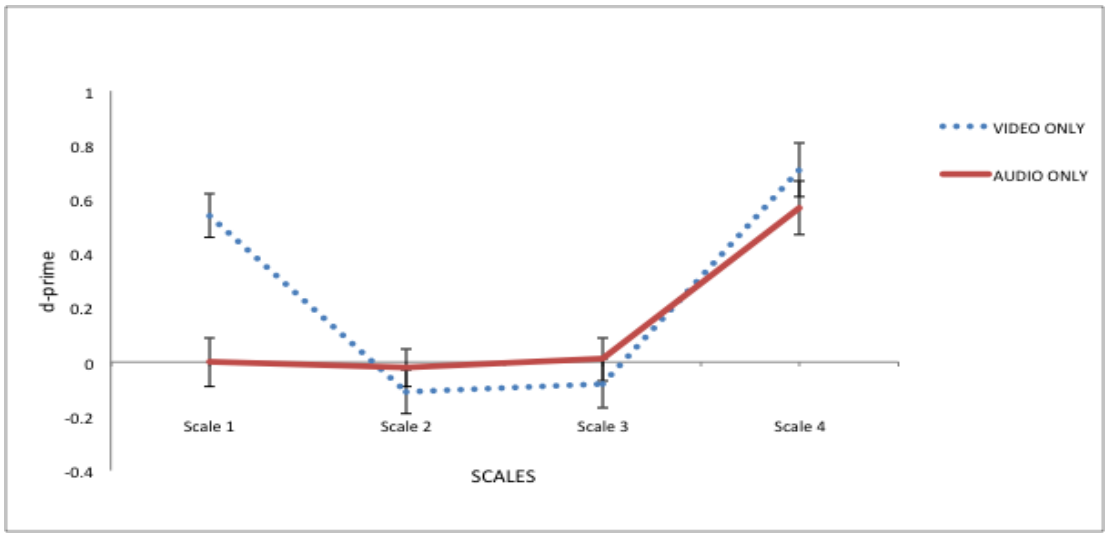


Fig. 4.4. Mean d-prime of each scale in each condition (Video, Audio) in Study 5. Error bars represent standard error of the mean.

To examine whether the perceivers performed differently across scales and conditions in judging the EQ, a 2 × 4 mixed design ANOVA was computed, with the two conditions (Video – Audio) as the between-subjects factor and the four EQ scales as the within-subjects factor; the dependent variable was d'. There was a main effect associated with the scales, F (3, 174) = 29.61, p < .001, a marginally significant main effect of condition, F (1, 58) = 4.16, p = .05, and an interaction between the scales and the conditions, F (3, 174) = 6.10, p = .001. Simple effects analyses revealed two things. First, as with the previous studies, there were significant differences among the four scales and this trend was evident for each condition: Video Condition, F (3, 87) = 22.73, p < .001; Audio Condition, F (3, 87) = 12.17, p <. 001. Post hoc LSD analyses confirmed higher d' values in Scale 1 and Scale 4 compared with Scales 2 and 3 in the video condition (ps < .001) but higher d' values only in Scale 4 in the audio condition (ps < .001). Second, as seen in Table 4.3 and Fig. 4.4, the d' value in Scale 1 was much higher in the video condition than in the audio condition, and an independent-samples t test provided confirmation, t

(58) = 4.46, $p < .001$; there was no difference between video and audio conditions in Scale 4.

To summarize, Study 5 provided new evidence for perceivers' ability to make judgments of empathic traits: Perceivers also systematically identified targets with high EQ (but not with low or middle EQ) on merely listening to the target's voice for about 9 seconds as he or she read aloud a joke. Evidently, perceivers stood a better chance of estimating EQ in the video condition, especially when the target's EQ was low.

How could perceivers form accurate impressions of empathic traits even if they only heard a soundtrack spanning less than ten seconds? In the audio condition, the content of the verbal information is not about the targets' personal lives, and is merely several lines of a joke. Hence, there was no possibility for perceivers to obtain information about the life of the targets in making judgments of their EQ; instead, the only available information was the target's voice characteristics such as tone pitch, as well as mannerisms, such as laughing. Even based on this scant auditory information, overall accuracy of the EQ judgments was still significantly above chance though it was slightly lower than the accuracy in the video condition. Moreover, on hearing the soundtrack, perceivers performed as well as in the video condition when judging the targets with high EQ. These results hence suggest that visual information is not the only channel that perceivers might utilize in making psychological inferences, and auditory cues can also play a key role in forming an accurate first impression of empathic traits, especially in the case of high empathy.

Why were perceivers able to guess the low empathic targets in the video condition but not in the audio condition? In comparison with auditory cues, it seems visual cues provided better information, enabling the perceiver to identify the targets

who had low empathic capacity. Table 4.4 summarizes the behaviour of the 12 targets who were deemed to have low EQ and shows how many perceivers correctly identified these targets as having low EQ. It seems that targets who smiled infrequently, who had few bodily movements and who seldom looked at the camera were likely to be accurately perceived as having low EQ, whereas those who performed in more positive ways while reading the joke, such as looking at the camera, responding to the punch lines of the joke (smiling), and showing more bodily gestures, were less likely to be accurately judged as having low empathising capacity. Such telltale behaviour would not have been apparent in the audio condition and perhaps this is the reason why perceivers were unable to reliably estimate low empathising capacity in this condition.

Table 4.4. The number of perceivers' correct judgments of the low EQ targets and the coding of the targets' visual behavioral cues in the video condition (without sound) in Study 5

Target No.	Number correct out of 30 (%)	Visual behavioral cues				
		Look at the camera	Smiling while telling the joke	Smiling after the joke	A expressive face	Bodily movements
16	16 (53.33)	No	No	No	Less	No
25	15 (50.00)	No	No	No	Less	No
15	13 (43.33)	Yes	No	A little	Less	No
33	12 (40.00)	Yes	No	A little	Middle	No
1	11 (36.67)	No	No	A little	Less	No
14	10 (33.33)	No	No	No	Less	Middle
9	9 (30.00)	Yes	No	A little	Less	No
39	8 (26.67)	Yes	Yes	A Little	Middle	Less
42	7 (23.33)	Yes	Yes	Middle	Middle	Middle
31	6 (20)	Yes	Yes	A Little	Middle	Middle
17	5 (16.67)	Yes	Yes	More	Middle	Less
7	2 (0.06)	No	No	More	Less	Middle

4. 5 General Discussion

The three studies replicated and extended our previous findings of empathic trait judgments based on short samples of behaviour. The video conditions in Study 3 and Study 5 replicated the results of Study 1 and Study 2, suggesting that perceivers generally could guess EQ, and were especially good at identifying the targets with either low or high EQ. Using a variety of different kinds of information, including photographs and sound, the research presented in this chapter has extended the evidence of perceivers' ability to infer empathic traits. The results of Studies 3 and 4 seem to suggest that the capacity to guess the EQ of the target is based on the

behaviour of the target. A neutral image that does not give any information about activity does not contain enough information for the perceiver to guess the target's EQ.

Either visual or auditory information of the target's behaviour is sufficient for perceivers to make an accurate judgment of the EQ though visual information might be more effective in helping perceivers to infer which targets have low EQ. How could perceivers link the visual and vocal behaviour with empathic traits? According to a study by Zaki, Bolger, and Ochsner (2009), perceivers were better at inferring the targets' affect when the targets were expressive and allowed their thoughts and feelings to be read. Is it possible that perceivers linked the observable behaviour cues (either visual or auditory) with the targets' expressivity, and accordingly guessed the EQ on this basis? To examine this possibility, we carried out two surveys, each asking 10 independent judges to evaluate how expressive they thought the target was (on a four-point scale, from low expressive to high expressive) after either watching the videos of the Joke Scenario (without sound) or hearing the sound extracted from the same videos. There was no difference of the evaluation of the targets' expressivity between the conditions of video and sound (Video: $M = 2.10$, $SD = .70$; Audio: $M = 2.21$, $SD = .64$). Moreover, there were significant correlations between perceivers' average assessments of EQ ($M = 2.38$, $SD = .34$) and judges' mean ratings of expressivity in the video condition ($r = .70$, $p < .001$), and between perceivers' average evaluation of EQ ($M = 2.41$, $SD = .50$) and judges' mean ratings of expressivity in the audio condition ($r = .71$, $p < .001$). These preliminary data seem to suggest that perceivers indeed judged if the target was expressive by observing the visual or auditory behaviour associated with telling a joke; furthermore, how they evaluated expressivity of the target was associated with how they guessed

the EQ of the target. Those targets who were assessed as having low expressivity were more likely to be perceived as having low EQ, while those who were considered as having high expressivity were likely to be perceived as having high EQ. In other words, it seems that expressivity is an important indicator allowing perceivers to form a first impression of empathic traits based on thin slices of observable behaviour.

In conclusion, the phenomenon of being able to guess the EQ, with special sensitivity to identifying low and high EQ, is highly replicable and therefore highly robust: Perceivers can perform well above chance on the basis of merely 9 seconds of evidence. The evidence can be a video with sound, a video without sound or three static photographs taken from the videos that sample three different moments in the target's behaviour. The evidence can be a photograph taken from the videos that captures the last moment in the target's behaviour, as well as merely hearing the target's voice for approximately 9 seconds (but in this case perceivers are effective only in identifying cases of high EQ).

CHAPTER FIVE

Study 6: The Correlation Between Empathic Trait Judgments and Assessments of Expressivity

5.1 Introduction

According to Baron-Cohen and Wheelwright (2004), empathy is the ability to share other people's thoughts and emotions, allowing us to respond to their emotions appropriately. Further, it is proposed that thought or emotion experienced by a target resonates in the observer. This definition suggests that capacity for empathizing is determined by the way in which we react to what someone else might be thinking and feeling. For this to be possible we would need to sense what the target is thinking or feeling and what would in turn depend on these states leaking out in the behaviour of the target. Hence, we might suppose that a state of empathy is more or less "visible". In social life, if a person shows seldom acts in an empathic way, for example, consoling a friend on her difficult day or expressing sympathy for someone who is suffering, then the person would be probably perceived as less empathising. In contrast, an empathic person would be expected to demonstrate empathy-related responding, such as having a sympathetic facial expression, giving an understanding eye gaze, offering helps for those in need, and so forth.

Different social situations might induce different experiences of empathy. For example, people might be more likely to demonstrate empathising in a scenario of hearing an empathic story than in a scenario of being detained without reason.

Researchers have argued that empathy relevant emotions can be determined from physiological and facial indexes in some situations. Moreover, it seems that other-oriented sympathetic responding is positively related to behaviour of altruism (Eisenberg & Fabes, 1990). Besel and Yuille (2010) have also reported a significant correlation between individual differences in empathy and accuracy of facial expression recognition. Taken together, this evidence indicates that empathic traits can be expressive in some social contexts, which can be revealed in behavioral evidence including facial reactions, bodily gestures, prosocial actions and so on.

Zaki, Bolger, and Ochsner (2009) reported correlations between targets' expressivity and perceivers' empathic accuracy and targets' behaviour. Specifically, targets completed a questionnaire to measure the trait emotional expressivity, and were videotaped while relating emotional autobiographical events. Perceivers were randomly assigned to either view the videos with sound, videos only or hear the sound only; in each information condition, they had to infer the targets' affect on a 9-point Likert scale (on which 1 = "very negative" and 9 = "very positive"). The empathic accuracy was assessed by the correlation between perceivers' inference and targets' self-ratings. The results show that targets' expressivity predicted both targets' behaviour and empathic accuracy. Expressive targets tended to produce more nonverbal negative cues in the process of narrating negative emotional experience and to present more positive verbal cues while discussing positive experience. By observing this, perceivers were more likely to draw an accurate inference about the target's affect. This study suggests two things: (1) Some people are more expressive

than others, and expressivity can be revealed in both visual and verbal behaviours; (2) targets' expressivity provides a fairly reliable clue for perceivers to guess their emotional sates.

Across the five studies reported in Chapters 3 and 4, there is consistent evidence for people's ability to make fairly accurate judgments of empathic traits based on a brief sample of behaviour. Intriguingly, according to the exploratory analysis of the ratings of targets' expressivity (see Study 5, Chapter 4), the behaviour might have delivered the signals indicating the extent to which the target was expressive. After observing either visual or auditory behaviour lasting for several seconds, judges were able to assess whether a target was expressive, and these assessments of expressivity were significantly associated with perceivers' judgments of the targets' EQs. Following this preliminary result, Study 6 aimed to investigate whether perceivers' empathic trait judgments are correlated with their ratings of the targets' expressivity.

In summary, the main purpose of Study 6 was to investigate the basis on which perceivers judge that a target has high or low empathy. We already know from Study 5 that targets who are rated by independent judges to be expressive are the same who are rated by perceivers as having a high level of empathy; conversely, those rated by judges as being less expressive are rated by perceivers as having low empathy. Based on that finding, we take a further step in Study 6 by investigating whether or not targets who are judged by perceivers to be expressive are also judged by perceivers to have a high level of empathising capacity; and conversely we investigate whether or not targets who are judged by perceivers to be un-expressive are also judged by perceivers to have a low level of empathising capacity. If so, then

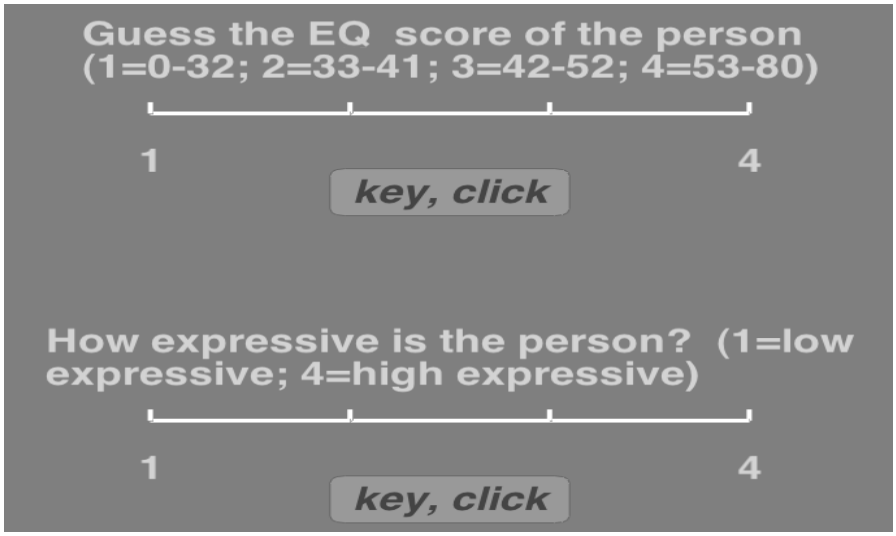
we should conclude that when perceivers judge that the target has high empathising capacity, this is tantamount to the perceiver judging that the target is expressive.

5.2 Method

5.2.1 Summary

For the sake of simplicity this study only used the videos (with audio) in the Joke Scenario (in Study 2) because the results in the previous studies based on this scenario were very clear. After completing the self-report EQ questionnaire, the perceiver was given brief definitions of empathy, EQ and expressivity. Expressivity is defined as the extent to which a person expresses themselves, which can be reflected in facial expressions, bodily movements, voice cues and the like. After that, the perceiver was asked to evaluate how expressive the target was and guess the EQ of the target on two separate four-point scales after viewing the video (see Fig.5.1). The responses were automatically recorded by the software PsychoPy for later retrieval.

Video presentation screen



Response screen

Fig. 5.1. An example of a trial in Study 6.

5.2.2 Participants

Thirty students (15 females & 15 males) aged between 19 years and 28 years (mean age 21 years, *SD* = 1.96) were recruited from the University of Monash Sunway Campus. To assure that the perceivers were unacquainted with one or more targets, they were asked whether they had seen any of the targets and all perceivers denied having met any of them.

5.2.3 Materials and Procedure

The 47 video clips with audio in the Joke Scenario taken from Study 2 (see Chapter 3) were used as stimuli, presented in 800 × 650 pixels in full colour on the 14-inch HP laptop with the software PsychoPy (1.74.00 windows version). The procedure was similar to that in the earlier studies except that participants were required to evaluate expressivity of each target in addition to rating his/her EQ. The order of the two categories of ratings was counterbalanced across participants: Half the participants first guessed the EQs of the targets and then evaluated the expressivity; the remaining participants rated the expressivity and then assessed the EQs of the targets.

5.3 Results

5.3.1 Preliminary Analysis – The EQ Scores of the Perceivers and the Targets

The perceivers' average EQ was 39.53 (*SD* = 12.45, ranging from 16 to 71). Preliminary analyses did not find a difference between the average EQs of targets and perceivers.

5.3.2 Accuracy in Guessing the EQ of the Target

As with the previous studies, the index of accuracy was assessed as d' in each category of the four-point scale: Scale 1: *d*' = .51, *SD* = .46; Scale 2: *d*' = -.11, *SD* = .34; Scale 3: *d*' =. 02, *SD* = .48; Scale 4: *d*' = .60, *SD* = .73. Averaged across all four scales the *d*' was .25, *SD* = .32. The results replicated previous studies with video stimuli, where d' values in Scales 1 & 4 and the average d' were significantly above zero according to one-sample t tests (Scale 1: t (29) = 6.00, $p < .001$; Scale 4: t (29) = 4.51, $p < .001$; the average d': t (29) = 4.39, $p < .001$). A repeated measures one-way ANOVA demonstrated a main effect associated to the four EQ scales, F (3,

87) = 16.28, $p < .001$. Post hoc LSD tests showed that d' values were higher in Scale 1 than in Scale 2 ($p < .001$) and Scale 3 ($p < .001$). Also, d' values were higher in Scale 4 than in Scale 2 ($p < .001$) and Scale 3 ($p < .001$); there was no difference between Scale 1 and Scale 4 and neither was there a difference between Scale 2 and Scale 3. In sum, again, this study replicated the earlier findings: Overall, perceivers could guess the EQs of the targets, and they were better at judging the targets with either high or low EQ than the targets with average EQ.

5.3.3 The Correlation Between Ratings of EQ and Assessments of Expressivity

In order to examine whether ratings of EQ related with ratings of expressivity for each target, we calculated the average rated EQ (as rated by perceivers) and the average assessed expressivity (based on the 4-point scale). The mean of the average rated EQ and the average evaluated expressivity was 2.41 (SD = .45) and 2.29 (SD = .64) respectively. There was a highly significant association between the ratings of EQs and the evaluations of expressivity (r (45) = .85, $p < .001$) according to a Pearson correlation.

The order between perceivers' ratings of the EQ and the expressivity was counterbalanced, and there was neither an order effect associated with perceivers' ratings on the EQs of targets nor an order effect related to perceivers' assessments of how expressive each target was according to paired-samples t tests. Having said that, to be cautious, we calculated the average rated EQ across the 15 participants who first guessed the EQ and the average rated expressivity across the other 15 participants who first evaluated the expressivity (EQ: M = 2.42, SD = .44; Expressivity: M = 2.25, SD = .65). The Pearson correlation between these two categories was r (45) = .75, $p < .001$. Indeed, more than 70 percent of the variance in judgments of EQ was accounted for in assessments of expressivity.

5. 4 Discussion

All in all, these results provided compelling evidence for a strong relationship between empathic trait judgments and assessments of expressivity: Targets judged to be less expressive were also judged to have low EQ; targets judged to be more expressive were also judged to have high EQ. Cross contamination from answering one question to another cannot explain this high correlation because ratings (of expressivity or of EQ) were based on judgments made to the first question (whether the question asked for a rating of EQ or expressivity). Still, is it possible that perceivers thought that the question about EQ and the question about expressivity were for all intents and purpose asking the same thing? If so, then we might expect the means of these two scales to be identical but in fact the mean for judged expressivity was lower than the mean for the judged EQ, t (46) = 2.38, p = .022. Hence, apparently perceivers treated these two questions as being different from each other.

In conclusion, the results seem to suggest that perceivers made judgments of EQ based on how expressive they thought the target was. But correlation is not the same as causation and an alternative possibility is that perceivers based their judgment of expressivity on their judgment of the target's EQ. This seems an implausible explanation when considering the results of Study 5. In that, raters who were independent of the perceivers assessed the expressivity of the targets and the rated expressivity (as determined by these independent raters) correlated strongly with the ratings of EQ made by the perceivers. Therefore, it is reasonable to suppose that ratings of expressivity made by the perceivers are corroborated by ratings of expressivity made by independent judges. Of course, we cannot completely rule out the possibility that the independent judges based their ratings of expressivity on their

estimation of the targets' EQ but it is hard to imagine why they would have done such a thing. The most plausible explanation is that judgments of EQ are based on assessments of expressivity.

CHAPTER SIX

Can People Guess the Big Five Personality Traits Based on Thin Slices of Behaviour?

6.1 Introduction

6.1.1 The Big Five Personality Traits

The Five-Factor Model of personality (often termed the "Big Five") depicts the fundamental structure of human beings' personality traits in terms of five basic dimensions: Neuroticism (N), Extraversion (E), Agreeableness (A), Conscientiousness (C), and Openness to Experience (O) (McCrae & John, 1992). Research using natural language adjectives and theoretically based personality questionnaires has supported the comprehensiveness of the model and its applicability across observers and cultures (McCrae & John, 1992, for a review; Borkenau, 1998; McCrae & Costa, 2004; McCrae, Terracciano, et al., 2005; Schmitt, Allik, McCrae, & Benet-Mart ńez, 2007).

Several inventories have been developed to measure the Big Five personality factors, including the most comprehensive instrument NEO Personality Inventory, Revised (NEO-PI-R; Costa & McCrae, 1992) and its corresponding shortened version NEO Five-Factor Inventory (NEO-FFI; Costa & McCrae, 1992). According to the manual of NEO-PI-R (Costa & McCrae, 2011), the core of the domain N is the

general tendency to experience negative affects, such as fear, sadness, embarrassment, anger, guilt and disgust; individuals high in N are also prone to have irrational ideas, to be less able to control their impulses, and to cope poorly with stress compared with others; the dimension E includes traits in the tendency to be sociable, assertive, active, talkative, excitement-seeking and to experience positive emotions; the elements of O involve imagination, aesthetic sensitivity, attentiveness to inner feelings, preference for variety, intellectual curiosity, and independence of judgment; the domain A means a tendency to be straightforward, altruistic, compliant, modest and tender-minded; the domain C is pertinent to the tendency to be purposeful, strong-willed and determined, to act dutifully, strive for achievement, and to be orderly rather than disorganized.

6.1.2 How Well Can We Form First Impressions of the Big Five Traits?

In daily life, we have many opportunities to meet people we are unacquainted with. How well can we form a first impression of the Big Five personality traits of them? Norman and Goldberg (1966) reported the first major study of zero acquaintance to seek for solutions to this question. A sample of college students with no prior acquaintance nor any opportunity to interact with one another were randomly assigned to a 6- to 9-person group and then they were asked to independently evaluate each member of their group and themselves on a set of 20 bipolar scales measuring the five factors of traits (Extraversion, Agreeableness, Conscientiousness, Emotional Stability (equal to ‘N’), and Culture (equal to ‘O’)). This study yielded striking results, indicating significant correlations between the self-ratings and averaged peer ratings for three of the five personality factors, including Extraversion, Conscientiousness and Culture.

Following the same face-to-face procedure of zero acquaintance, a body of subsequent research has replicated these initial findings to some extent (e.g., Albright et al., 1988; Watson, 1989; Kenny, Horner, Kashy, & Chu, 1992; Kenny et al., 1994): Researchers consistently revealed significant self-other agreement and high consensus among perceivers on Extraversion and Conscientiousness but the results on the dimension of Culture were inconsistent (e.g. Albright et al., 1988; Watson, 1989).

Using videotapes recording the targets while performing one or more certain activities, Borkenau and colleagues have conducted a series of studies examining accuracy (defined as self-other agreement or consensus among judges) in first-impression formation of personality (Borkenau & Liebler, 1992a; 1992b; 1993; Borkenau et al., 2004). In one study (Borkenau & Liebler, 1992b), a sample of 100 targets was videotaped while entering and walking through a room, sitting down, looking into the camera, and reading a standard text, followed by providing self-ratings of the Big Five personality traits on a set of 20 bipolar scales. Each bipolar scale was labeled with two opposite trait-descriptive terms (e.g., talkative vs. silent) along with a 7-point rating scale. A sample of 24 perceivers with no prior knowledge of the targets was asked to evaluate the targets' personalities on the same set of 20 bipolar scales when given one of four types of information on the targets: (a) sound film, (b) silent film, (c) still (a 60s still of each target looking into the camera) or (d) audiotape. The results demonstrated that performance in judging the five traits of the targets was affected by different types of information – In the sound film condition, there was significant self-other agreement on each of the five traits except for N, in the silent film condition, the self-ratings of the targets significantly agreed with the judgments by perceivers on three of the five personality traits (E, C and N), in the

still condition, self-other agreement was only found on E and C, and in the audiotape condition, there were significant correlations between targets' self-ratings and perceivers' ratings on E and A.

Focusing on thin slices of behaviour as cues of personality, Borkenau et al. (2004) reported a comprehensive study investigating inter-judge agreement on personality judgments, suggesting that trait inferences can be drawn from thin slices of behaviour in seemingly trivial scenarios. In particular, after completing self-ratings of the Big Five personality traits (on the NEO-Five Factor Inventory) and intelligence, the targets were individually videotaped while performing 15 different tasks lasting from 1 minute to 12 minutes, in which they were asked to introduce themselves, tell a joke to an experimenter confederate, introduce a stranger to the experimenter, invent a neologism, read newspaper headlines, sing a song and so on. These videotapes were later presented to a sample of independent judges who had never seen the targets before. Every four independent judges observed each target in one of the 15-videotaped episodes only, and then rated the target along the Big Five traits on a set of 20 bipolar adjective scales. Results showed that the average agreement among judges across the tasks was highest on E and lowest on C and O, and personality inferences from thin slices of behaviour were significantly associated with the ratings by close acquaintances of the targets and this association became stronger when more episodes were included (Borkenau et al., 2004).

Would perceivers perform better in drawing personality inferences if they had a chance to observe a trait-relevant context instead of mundane scenarios? Based on carefully structured trait-related situations allowing one or more traits to be expressed, McLarney-Vesotski, Bernieri and Rempala (2006) investigated the ability to infer the Big Five traits from thin slices of behaviour. Specially, targets were

videotaped in five carefully crafted situations, each situation designed to maximally induce a dimension of the Big Five traits. Perceivers evaluated the personalities of the target using a rating form involving five questions (each along with a 5-point rating scale) corresponding to the five personality traits (e.g., (For the trait A) How pleasant and positive is the person?). The accuracy was defined as perceivers' agreement with the personality assessments of the targets provided by knowledgeable informants. Results suggested well-above chance judgmental accuracy in each trait except for A, with a strongest correlation of perceiver-informant agreement on E (r = .57), a relatively weaker correlation on C (r = .25), and a weakest correlation on N (r = .12). Apparently, perceivers did not demonstrate more accurate performance in making trait inferences compared with the findings of prior studies based on mundane situations, such as an unstructured get-acquainted conversation, reading the standard text of a weather forecast, or an unstructured dyadic interaction.

To examine a thin slice perspective on the accuracy of first impressions, Carney et al. (2007) conducted a study investigating judged construct (the Big Five personality traits and other psychological characteristics), exposure time (5, 20, 45, 60 and 300s) and slice location (beginning, middle, end) using video stimuli (with sound) filming a 5-minute unstructured dyadic interaction. Accuracy was indexed as self-other agreement on the personality evaluation of the targets. Results indicated that judgments of perceivers were moderately correlated with the ratings of the targets on E and C after 5s exposures, but N, O and A required more exposure time to achieve similar levels of accuracy. Overall, accuracy increased with exposure time, personality judgments based on later segments of the 5-minute interaction were more accurate, and 60s yielded the optimal correlation between accuracy and slice length

(Carnet et al., 2007). Researchers thus conclude that accuracy of first impressions relies on the personality factors (N, E, O, A, & C), amount of exposure, and temporal location of the slice of observable social behaviour.

Taken together, the literature on accuracy research of personality judgments has suggested that we can form somewhat accurate first impressions on some dimensions of the Big Five traits. Overall, perceivers can draw accurate inferences about E and C across different contexts but their performance in inferring the other three personality dimensions seems to be inconsistent across different studies (e.g., Norman & Goldberg, 1966; Watson, 1989; Borkenau & Liebler 1992b; McLarney-Vesotski et al., 2006; Carney et al., 2007). Furthermore, even with minimal behavioral observation, according to the research adopting the paradigm of 'thin slices', perceivers are still able to form an accurate first impression on E and C or other traits based on observations of people in either trivial scenarios or crafted trait-relevant situations (e.g., Borkenau et al., 2004; McLarney-Vesotski et al., 2006; Carney et al., 2007).

6.1.3 The Present Study

There are some limitations in the extant research into accuracy in first impressions of personalities. First of all, comparison of judgmental accuracy should be made with caution due to methodological variations between studies (Hall, Andrzejewski, Murphy, Mast, & Feinstein, 2008), such as different indexes of accuracy (self-other agreement or peer consensus), diverse measurements of the Big-Five traits (e.g., a set of 20 bipolar scales or the NEO-Five Factor Inventory) and various procedures of zero acquaintance (e.g. face-to-face or video-based contexts). Second, in the earlier studies, participants usually had to act as a perceiver and a target at the same time. In this case, the accuracy indexed as correlations of self-other

agreement might be inflated by participants' consistent response patterns in the personality questionnaires and their shared stereotypes of generalized people. Third, judgmental accuracy (self-other agreement or consensus among perceivers) was usually calculated based on the ratings from a small group of judges (including 2 to 7 persons). However, some researchers have suggested that self-other agreement correlations tended to rise as the number of peer raters increased; for example, ratings of A displayed significant convergent validity when a sufficient number of judges rated the target (Watson, 1989). Fourth, there is still controversy about people's abilities to infer some of the Big Five traits. For instance, even in a carefully crafted trait-relevant situation, perceivers failed to make an accurate judgment of A (McLarney-Vesotski et al., 2006), whereas in another study (Borkenau et al., 2004), after briefly observing trivial scenarios the target experienced, perceivers could achieve significant agreement on the evaluation of A. Besides, some researchers reported a positive connection between accuracy of personality judgments and exposure time of thin slices of behaviour (e.g., Carnet et al., 2007), but a meta-analysis revealed that studies using longer periods of behaviour observation did not differ significantly from predictions based on 4- and 5-minute observations (Ambady & Rosenthal, 1992). Finally, accuracy, operationally defined as correlations of self-other agreement or correlations among perceivers allows examination of an overall level of accuracy to some extent but does not always allow us to investigate perceivers' fine-grained abilities to identify different levels (e.g., low, average, & high) of the same personality dimension.

Considering these factors, it is necessary to confirm how well we can guess personalities from thin slices of behaviour in naturalistic social settings using a novel method that overcomes some of the limitations identified in the previous research.

Due to its high ecological validity, the method we developed for investigating accuracy of empathic trait judgments is also well suited for studying accuracy of first impressions on the Big Five traits. From Chapters 3 to 5, we already know that perceivers can systemically guess the targets' empathic traits by observing a brief sample of behaviour, and performance has a U-shaped pattern in which perceivers are more effective in evaluating the targets who had a low or high EQ than the targets who had average EQ. Adopting the same accuracy-oriented methodology, we aim to find out whether the same U-shaped pattern would be apparent when assessing the Big Five personality traits. Thus, the two studies reported in this chapter tried to explore people's fine-grained abilities to make personality inferences about others on first meeting.

6.2 Study 7

Apart from the study by Borkenau et al. (2004), previous research on accuracy of personality judgments usually examines the capacity to infer the Big Five traits on the basis of behaviour observed from only one situation. Nonetheless, Borkenau et al (2004) have found that correlations of personality judgments between perceivers and knowledgeable informants became stronger when observable actions involved more episodes. This implies that it may be more ecologically valid to investigate people's capacity for trait inferences based on a broad range of behaviour cues extracted from more than one situation. In Study 2 (see Chapter 3), after watching a sample of video, perceivers were capable of making accurate judgments of empathic traits across three scenarios. Using the same video stimuli (involving three scenarios), the main goal of Study 7 was to examine how well people could guess the Big Five traits of the targets after watching a short video clip.

6.2.1 Method

Summary

Three sets of video clips (with sound) were derived from Study 2 (see Chapter 3), with 47 clips in each set where the targets were recorded during the conversation, reading a joke or doing the screen test. A 5 ×3 × 3 mixed design was used, with the Big Five traits and the three scales of each trait (low, average, & high) as the within-subjects factors and the three scenarios (Conversation, Joke, and Screen Test) as the between-subjects factor. The video clips were displayed as with the studies on empathic trait judgments. After viewing each video clip, perceivers had to guess the Big Five traits of the target. The procedure was scrutinized and approved by the Faculty of Science ethics committee in the University of Nottingham Malaysia Campus, which was constituted and operated according to guidelines prescribed by the British Psychological Society.

Participants

The perceivers were ninety students (45 males) aged 18 to 25 years (mean age 20 years, *SD* = 1.59), recruited from the University of Nottingham Malaysia campus. To make sure that the perceivers did not know the targets, they were asked to identify whether they had met any of the targets after viewing the pictures of them (taken from their videos). Another fifteen participants who reported one or more acquaintances were replaced. Hence, the sample of 90 perceivers who proceeded to the testing phase were assumed to be unacquainted with the targets. Each perceiver was randomly assigned to watch the targets either in the Conversation, the Joke or the Screen Test Scenario.

Materials

Video stimuli

One-hundred-and-forty-one video clips were obtained from Study 2, with 47 clips in each of the three scenarios (the Conversation, the Joke, and the Screen Test Scenario). Each target experienced the three scenarios. The details of the stimulus video collection and editing appeared in Chapter 3.

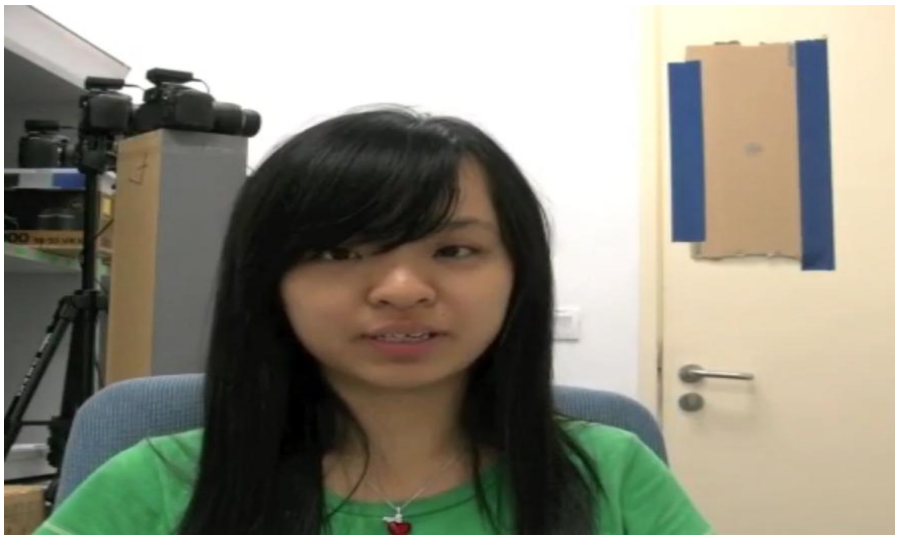
The NEO Five-Factor Inventory-3 (NEO-FFI-3) and the Targets' Big Five Personality Traits

The NEO-FFI-3 is an updated and revised shortened version of the self-report form of the NEO Personality Inventory-3 (NEO-PI-R-3), with 12-item scales to measure each of the five personality factors, N, E, O, A, and C, providing a comprehensive portrait of an individual in respect of global personality with high reliability and validity (McCrae & Costa, 2004). Responses were collected on 5-point Likert scales ranging from strongly disagree to strongly agree.

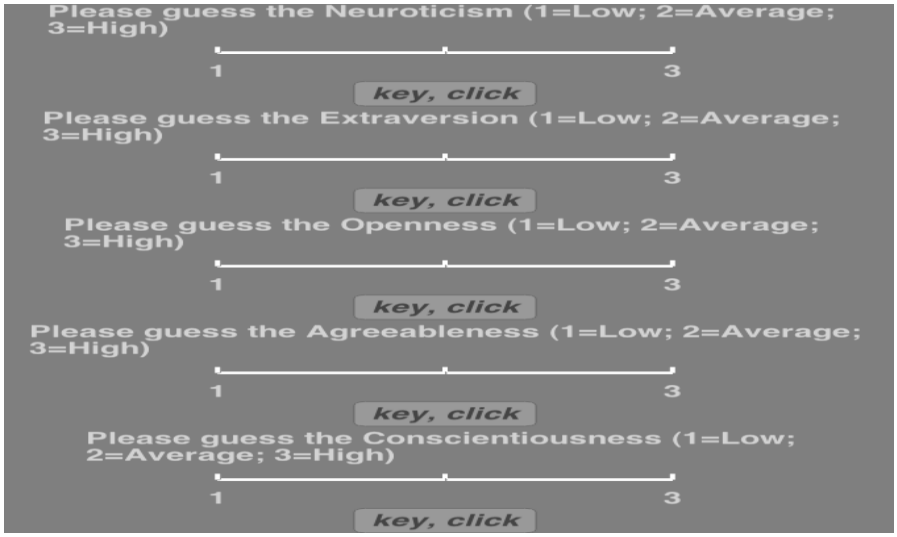
After recording and a short break for several minutes, the target was asked to fill in the NEO-FFI-3 questionnaire (plus the EQ questionnaire but this had no relation with the current study). On average it took about 15 minutes to complete the NEO-FFI-3. According to the scoring instruction in the manual (Costa & McCrae, 2011), the targets' self-report ratings on each of the five personality traits were classified into a scale of low, average or high: Eight, 13 and 26 targets respectively scored low, average and high on N; 14, 23 and 10 were respectively low, average and high in E; 7, 22, and 18 were low, average and high O; 20, 20 and 7 scored low, average and high on A; 24, 13 and 10 were low, average and high on C.

Procedure

Perceivers were tested individually. After receiving the instruction information sheet and offering consent, the perceiver was given a personality information sheet that includes a brief definition of personality along with concise explanations of each personality trait in line with the description in the NEO manual (Costa & McCrae, 2011). After reading the personality information sheet without any further questions, the perceiver proceeded to the trials of guessing the targets' Big Five traits. Each video clip was displayed (in size of 600 × 400 pixels) to the perceiver in full colour with audio in random orders with the software PsychoPy (Peirce, 2007) on a laptop. In each trial, following a video presentation screen, a response screen appeared showing five 3-point scales (from low to high) as response options corresponding to the five traits of N, E, O, A, and C in a fixed order (see Fig.6.1). The perceiver guessed the five traits of each target by using the mouse to click one of the three scales of each trait. After the perceiver made judgments on all the five traits, the screen moved immediately to the next video clip. The procedure allowed the perceivers to take as long as they needed to respond. Responses were automatically recorded by the software for later retrieval.



Video presentation screen



Response screen

Fig. 6.1. An example of a trial in Study 7.

6.2.2 Results and Discussion

Trends in Judgmental Accuracy of the Big Five Traits in Each Scenario

Following the procedure developed in the previous studies examining judgmental accuracy of empathic traits, the accuracy of the guesses perceivers made about each scale in each of the five traits was calculated using signal detection. This method allows an assessment of accuracy that is independent of underlying base

rates of targets' actual personality scales in each trait and perceivers' response bias in a particular scale of each trait. The index of accuracy was computed as d-prime (d'). Table 6.1 displays the means of d' in each scale of each trait in each scenario (Conversation, Joke, & Screen Test) and the corresponding results of one-sample t tests of each d', where the comparison value is zero. If a perceiver guessed at random when evaluating the traits of the target, this would yield a d' of zero.

Table 6.1. Means (and standard deviations) of d-prime (d') in each scale of each of the five traits (N, E, O, A, C) in each scenario (Conversation, Joke, & Screen Test), along with values of one-sample t tests of each d' in Study 7

Trait	Scale	Scenario					
		Conversation		Joke		Screen Test	
		d' (*SD*)	*t*	d' (*SD*)	*t*	d' (*SD*)	*t*
N	Low	.22 (.46)	2.58*	.31 (.49)	3.45**	.35 (.35)	5.38***
	Average	-.08 (.35)	-1.24	.01 (.37)	.15	.05 (.44)	.59
	High	-.05 (.35)	-.80	.21 (.42)	2.81**	.09 (.39)	1.30
E	Low	.17 (.41)	2.24*	.21 (.43)	2.64*	-.23 (.81)	-1.53
	Average	.0 (.39)	.07	-.09 (.38)	-1.34	-.03 (.39)	.37
	High	.53 (.50)	9.45***	.51 (.37)	7.47***	1.11 (.69)	8.82***
O	Low	.30 (.39)	4.28***	.54 (.11)	2.81**	.42 (.39)	5.88***
	Average	.09 (.33)	1.53	-.05 (.74)	-.04	-.04 (.30)	-.67
	High	.35 (.38)	4.98***	.29 (.82)	1.97	.30 (.38)	4.38***
A	Low	.14 (.38)	2.01	.06 (.54)	.64	.15 (.45)	1.87
	Average	-.08 (.36)	-1.16	-.02 (.47)	-.21	.04 (.49)	.46
	High	.33 (.43)	4.26***	-.10 (.66)	-.83	-.06 (.51)	-.62
C	Low	.26 (.49)	2.97**	.07 (.43)	.95	.15 (.42)	1.91
	Average	-.06 (.31)	-.97	-.06 (.45)	-.71	.12 (.38)	1.74
	High	.24 (.43)	3.11**	.04 (.65)	.34	.18 (.44)	2.29*

Note: Three groups (n = 30 in each group) of perceivers each viewed targets in one of the three scenarios; *. p < .05, **.p < .01; ***.p < .001, two-tailed.

As shown in Table 6.1 and Fig. 6.2 – Fig. 6.4, perceivers seemed to demonstrate a similar trend of being able to guess the five personality traits in each scenario: Perceivers could not make systemically accurate judgments on the average scale of each trait, but they were fairly effective in guessing one or both of the two extreme scales (low and high) in one or more traits; that is, accuracy of judgments about the five traits in each scenario tended to present a U-shaped pattern.

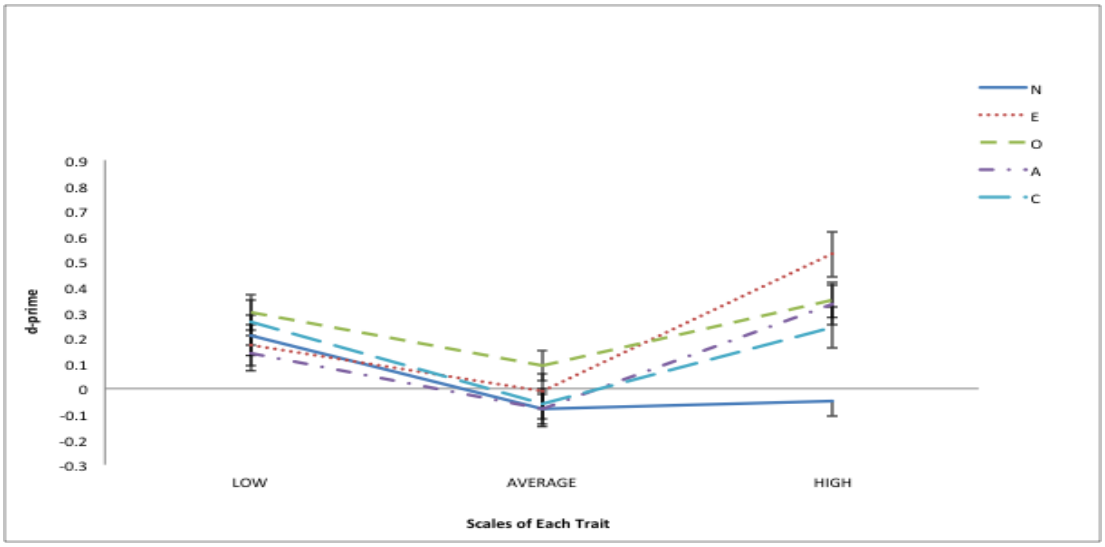


Fig. 6.2. Mean d-prime of each scale in each trait in the Conversation Scenario in Study 7. Error bars represent standard error of the mean.

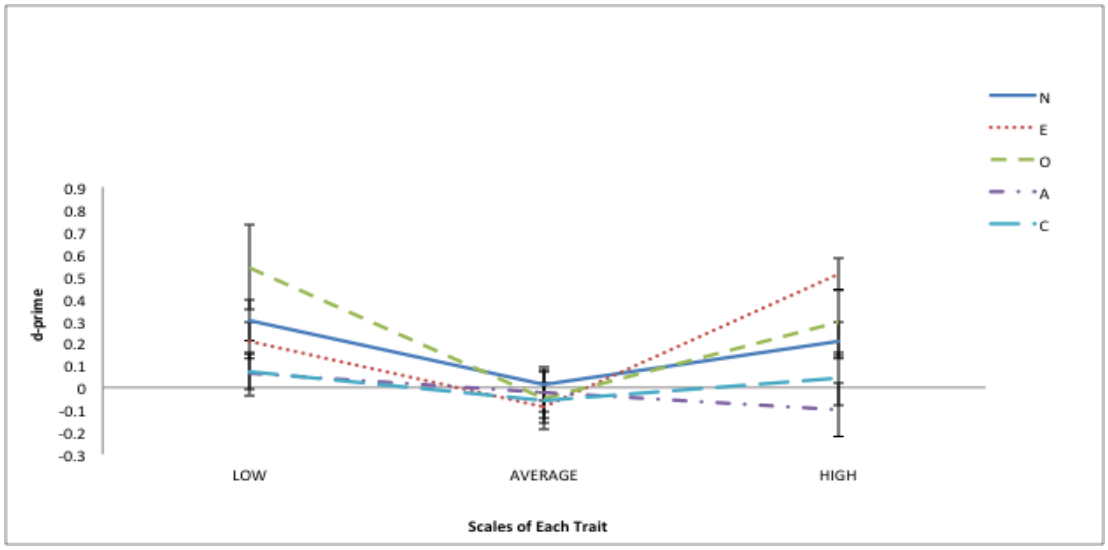


Fig. 6.3. Mean d-prime of each scale in each trait in the Joke Scenario in Study 7. Error bars represent standard error of the mean.

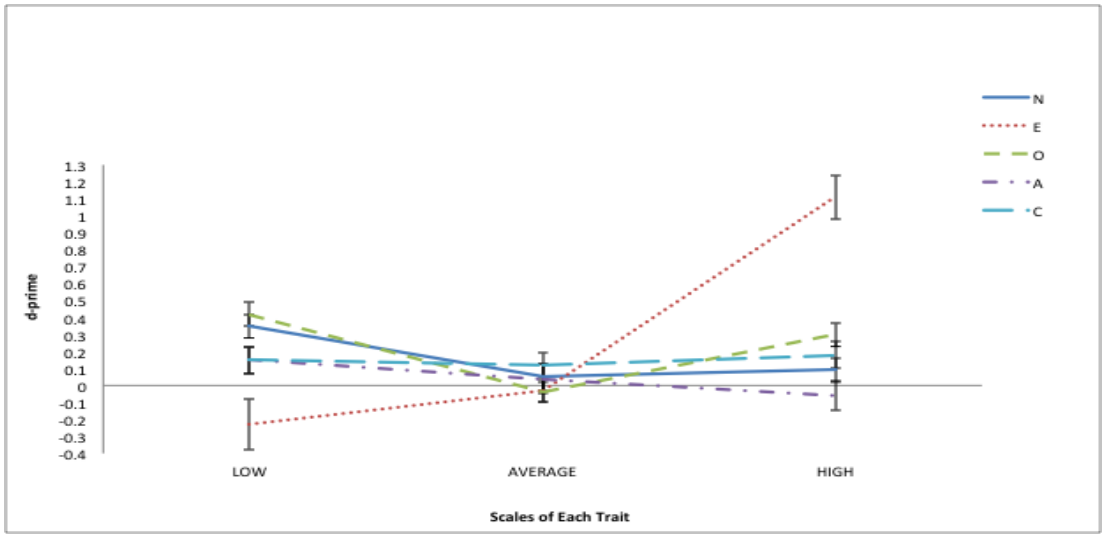


Fig. 6.4. Mean d-prime of each scale in each trait in the Screen Test Scenario in Study 7. Error bars represent standard error of the mean.

According to Table 6.1 and Fig. 6.2 – Fig. 6.4, in the Conversation Scenario, perceivers were good at identifying the extreme scales of each trait with the exception of high N and low A; in the Joke Scenario they were effective in guessing low and high N, low and high E and low O. In the Screen Test Scenario they made systemically accurate judgments of high N, high E, low and high O, and high C. In

other words, the U-shaped curve of each trait might be affected by different scenarios – For example, the curve of E in the Screen Test is rather different from that in the other scenarios; meanwhile, the strength of the U-shaped curve might be dependent on traits: For instance, in general, the U shape of the curves of O across the three scenarios seemed pretty strong while the curves of A appeared rather weak.

In order to examine these preliminary impressions, we carried out a $5 \times 3 \times 3$ mixed design ANOVA, with the five traits and the three scales of each trait as the within-subjects factors and the three scenarios as the between-subjects factor; the dependent variable was d'. There were main effects associated with the five traits (F (4, 348) = 7.88, $p < .001$) and the three scales (F (2, 174) = 42.33, $p < .001$), but no main effect of scenario (F (8, 348) = 1.55, $p = .14$). Meanwhile, an interaction between the traits and the scales (F (8, 696) = 14.32, $p < .001$), and an interaction among the traits, scales and scenarios (F (16, 696) = 3.69, $p < .001$) were also significant but there was no interaction between the traits and the scenarios (F (8, 348) = 1.55, $p = .14$) or between the scales and the scenarios (F (4, 174) = 1.90, $p = .11$).

According to Fig. 6.2 – Fig. 6.4, the three-way interaction among the traits, scales and scenarios might result to some extent from a stronger significant interaction between the traits and the scales in the scenario of Screen Test. We therefore conducted a 5×3 repeated measures ANOVA in each scenario to examine the trends of perceivers' performance in guessing the Big Five traits, with the five traits and the three scales of each trait as the within-subjects factors. As displayed in Table 6.2, which reports the results of the ANOVA tests in the three scenarios, there was the same kind of trend towards perceivers' performance in each scenario: There were main effects related to the traits and the scales as well as an interaction between

the traits and the scales. In addition, the interaction in the Screen Test Scenario was much stronger than the other scenarios, suggesting that the three-way interaction was associated with the interactions between the traits and the scales in different scenarios.

Table 6.2. F values of the repeated measures ANOVA in each of three scenarios (Conversation, Joke, & Screen Test) in Study 7

ANOVA Test	Conversation	Joke	Screen Test
Main Effect of Traits	F (4, 116) = 2.88*	F (4, 116) = 4.45**	F (4, 116) = 3.43**
Main Effect of Scales	F (2, 58) = 38.4***	F (2, 58) = 10.54***	F (2, 58) = 12.47***
Interaction (Traits × Scales)	F (8, 232) = 3.69***	F (8, 232) = 2.15*	F (2, 232) = 17.90***

Note: * $p < .05$, ** $p \leq .01$, & *** $p < .001$

The U-shaped Pattern of Judgmental Accuracy in Each Trait Across the Three Scenarios

Given no main effect of scenario and given that there was an interaction between Trait and Scale for each scenario, we calculated the average d' of each scale in each trait across the three scenarios combined and reported the corresponding *t* values of one-sample t tests of each d', as shown in Table 6.3. Additionally, the average d' of each trait across the scales and the scenarios was computed as overall judgmental accuracy in each trait, and the corresponding results were .12 (*SD* = .30), .24 (*SD* = .29), .25 (*SD* = .22), .05 (*SD* = .35) and .11 (*SD* = .31) for N, E, O, A, and C respectively, which were significantly above chance except for the d' of A according to one-sample t tests (N: t (89) = 3.89, $p < .001$; E: t (89) = 7.89, $p < .001$; O: t (89) = 10.40, $p < .001$; A: t (89) = 1.43, $p = .16$, and C: t (89) = 3.22, $p = .002$),

indicating that generally perceivers could make an accurate judgment of each of the five personality traits except for the trait A.

Table 6.3. Means (and standard deviations) of d-prime (d') in each scale of each trait across the three scenarios combined, along with the corresponding *t* values of one-sample t tests for each d' in Study 7

Trait	Scale					
	Low		Average		High	
	d' *(SD)*	*t*	d' *(SD)*	*t*	d' *(SD)*	*t*
N	.29 (.44)	6.33***	0 (.39)	-.17	.09 (.40)	2.03*
E	.05 (.61)	.78	-.04 (.38)	-1.02	.72 (.60)	11.33***
O	.42 (.68)	5.83***	0 (.50)	0	.31 (.56)	5.34***
A	.12 (.46)	2.45*	-.02 (.44)	-.38	-.06 (.57)	.98
C	.16 (.45)	3.42**	0 (.39)	.05	.16 (.52)	2.85**

Note: ***p < .001, **p < .01, *p ≤ .05; N = 90; two-tailed.

However, as we see in Table 6.3, perceivers were not uniformly effective in guessing each scale of each trait – They made systematically accurate judgments (indicated by a d' statistically higher than zero) in the cases of low and high N, low and high O, low and high C, high E and low A, but in all there was no evidence suggesting systematic judgments when the targets belonged to the middle scale of each trait. Fig. 6.5 summarizes these results, demonstrating a U-shaped curve of different strengths in each trait.

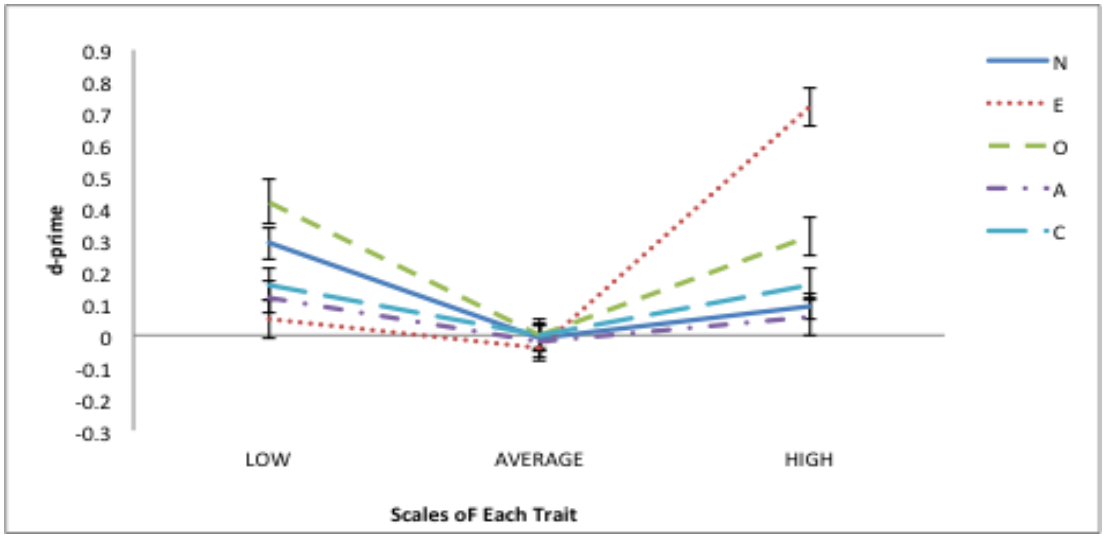


Fig. 6.5. Mean d-prime of each scale in each trait across the three scenarios combined (Conversation, Joke, & Screen Test) in Study 7. Error bars represent standard error of the mean.

To examine the U-shaped pattern in the performance of each trait, we conducted a one-way repeated measures ANOVA for each trait, with the three scales as the within-subjects factor. The results showed that d' values among the three scales of each trait were significantly different except for A (N: F (2, 178) = 18.27, p < .001; E: F (2, 178) = 49.30, p < .001; O: F (2, 178) = 9.78, p < .001; A: F (2, 178) = 2.30, p = .103; C: F (2, 178) = 4.52, p = .012). Post hoc Bonferroni tests for each trait (except for A) revealed the following results. For the trait N, perceivers were better at detecting the targets with low N than the targets with average N (p < .001) or high N (p < .001), but there was no difference between the performance in judging the cases of average N and high N. For the trait E, perceivers performed better in identifying the targets who were high in E than the targets who fell into either low E (p < .001) or average E (p < .001), but they performed similarly in the scales of low and average E. For the trait O, perceivers were equally effective in judging the targets who had either low O or high O, and the performance in both scales were better than the performance in the average O ($ps \leq .001$). For the trait C, perceivers

made more accurate judgments of the targets either with low C (p = .009) or high C (p = .028) in comparison with the targets who had average C, and the judgmental accuracy was not different in the cases of low and high C.

In summary, accurate judgments that perceivers made about each trait (except for the trait of A) presented a similar type of U-shaped pattern, in which accuracy in one or both of the extreme scales was greater than that in the average scale. Despite the finding that the form of the U-shaped curve of A was relatively flat, according to Table 6.3, perceivers were still effective in making judgments about the targets with low A.

6.3 Study 8

Study 7 has already revealed that perceivers could make an overall accurate judgment of each trait with the exception of the trait A across the three scenarios while having access to visual and sound information; moreover, there was the same kind of U-shaped pattern for each trait across the scenarios, in which perceivers consistently failed to detect the targets who were average but were fairly good at identifying the targets who were extreme. Would these principal findings be replicated with different types of information channels (sound video, silent video, and soundtrack)? Study 8 was conducted in order to gain an insight into this matter.

6.3.1 Method

Summary

For the sake of simplicity the present study only used the scenario of Conversation where the targets were unobtrusively filmed while answering some questions about themselves – This scenario might reflect informative spontaneous behaviors of the targets, which could explain why some of the strongest U-shaped

curves were associated with this scenario. In each trial, perceivers either viewed a video with sound, a video without sound, or heard a soundtrack taken from the same video clip. Subsequently, the perceiver was asked to make first-impression judgments of the Big Five traits of the target.

Participants

Ninety students (43 males & 47 females) aged 18 to 27 years (mean age 21 years, *SD* = 1.74) from the University of Monash Sunway Campus volunteered to participate in the study. There were no participants reporting acquaintanceships with one or more targets. Participants were randomly assigned to three groups of 30 in each condition (Video & Audio, Video Only, and Audio Only).

Materials and Procedure

The 47 video clips in the Conversation Scenario were used with sound or without sound. The 47 audio tracks were extracted from the corresponding video clips. Thus, 30 perceivers viewed 47 video clips with sound, 30 perceivers viewed the same 47 video clips without sound, and 30 perceivers listened to 47 audio tracks. The procedure was the same as that in Study 7. After watching an audio video or a silent video or hearing a soundtrack, perceivers were asked to guess the Big Five traits of each target.

6.3.2 Results and Discussion

Trends in Judgmental Accuracy of the Big Five Traits in Each Condition

As with Study 7, the accuracy of the perceivers' ratings of the targets' Big Five traits was analyzed using signal detection, indexed as d'. Table 6.4 shows the means of d' in each scale of each trait in each information condition, together with the corresponding values of one-sample t tests of each d', where the comparison

value is zero. If a perceiver guessed randomly when assessing the personality traits of the target, the corresponding d' would not be significantly above zero.

Table 6.4. Means (and standard deviations) of d-prime (d') in each scale of each of the five traits (N, E, O, A, C) in each condition (Video & Audio, Video Only, & Audio Only), along with values of one-sample t tests of each d' in Study 8

Trait	Scale	Condition					
		Video & Audio		Video Only		Audio Only	
		d' (*SD*)	*t*	d' (*SD*)	*t*	d' (*SD*)	*t*
N	Low	.19 (.43)	2.40*	.05 (.39)	.74	.13 (.58)	1.26
	Average	.08 (.35)	1.32	-.04 (.37)	-.54	.01 (.46)	.12
	High	-.10 (.48)	-1.14	-.07 (.41)	-.98	.15 (.37)	2.16*
E	Low	.24 (.44)	2.94**	-.07 (.35)	-1.16	.51 (.42)	6.66***
	Average	.07(.42)	.87	-.16 (.33)	-2.69*	.06 (.40)	.82
	High	.77 (.44)	9.55***	.29 (.48)	3.35**	.37 (.50)	4.01***
O	Low	.32 (.59)	2.97**	.30 (.53)	3.16**	.29 (.42)	3.82***
	Average	.13 (.43)	1.68	0 (.38)	.04	.13 (.38)	1.89
	High	.32 (.40)	4.29***	.07 (.42)	.98	.08 (.41)	1.06
A	Low	.05 (.39)	.67	.03 (.42)	.34	-.05 (.48)	-.55
	Average	-.10 (.44)	-1.26	-.12 (.34)	-1.93	-.14 (.78)	-1.02
	High	.46 (.58)	4.29***	-.09 (.42)	-1.15	.36 (.54)	3.72***
C	Low	.30 (.50)	3.30**	-.04 (.28)	-.71	-.07 (.45)	-.81
	Average	-.06 (.47)	-.71	-.27 (.44)	-3.32**	-.04 (.37)	-.65
	High	.51(.51)	5.47***	-.06 (.43)	-.72	-.07 (.42)	-.96

Note: Three groups (n = 30 in each) of perceivers viewed or heard targets in one of the three conditions; *. $p < .05$, **.$p < .01$; ***.$p \leq .001$, two-tailed.

As we can see in Table 6.4 and Fig. 6.6 – Fig. 6.8, similar to Study 7, perceivers seemed to demonstrate a U-shaped trend in being able to make an accurate judgment of the Big Five traits in each condition. That is, when judging the targets with an average trait, perceivers generally demonstrated accuracy of chance

and accuracy even worse than chance in the judgments of E and C in the Video Only Condition, but they could systemically identify the targets who had a low or high scale in one or more traits in each condition. Furthermore, this U-shaped pattern seemed to be influenced by information conditions – The trends seemed strongest in the condition of Video and Audio where perceivers could guess the extreme scales in more types of traits than the other information conditions; also, the strength of the U shape varied depending on traits – The U shape of the curves of E and O across the three conditions appeared relatively strong as opposed to the curves of N and A, for example.

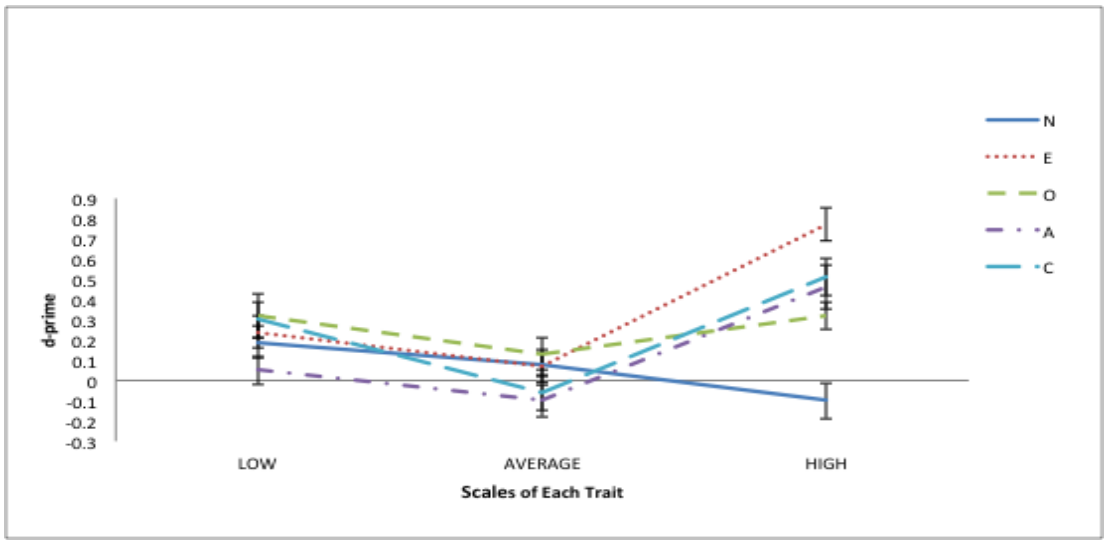


Fig. 6.6. Mean d-prime of each scale in each trait in the Video and Audio Condition in Study 8. Error bars represent standard error of the mean.

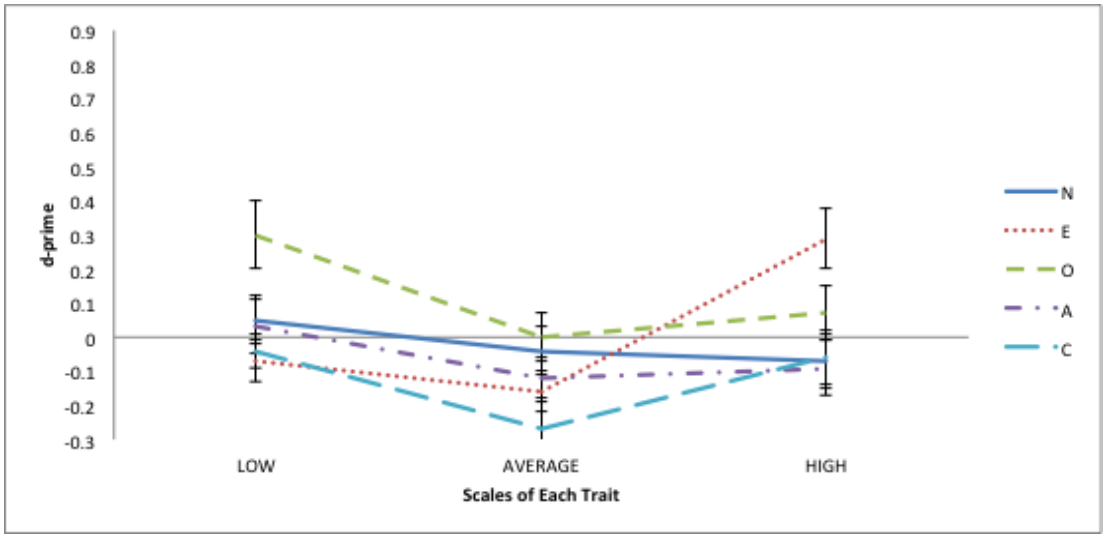


Fig. 6.7. Mean d-prime of each scale in each trait in the Video Only Condition in Study 8. Error bars represent standard error of the mean.

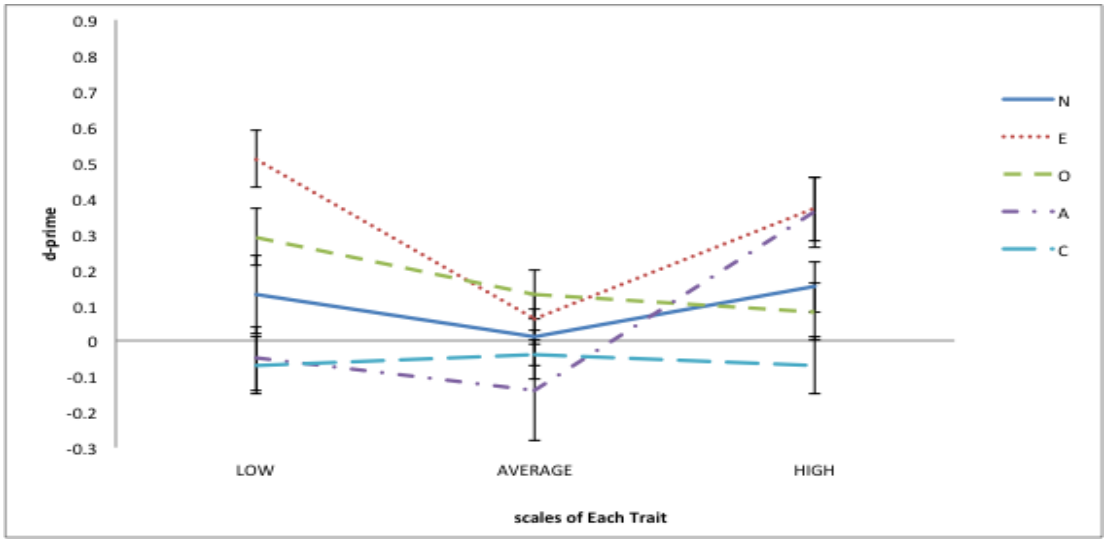


Fig. 6.8. Mean d-prime of each scale in each trait in the Audio Only Condition in Study 8. Error bars represent standard error of the mean.

To confirm these findings reported in Table 6.4 as well as Fig. 6.6 – Fig. 6.8, we conducted a 5 × 3 × 3 mixed design ANOVA, with the five traits and the three scales of each trait as the within-subjects factors and the three conditions as the between-subjects factor; the dependent variable was d'. The results revealed main effects associated with the traits (F (4, 348) = 8.19, p < .001), the scales (F (2, 174) =

42.17, $p < .001$), and the conditions (F (2, 87) = 15.37, $p < .001$); additionally, there were interactions between the traits and the conditions (F (8, 348) = 2.19, $p = .027$), between the scales and the conditions (F (4, 174) = 4.14, $p = .003$), between the traits and the scales (F (8, 696) = 8.75, $p < .001$), and an interaction among the traits, the scales and the conditions (F (16, 696) = 4.12, $p < .001$).

For the main effect of condition, we carried out post hoc Bonferroni tests to examine the differences of judgmental accuracy among the three conditions. The findings revealed that overall perceivers performed best in the condition of Video and Audio and worst in the condition of Video Only (Video & Audio > Video Only, $p < .001$; Audio Only > Video Only, $p = .007$; Video & Audio > Audio Only, $p = .056$). These data suggest that perceivers were generally most effective in making trait inferences of the targets while having access to both visual and auditory information of the targets, and less effective while having access to only visual information.

As with Study 7, to examine whether there was a similar type of trend in judgmental accuracy of the Big Five traits for each type of information, we carried out a 5×3 repeated measures ANOVA in each condition, with the five traits and the three scales as the within-subjects factors. As demonstrated in Table 6.5, which reports the corresponding results of the ANOVA tests in the three conditions, there was the same kind of trend in judgmental accuracy in each condition – a main effect of traits, a main effect of scales as well as an interaction between the traits and the scales. Additionally, the main effect associated with the scales and the interaction between the traits and the scales were strongest in the Video and Audio Condition.

Table 6.5. F values of the repeated measures ANOVA in each information condition (Video & Audio, Video Only, and Audio Only) in Study 8

ANOVA Test	Video & Audio Condition	Video Only Condition	Audio Only condition
Main Effect of Traits	F (4, 116) = 3.90*	F (4, 116) = 3.41*	F (4, 116) = 4.95**
Main Effect of Scales	F (2, 58) = 40.87**	F (2, 58) = 9.34**	F (2, 58) = 7.10*
Interaction (Traits × Scales)	F (8, 232) = 7.83**	F (8, 232) = 3.83**	F (2, 232) = 4.95**

Note: * $p \leq .01$, ** $p \leq .001$

The U-shaped Pattern of Judgmental Accuracy in Each Trait in Each Condition

Video & Audio Condition

In accordance with Study 7, we have already know that perceivers could make above-chance accurate judgments about the Big Five traits (except for A) based on a series of videos with sound across different scenarios, and the performance of guessing each trait formed a U-shaped pattern. Would these findings be replicated based on the Conversation Scenario only?

In the Video and Audio Condition, the average d' of each trait across the three scales was calculated as overall judgmental accuracy in each trait; the average d' was .06 (*SD* = .29), .36 (*SD* = .36), .26 (*SD* = .37), .14 (*SD* = .32) and .25 (*SD* = .33) for N, E, O, A, and C respectively. These values were significantly above chance except for the d' of N according to one-sample t tests (N: t (29) = 1.07, p = .295; E: t (29) = 5.37, $p < .001$; O: t (29) = 3.76, p = .001; A: t (29) = 2.30, p = .029, and C: t (29) = 4.17, $p < .001$), suggesting that perceivers could correctly guess the five traits except for N after viewing a short video with sound in the Conversation Scenario. Table 6.4 and Fig. 6.6 show that, as with Study 7, perceivers

did not uniformly make accurate judgments of each scale in each trait: There was no evidence of systematic judging for the targets who were in the middle of each trait, but perceivers were effective in judging the cases of low and high E, low and high O, low and high C, as well as low N and high A.

A one-way repeated measures ANOVA was conducted for each trait, with the three scales as the within-subjects factor. The results revealed that the d' values across the three scales in each trait were significantly different except for the trait O (N: F (2, 58) = 4.65, p = .013; E: F (2, 58) = 46.80, p < .001; O: F (2, 58) = 2.52, p = .089; A: F (2, 58) = 13.23, p < .001; C: F (2, 58) = 12.17, p < .001). Post hoc Bonferroni tests for each trait (except for O) revealed the following results. For the trait N, performance in guessing the targets with low N was greater than the targets with high N (p = .032); there were no differences between the judgments of the other scales. For the trait E, perceivers were most effective in making judgments about the targets with high E than the targets who either had low E (p < .001) or average E (p < .001), and they performed slightly better in judging the targets with low E than the targets with average E (p = .056). For the trait A, perceivers were better at identifying the targets with high A than targets who were low (p = .012) or average (p < .001) in A, but there was no difference between performance in low N and average N. For the trait C, perceivers performed equally well in identifying the targets in the two extreme scales, which was better than guessing the targets within the average scale (ps = .001).

Overall, perceivers were able to make accurate judgments of each trait except for N, and the accuracy in judging these traits formed the same kind of U-shaped trend as found in Study 7. Though U-shape curve for N was absent (see Fig. 6.6), perceivers were still relatively good at guessing the targets with low N. The

inconsistencies in the inferences of N and A between Study 8 and Study 7 might be because the current study was based only on the Conversation Scenario whereas Study 7 calculated the aggregate accuracy across the three scenarios.

Video Only Condition

Would the findings in the Video and Audio Condition be generalized to the other information conditions? In the Video Only Condition, the average d' of each trait across the scales was -.02 (*SD* = .26), .02 (*SD* = .30), .13 (*SD* = .29), -.06 (*SD* = .25) and -.12 (*SD* = .26) for N, E, O, A, and C respectively, among which only the d' of O was significantly above chance (t (29) = 2.39, p = .024) while the d' of C was worse than chance (t (29) = -2.55, p = .016) according to one-sample t tests of each d', suggesting that perceivers could only correctly guess the trait O and had special difficulty guessing the trait C after observing the target in a silent video. According to Table 6.4 and Fig. 6.7, perceivers were merely effective in guessing the cases of high E and low O and were rather poor at guessing the cases of average E and average C.

A one-way repeated measures ANOVA was conducted for each trait, with the three scales as the within-subjects factor. The results revealed that the d' values across the three scales of each trait were significantly different except for the traits N and A (N: F (2, 58) = .98, p = .382; E: F (2, 58) = 18.06, p < .001; O: F (2, 58) = 4.39, p = .017; A: F (2, 58) = 1.29, p = .284; C: F (2, 58) = 3.70, p = .031). Post hoc Bonferroni tests for each trait (except for N and A) revealed the following things. For the trait E, perceivers were better at judging the targets with high E than the targets who either had low E (p = .002) or average E (p < .001), and they performed equally in judging the targets who were low or average in E. For the trait O, perceivers were only more effective in identifying the targets with low O than the targets with

average O (p = .035). For the trait C, perceivers were better at identifying the low C than the average C (p = .035) though they were generally poor at guessing the trait C.

In summary, when only visual information of the target was available, overall, perceivers could only guess the trait O; the performance in judging the traits E, O and C demonstrated a U-shaped trend where judgmental accuracy was greater at the extremes (low or high) than in the middle.

Audio Only Condition

In the Audio Only Condition, the average d' of each trait across the scales was .10 (SD = .35), .31 (SD = .33), .17 (SD = .30), .06 (SD = .44) and -.06 (SD = .30) for N, E, O, A, and C respectively; except for the d' of E (t (29) = 5.10, p < .001) and O (t (29) = 2.39, p = .024), the d' values of the other traits were not above chance according to one-sample t tests, suggesting that perceivers could only correctly guess E and O while having access specifically to auditory information of the target. As displayed in Table 6.4 and Fig. 6.8, perceivers were good at identifying the cases of low and high E, high N, low O as well as high A but consistently failed to guess the targets who were in the average scale of any trait.

A one-way repeated measures ANOVA was conducted for each trait, with the three scales as the within-subjects factor. The results showed that the d' values across the scales of each trait were significantly different except for N and C (N: F (2, 58) = 1.10, p = .341; E: F (2, 58) = 12.60, p < .001; O: F (2, 58) = 3.22, p = .047; A: F (2, 58) = 8.00, p < .001; C: F (2, 58) = .06, p = .939). The results of post hoc Bonferroni tests for each trait (except for N and C) were as follows. For the trait E, perceivers were equally good at guessing the targets within low and high E, and the performance in both cases was better than that in average E (low E > average E, p < .001; high E > average E, p = .022). For the trait O, there was no difference

between each pairwise comparison. For the trait A, perceivers were better at judging the targets with high A than the targets who had low (p = .004) or average A (p = .002).

In summary, in the Audio Only Condition, perceivers were generally able to guess the traits E and O, and the performance in judging E formed a typical U-shaped pattern where the accuracy in the average scale was not above chance but the accuracy in both extreme scales was greater than chance and better than that in the average scale. In contrast, although the trends in the traits N, O, and A were not as strong as E, perceivers still demonstrated effective performance in the cases of high N, low O and high A.

6.4 General Discussion

6.4.1 How Well Can Perceivers Form a First Impression of the Big Five Traits From Thin Slices of Behavior?

As demonstrated in Study 7, after watching a brief video with sound, perceivers were generally capable to form an accurate first impression on each of the Big Five traits with the exception of A. Using the video with sound in the Conversation Scenario, Study 8 generally replicated these results but with a few small exceptions.

These findings are consistent with the previous accuracy research of personality judgments, revealing perceivers' abilities to evaluate E and C by observing thin slices of behaviour (e.g., McLarney-Vesotski et al., 2006; Carnet et al., 2007). Moreover, the present study suggests that perceives could even make an accurate inference about the other three traits, which has not been inconsistently demonstrated in the extant literature on accuracy of first impressions of personalities

(e.g., Albright et al., 1988; Watson, 1989; Brokenau et al., 2004; McLarney-Vesotski et al., 2006; Carnet et al., 2007).

Why is perceivers' capacity for judging E and C consistently replicated whereas their ability to infer the other traits tends to be inconsistent depending on the combination of visual and auditory information across various studies? According to Funder's Realistic Accuracy Model (RAM; 1995), achievement of accuracy in personality judgments requires, in part, valid and available behavioral cues. In a study concerning connections between the Big Five personality traits, the behaviours by which they are manifest, and the behaviors by which they are judged, Funder and Sneed (1993) suggest that for the traits E, C and A, perceivers draw upon the behavioral cues they think they use, and the cues they use the most are those that are most likely to be valid, but these conclusions less apply to the traits of N and O. Meanwhile, the study reveals that the number of accessible behaviour cues varies with different traits: The trait E had the greatest number of valid and available behaviour cues followed by the traits C and A while the traits N and O possess relatively few or conflicting behaviour cues.

Given these connections between the five traits and their behavioral manifestations, research on accuracy of personality judgments might involve situations (which the targets experienced) varying in valid and available behaviour cues in relation to different types of traits. Many situations involving visual and auditory cues could have offered sufficient information to make an accurate judgment of E and C but not necessarily have supplied valid and available cues for guessing the traits N, O, and A. For example, in Borkenau et al.'s study (2004), only some of the 15 tasks seemed to be especially diagnostic of the trait O: Perceivers' ratings of O agreed most strongly with self and acquaintance ratings based on the

pantomime task where the targets described multiple uses of a brick and demonstrated them by pantomime. In the current study, after briefly observing the targets in videos with sound, either across the three scenarios or only in the Conversation Scenario, perceivers seemed to have access to sufficient behavioral cues of each trait (except for A across the three scenarios and N in the Conversation Scenario), and were duly able to form an accurate first impression on the Big Five traits of the targets. The capacity for personality inferences was investigated in diverse naturalistic social contexts and measured with d' values using the more sensitive and powerful signal detection method, thus offering compelling ecological validity.

In addition, based on the Conversation Scenario, Study 8 extended the findings with new evidence suggesting that perceivers could correctly judge the trait O while having access to visual only information and could accurately identify the traits E and O after hearing the target speaking in the conversation. Furthermore, perceivers' capacity for inferences of the five traits was best in the Video and Audio Condition and worst in the Video Only Condition. In other words, it seems that a video with sound provided informative visual and auditory behaviour cues by which perceivers were able to form a first impression on the Big Five traits (except for N) of the targets. However, why did perceivers demonstrate fewer abilities to infer the five traits while viewing the same video without sound than hearing only the sound of the video?

In the Conversation Scenario, the targets were covertly videoed while having a brief conversation with the researcher by answering some questions about themselves, such as "where are you from", "what are you studying", "what do you usually do at weekends", "why did you choose to study in the university", "what do

you think of the food at the cafeteria" and so on. Accordingly, a sound video could contain plenty of behaviour cues, either visually or audibly observable actions like facial expressions, bodily gestures, loudness of the voice, along with everyday behaviours expressed in the verbal contents. Hence, while viewing a video with sound, perceivers could not only observe visual and auditory behaviours but also had a chance to perceive the target's daily behaviour and even to link the observable behaviour with the verbal contents. For example, if watching a video where a target was very expressive and made many bodily movements while talking about an exciting activity he/she experienced, we would be likely to attribute the targets' actions to signs of being extrovert. By contrast, the visual only behavioral cues might be relatively less informative owing to underlying ambiguities of the actions. In the above example, if watching the same video without sound, we might perceive the same actions in different ways – Judging from the frequency of the target's bodily movements, we might regard the target as being nervous (leading to a judgment of high N) or being excited (leading to a judgment of high E). In spite of the lack of visual cues, auditory information could be less vague and deliver fewer conflicting cues as opposed to visual only information in that verbal contents usually involve much clearer information, which enables perceivers to make more reliable personality judgments of the target. In the above case, even if perceivers could not visually observe targets' behaviour, they would still be able to form somewhat accurate impressions of the target by drawing upon the verbal contents such as the exciting activity the target related.

6.4.2 The U-Shaped Pattern of Judgmental Accuracy in the Big Five Personality Traits

Apart from examination of overall accuracy for each of the Big Five traits, the present study examined the capacity to infer different scales of each trait. Generally, there was a similar kind of U-shaped pattern in perceivers' accuracy of personality judgments, in which perceivers could not guess the targets who scored average on a trait but were relatively good at detecting the targets who scored either low or high on one or more traits.

As shown in Study 7, after watching a short video (with sound) across the three scenarios, perceivers were especially effective in guessing the targets who were low or high in N, low or high in O, low or high in C, high in E and low in A. Based on the videos (with sound) in the scenario of Conversation, Study 8 replicated these findings to a large degree: Perceivers demonstrated systemic performance in identifying the targets who scored either low or high on the traits of E, O and C and the targets who were low in N and high in A. To summarize these data another way: In the video and audio condition, there is the same kind of trend in perceivers' capacity to infer each trait – with worst performance in the average scale and better performance in either one or both of the two extreme scales (low and high).

Why would this U-shaped trend occur in forming first impressions of the Big Five traits? As with the explanation for the U-shaped pattern in the ability to infer empathic traits (details appear in Chapter 3), people who have an unusual personality might behave in some ways different from the average person; for instance, an extravert usually is more talkative while an introvert is rather quiet in social interactions, by which perceivers could distinguish them from average persons. An alternative possibility is that people could have developed to be able to recognize

persons who are low or high in personality dimensions during the everyday practice of person perception in their social lives. For example, while communicating with an agreeable person, we usually could experience calm and enjoyment; while interacting with a person high in neuroticism we might feel stressful and nervous. From such experiences of social interactions we might have learned to identify those who have unusual personalities.

Both studies presented in this chapter suggest that the strength of U shape differs depending on traits – For some traits, the U shape is strong with better performance in both low and high scales while for the others the U shape is relatively weak with effective performance in only one of the extreme scales. Why are some traits easier to infer than others? As discussed, the Big Five traits are related to different behavioral manifestations (Funder & Sneed, 1993). That is, access to sufficient observable behaviour cues may vary across the traits depending on the context. In addition, according to previous studies (e.g., Albright et al., 1988; Watson, 1989; Norman & Goldberg, 1966; Borkenau & Liebler, 1992b; Funder & Sneed, 1993; Carney et al., 2007), some traits (such as E and C) seem to be relatively easy to infer due to being less inward and more visible while other traits (such as N) are more difficult to judge because of being more inward and less visible. Taking these factors into account, it seems understandable that perceivers did not demonstrate a uniform U-shaped pattern in the judgments of each trait.

Finally, based on the Conversation Scenario, there was a similar kind of U-shaped pattern in the Video only Condition where perceivers were better at identifying high E and low O and in the Audio Only Condition where perceivers made systemically accurate judgments of low and high E, low O and high A. Taken together, these results suggest that visual or auditory information is sufficient to

make an accurate judgment about high E and low O, but it seems that auditory cues (including verbal contents) played a key role in detecting the targets low in E and high in A.

6.5 Conclusion

To conclude, this chapter reported two studies based on thin slices of behaviour, revealing that perceivers can form well-above chance accurate first impressions of the Big Five traits while having access to both visual and sound cues, and also can make an accurate judgment of one or more traits while being exposed to visual only or audio only information. Moreover, performance of guessing each trait demonstrates a similar kind of U-shaped pattern where perceivers could not guess the average scale of each trait but were effective in guessing either one or both of the extreme scales in one or more traits. Finally, while performance in making judgments about personalities was not different across the three scenarios, performance was best in the Video and Audio Condition and worst in the Video only Condition.

CHAPTER SEVEN

Ordinary People's Commonsense Views About First Impressions of Personalities

"Though the full significance of man's relations to man may not be directly evident, the complexity of feelings and actions that can be understood at a glance is surprisingly great. It is for this reason that psychology holds a unique position among the sciences. "Intuitive" knowledge may be remarkably penetrating and can go a long way toward the understanding of human behavior, whereas in the physical sciences such common-sense knowledge is relatively primitive. If we erased all knowledge of scientific physics from our world, not only would we not have cars and television sets and atom bombs, we might even find that the ordinary person was unable to cope with the fundamental mechanical problems of pulleys and levers. On the other hand, if we removed all knowledge of scientific psychology from our world, problems in interpersonal relations might easily be coped with and solved much as before." (Heider, 1958, p.2)

7.1 Introduction

We have a great deal of naïve physics knowledge that enables us to master our daily life successfully; likewise, as Heider claimed in the above quotation, we also possess a lot of commonsense knowledge about the mind, which allows us to perceive the inside of another person that is physically invisible, enabling us to navigate our social world smoothly. Psychologists like Hedier (1958), Fletcher (1984), and Kelly (1992) have paid much attention to commonsense psychology for both theoretical and applied reasons.

For theoretical considerations, commonsense psychology has been regarded as a valuable resource of scientific psychology (Heider, 1958; Fletcher, 1984; Kelly, 1992). Fletcher (1984) argues that psychological theories at least embrace the class of fundamental assumptions including a set of shared basic commonsense knowledge about the nature of the social and physical world even in the absence of explicit recognition of the fact. Heider (1958) not only debated the importance of commonsense psychology in studying person perception, but also proposed an elaborate conceptual framework built upon naïve psychology to interpret the processes by which people perceive each other in daily life.

For practical reasons, commonsense psychology guides our everyday actions towards other people (Heider, 1958). We talk about other people using everyday mentalistic terms and personality labels; we explain what other people did in the past and predict what they will do in the future using various strategies of mentalising; we form impressions of other people by observing what and how they do in their lives. Furthermore, our commonsense views about psychological events seem to be accurate to some degree. For instance, to study "lay knowledge of the principles of psychology", Houston (1983; 1985) invented a questionnaire including an array of

multiple-choice questions about various memory and learning phenomena, with each item stated in everyday language. Participants were required to respond to each item by picking a choice from the given options. Results show that they answered the majority of items more accurately than would be expected by chance. Houston thus claimed, "A great many of psychology's basic principles are self-evident. One gets the uneasy feeling that we have often been dealing with the obvious and did not know it" (1983, p.207).

In contrast, some findings are counter-intuitive. In the classic research of obedience to authority (Milgram, 1963), participants were asked to obey an authority figure who instructed them to give severe electric shocks to a learner each time the leaner answered a question incorrectly. Milgram (1963) found that about 65% of participants obeyed authority and gave such apparently dangerous shocks, and only 35% refused. Milgram subsequently investigated whether or not these results are intuitively surprising by conducting an attitude scale to examine how people predict others would perform when commanded to administer powerful shocks to another person. Results demonstrated that respondents underestimate the scale of obedience to authority in wrongly predicating that only a small minority would go through to the end of the shock series.

Do people hold accurate commonsense views about people's capacity to form an accurate intuitive impression of personality? Inspired by the ideas of Milgram, we created a survey to address this question.

Chapters 3 to 6 reported a series of studies examining how well people can guess personalities of non-acquaintances based on thin slices of behaviour. The findings reveal that people are capable of inferring the trait of empathy as well as some dimensions of the Big Five traits after observing a brief sample of behaviour,

whether the behaviour is visual or auditory, dynamic or static. Moreover, people have a great talent for guessing those who have an unusual personality in contrast with those who have an average personality. Additionally, Study 2 (see Chapter 3) suggests that people are inclined to have confidence in the first impressions that they formed of other people even though in fact the level of confidence was not related with accuracy in personality judgments. As with Milgram, when designing the attitude scale according to the experimental results, we invented a survey in order to explore the extent to which people possess accurate commonsense views that are consistent with the actual findings we obtained.

7.2 Method

7.2.1 Summary

The online survey system Qualtircs was used to carry out a survey ('Attitude Towards People's First Impressions on Personality') including 11 items designed in accordance with the experimental findings reported in Chapter 3 – Chapter 6. Participants were asked to fill in the survey on the website hosted by Qualtrics. The procedure was scrutinized and approved by the Faculty of Science ethics committee in the University of Nottingham Malaysia Campus, which was constituted and operated according to guidelines prescribed by the British Psychological Society.

7.2.2 Participants

Two-hundred and fifty-three participants from the University of Nottingham Malaysia Campus were recruited online, but 85 of these responses were incomplete and therefore deleted. Thus, there was a valid sample of 168 participants (64 males, 103 females, & 1 skipping the gender information), with a mean age of 22 years (three participants did not provide any information of age).

7.2.3 Survey

The survey ‘Attitude Towards People’s First Impressions on Personality’ consists of 11 items in the form of question, together with a 7-point Liker scale for each item. The 11 items concerned several aspects of people’s commonsense views about first impressions of personalities, involving the frequency (Item 7) and importance (Items 8, 9 & 11) of first-impression formation, and confidence (Item 10) and accuracy (Items 1, 2, 3, 4, 5, 6) of personality judgments. The 11 questions were as follows.

1. How accurate do you think people’s first impressions of personalities are? (1 = not accurate at all, 7 = extremely accurate)
2. People are more accurate in their first impressions of people who have unusual personalities than in their first impressions of people who have average personalities. Do you agree with this opinion? (1 = strongly disagree, 7 = strongly agree)
3. How accurately do you think people can assess personality after viewing a 10-second sample of behaviour in a video? (1 = not accurate at all, 7 = extremely accurate)
4. How accurately do you think people can assess personality after watching photographs (for ten seconds) of a person taken while carrying out some action? (1 = not accurate at all, 7 = extremely accurate)
5. How accurately do you think people can guess personality after looking at a passport photograph for ten seconds? (1 = not accurate at all, 7 = extremely accurate)

6. How accurately do you think people can assess personality after hearing a person speak for 10 seconds (in the absence of any visual information)? (1 = not accurate at all, 7 = extremely accurate)
7. How common is it for people to form an impression of personality on first meeting someone? (1 = not common at all, 7 = extremely common)
8. How important is it to form an impression of personality on first meeting someone? (1 = not important at all, 7 = extremely important)
9. How important is it to make an accurate assessment of someone's personality on first meeting? (1 = not important at all, 7 = extremely important)
10. How confident do you think people feel they have made an accurate assessment of personality on first meeting someone? (1 = very low confidence, 7 = very high confidence)
11. A first impression of someone's personality is enduring rather than being quickly replaced by better informed impression as acquaintance grows. Do you agree with this opinion? (1 = strongly disagree, 7 = strongly agree)

7.2.4 Procedure

We generated the survey using the online platform Qualtrics, along with an instruction form, participant's information form and a consent form. The 11 items in the survey were presented in random order determined by the Qualtirics system. After launching the survey on Qualtrics, the link was sent to all students in the

University of Nottingham Malaysia Campus by email. All data were collected and recorded by Qualtrics.

7.3 Results and Discussion

Table 7.1 displays the mean response to each item (according to the 7-point scale), along with the values of one-sample t tests comparing each mean score with the neutral point (four) of the scale. If the mean is significantly greater than 4, then participants hold a positive attitude towards the statement; if the mean is less than 4, then participants have a negative view.

Table 7.1. Means (and standard deviations) of responses to each item, along with values of one-sample t tests (df = 167) of each mean in the survey

Item No.	*M* (*SD*)	*t*	*p*
1	3.74 (1.30)	-2.62	.01
2	4.64 (1.35)	6.17	< .001
3	3.65 (1.39)	-3.29	.001
4	2.89 (1.40)	-10.34	< .001
5	2.23 (1.33)	-17.21	< .001
6	3.49 (1.53)	-4.29	< .001
7	5.78 (1.16)	19.88	< .001
8	5.20 (1.33)	11.68	< .001
9	4.52 (1.45)	4.69	< .001
10	4.51 (1.33)	4.93	< .001
11	4.17 (1.51)	1.43	.15

As shown in Table 7.1, participants believed that forming a first impression of personality is a very common phenomenon in everyday life (Item 7): t (167) = 19.88, p < .001. In their opinion, it is of importance to form a first impression of personality (Items 8 & 9), whether the impression is accurate or not. And when asked

whether a first impression would be quickly replaced by a better-informed impression as acquaintance increases (Item 11), they held a neutral attitude ($p = .15$). According to the results of Items 1, 3, 4, 5, and 6, participants held fairly negative views about whether people can form an accurate first impression of personality, either at an overall level ($p = .01$) or based on brief exposure at different types of cues such as videos, audios and images ($ps \leq .001$). However, they strongly agreed that people are more accurate in their first impressions of people who have unusual personalities than in their first impressions of people who have average personalities (Item 2): $t\ (167) = 6.17$, $p < .001$. Finally, participants agreed that people tend to feel confident of being able to form an accurate first impression of personality (Item 10): $t\ (167) = 4.93$, $p < .001$.

To organize these results, Table 7.2 demonstrates the range order of the means of the 11 items. Apparently, the means in the items that are relevant to accuracy of forming a first impression of personality (Items 1, 3, 4, 5, & 6) ordered closely at the bottom of the table while Item 2 (which relates to people's capacity for inferences of unusual personalities) was listed at the top. These results should not be explained by an order effect because all 11 items were presented to participants in a random order. Even so, given that we do not know the calibration of the 7-point scale, to be cautious, we compared related items by conducting correlation analyses and paired-samples t tests or a repeated measures one-way ANOVA.

Table 7.2. Range order of the means (and standard deviations) of responses to each item, along with values of one-sample t tests of each mean (df = 167) in the survey

Item No.	*M* (*SD*)	*t*	*p*
7	5.78 (1.16)	19.88	< .001
8	5.20 (1.33)	11.68	< .001
2	4.64 (1.35)	6.17	< .001
9	4.52 (1.45)	4.69	< .001
10	4.51 (1.33)	4.93	< .001
11	4.17 (1.51)	1.43	.15
1	3.74 (1.30)	-2.62	.01
3	3.65 (1.39)	-3.29	.001
6	3.49 (1.53)	-4.29	< .001
4	2.89 (1.40)	-10.34	< .001
5	2.23 (1.33)	-17.21	< .001

As shown in Table 7.2, Items 8 and 9, both concerning the importance of first impression formation, were in close proximity. A Pearson correlation revealed media significant association between the two items, r = .51, p < .001, and a paired-samples t test showed a significant difference (t (167) = 6.35, p < .001), suggesting that participants were more positive in response to the question about the importance of making a first impression (Item 8) than the question about the importance of forming an accurate first impression (Item 9).

According to Table 7.2, participants seemed to disagree that people can form accurate impressions of personality on first meeting (Item 1), but they agreed that people are prone to believe that they have formed an accurate first impression of another person (Item 10). A paired-samples t test confirmed that participants were more positive in response to Item 10 than in response to Item 1 (t (167) = 5.92, p < .001). These commonsense views are consistent with the empirical data reported in

Study 2 – Perceivers seemed to be overconfident in their personality judgments even though those who were most confident were not necessarily most accurate.

Participants held a negative attitude towards whether people can form an accurate first impression of personality (Item 1), as demonstrated in Table 7.2. Nevertheless, they seemed to have strong feelings that people are better at identifying extreme personality as opposed to identifying average personality (Item 2), t (167) = 6.37, $p < .001$.

Items 3, 4, 5 and 6 examined participants' attitudes towards judgmental accuracy based on different types of information. Pearson correlations revealed significant associations (Items 3 & 4: $r = .54$, $p < .001$; Items 3 & 5, $r = .40$, $p < .001$; Items 3 & 6: $r = .45$, $p < .001$; Items 4 & 5: $r = .51$, $p < .001$; Items 4 & 6: $r = .36$, $p < .001$; Items 5 & 6: $r = .38$, $p < .001$). In other words, those who thought that it is possible to make accurate judgments based on first impressions did so consistently; those who thought that it is difficult to make accurate judgments also did so consistently. A repeated measures one-way ANOVA showed a significant difference among these four items, F (3, 501) = 62.20, $p < .001$, and post hoc Bonferroni tests demonstrated the following results: (1) In participants' views, it is least likely to form an accurate first impression based on passport photograph (Item 5) ($ps < .001$); (2) accuracy based on the videos and sound (Items 3 & 6) is equally good ($ps < .001$), both of which are better than accuracy related to photographs capturing some action (Item 4) ($ps < .001$). In other words, these results suggest that participants held a position suggesting that people form an accurate impression of personality based on behaviour, whether the behavior is presented in video, sound, or photograph, but not based on still images (such as passport pictures) that are not informative about behaviour. In addition, participants seemed to judge that video and audio provide

equally useful information for forming an accurate first impression, which is more informative than a photograph that captures activity.

To summarize, the data of the survey suggest that ordinary people hold the following commonsense views about first impressions of personalities: (1) People believe it is very common to form a first impression of personality in social life, and this first impression is very important, whether it is accurate or not; (2) overall, they believe that people cannot infer personalities of others very well on first meeting, but they believe that people are inclined to have confidence in their first impressions of personality; (3) they believe that people are better at identifying those who have unusual personalities than those who have average personalities; (4) they believe that people cannot guess others' personalities after viewing a short sample of videos, or hearing snippets of audio, or briefly observing photographs, and that people make personality judgments according to the behaviour of others; (5) they believe that people can make more accurate first-impression judgments of personalities while having access to video or sound information than by observing photographs that are related to some action.

According to these findings, ordinary people apparently tend to underestimate the ability to form an accurate first impression of personality from a brief sample of behaviour. Apart from this, most of their commonsense views captured by the survey are inclined to be consistent with the evidence of actual performance (presented in the previous chapters) with the exception that the common sense view is not entirely consistent with the kinds of evidence that are more or less informative about personality.

CHAPTER EIGHT

General Discussion and Overall Conclusion

8.1 Summary of Empirical Studies

In responding to the question asked in Chapter 1 "how well can people infer personality traits on first meeting", I reported a series of studies in Chapter 3 to Chapter 6 and a survey in Chapter 7. In the following, I will summarize the empirical work presented in this thesis.

In Chapter 3, two studies were designed to explore whether people can guess how empathizing another person is after watching a short video. In Study 1, perceivers were asked to guess EQ after viewing a video (without sound) in which the target was experiencing one of four given scenarios (the Compliment, the Story, the Joke and the Waiting). Results demonstrate that overall perceivers could form an accurate first impression of target empathic traits, and they were especially good at identifying the targets who either had low or high EQ but were not effective in identifying the targets who scored average in EQ. Additionally, Study 1 revealed a significant association between perceivers' ratings of the targets' EQs and perceivers' self-reported EQs (which were fed back to the perceivers before proceeding to guess the EQ of the target), suggesting that perceivers to some extent assumed similarity when making first-impression judgments of empathic traits based on thin slices of behaviour.

Given the fact that the EQs of the targets varied according to scenarios experienced by the targets, the EQ might reflect not the target's enduring empathic trait but the target's temporary empathic state induced by a certain scenario. The purpose of Study 2 was thus to clarify whether or not perceivers genuinely assessed empathy as a trait. Three samples of new video clips were collected, each involving one of the three scenarios (the Conversation, the Joke, & the Screen Test) experienced by each target. Perceivers were asked to make a judgment of the target's EQ after viewing an audio video of the target either in the Conversation, the Joke or the Screen Test Scenario. Unlike Study 1, after completing the EQ questionnaire, perceivers were required to guess their own EQ instead of receiving informative feedback. In addition, they had to evaluate confidence in the EQ judgments they made about the targets and themselves. The results reveal the following things: (1) The principal findings reported in Study 1 were replicated; that is, perceivers could generally guess the EQs of the targets, and performance in detecting different EQ scales formed a U-shaped trend, in which perceivers were better at guessing the targets who fell into the extreme EQ scales than the targets who had average EQ; (2) there was no difference in the overall accuracy in estimations across the three scenarios; (3) perceivers' estimation of the targets' EQs significantly correlated with the EQ they guessed about themselves but not their actual EQs as determined by their responses to the EQ questionnaire; (4) perceivers generally felt confident in their judgments of the EQ about the targets and themselves though they had more confidence in self-perception; (5) nevertheless, judgmental accuracy of EQ was not associated with confidence in being able to make an accurate judgment.

Taken together, Studies 1 and 2 suggest that perceivers can form an accurate first impression of empathic traits of others after briefly observing them in a video,

whether the video is audible or silent. Does this mean that it is necessary for perceivers to see an animation of behaviour in inferring the empathic capacity of another person? Chapter 4 presented three studies based on the stimuli of the Joke Scenario (taken from Study 2) in order to address this question. In Study 3, perceivers were asked to guess the EQ of the target either after watching a video (without sound) or observing three photographs taken from the same video. The perceivers were equally good at guessing EQs of the targets in both conditions, and, as with Studies 1 and 2, the data formed the same U-shaped pattern in the sense that perceivers were effective in guessing who had extreme EQ but not effective in guessing who had average EQ.

On what basis were perceivers able to infer empathic traits? Did perceivers make accurate judgments of EQ based only on the facial appearance of the targets or based on the behaviour of targets as suggested by the targets' facial expressions? To examine this matter, Study 4 required perceivers to guess the EQ of the target either in the first photograph condition or the last photograph condition; the former involved a neutral pose before the target stated reading the joke whereas the latter captured the apex of the target's expression as he or she delivered the punch line of the joke. Results show that (1) the U-shaped pattern of judgmental accuracy was only replicated in the last photograph condition, (2) performance in the last photograph condition was much better than in the first photograph condition, and (3) perceivers could not make systemically accurate judgments of EQ in the first photograph condition, either at an aggregate level or at a fine-grained level. In summary, Studies 3 and 4 converge in suggesting that perceivers were able to form an accurate first impression of empathic traits based on the behaviour of the target and not based on

the facial appearance of the target, whether the behaviour was communicated in an animation or in still pictures.

Having demonstrated the capacity for inferences of empathic trait on the basis of visual behaviour cues, would perceivers be able to identify who has low EQ and who has high EQ after hearing the target talking for a few seconds? The goal of Study 5 was to answer this question. Perceivers were asked to judge the EQ of the target while watching a short video (without sound) or listening to the soundtrack derived from the same audio video. Results reveal that (1) performance in the video condition formed the same U-shaped trend as found in the previous studies, (2) perceivers were only effective in identifying the targets who had high EQ in the audio condition, and (3) perceivers performed slightly better in the video condition than the audio condition overall. Hence, Study 5 has extended the previous findings by providing new evidence on the ability to infer empathic traits merely from vocal behaviour cues.

Additionally, Study 5 included two surveys to examine the relationship between assessments of the targets' EQs and ratings of the targets' expressivity. Independent judges were asked to assess expressivity of the targets after either viewing a silent video or listening to an audiotrack (in the Joke Scenario). The results reveal that either in the video or the audio condition, independent judges' ratings of the targets' expressivity were significantly associated with perceivers' judgments of the targets' EQs. Following these preliminary results, Study 6 reported in Chapter 5 was conducted to further investigate the correlation between evaluation of expressivity and judgments of EQ. Perceivers had to rate both expressivity and the EQ of the target while watching a video with sound in the Joke Scenario (taken from Study 2). The results indicated a strong association between the ratings of EQ and

the estimates of expressivity, and this correlation was not caused by cross contamination from answering one question (of expressivity or of EQ) to another, and cannot be explained by saying that perceivers treated the two questions (of expressivity and EQ) identically.

Given the findings of the impressive U-shaped pattern for perceivers' capacity to infer empathic traits, would the U-shaped pattern also be apparent in inferences of the Big Five personality traits? Chapter 6 reported two studies investigating how well people form a first impression of the Big Five traits based on thin slices of behaviour.

In Study 7, using the same video stimuli as Study 2, perceivers were asked to guess the Big Five traits of each target after viewing an audio video either in the Conversation, the Joke or the Screen Test Scenario. The findings were summarized as follows. (1) Perceivers demonstrated the same kind of trend towards performance of guessing the five traits in each scenario: There were main effects associated with the traits and the scales and an interaction between Trait and Scale. The interaction between Trait and Scale was also affected by the scenarios. (2) Accuracy in the judgments of the Big Five traits was not different between scenarios. (3) Perceivers could make an overall accurate judgment of each of the five personality dimensions except for the dimension of A across the three scenarios. (4) Performance in guessing each personality dimension formed a similar type of U-shaped pattern, in which perceivers could not correctly guess the targets who belonged to the middle scale of each trait but were effective in identifying the targets who scored low or high in one or more traits, and the accuracy in one or both of the extreme scales was greater than the average scale in each trait with the exception of the trait A: Perceivers were better at identifying the cases of low N, high E, low and high O, and low and high C.

Though the U-shaped trend was not apparent for the trait A, perceivers still were effective in detecting targets with low A.

To examine whether this sort of U-shaped trend for judgmental accuracy in the Big Five traits would be apparent with different types of behavioral information, Study 8 required perceivers to guess the five personality traits of each target either after viewing a video with sound (the Video and Audio Condition), a video without sound (the Video Only Condition) or hearing a soundtrack taken from the same video clip (the Audio Only Condition) based on the Conversation Scenario (taken from Study 2). Results were reported as follows: (1) There was a similar kind of trend towards judgmental accuracy in the Big Five traits for each type of information, including main effects related with the traits and scales as well as an interaction between the traits and the scales. (2) Overall, perceivers demonstrated best performance in inferring the Big Five traits after watching a video with sound and performed worst while viewing the same video without sound. (3) Performance in the Video and Audio Condition replicated the findings in Study 7 with a few small exceptions: Perceivers could generally make accurate judgments of the five traits except for N, and performance in each trait formed a similar kind of U-shaped pattern, in which perceivers consistently failed to identify the targets who were average in each trait but were effective in identifying the targets who were low in N, low and high in E, low and high in O, high in A, and low and high in C. (4) In the Video Only Condition, perceivers could only infer the trait of O, and the U-shaped trend demonstrated that perceivers were better at guessing the cases of low O and high E. (5) In the Audio Only Condition, perceivers could make overall accurate judgments of the traits of O and E, and were more effective in identifying the targets with high N, low and high E, low O and high A.

The studies presented in Chapters 3 to 6 repeatedly revealed an intriguing phenomenon of being able to form an accurate first impression of personality from a brief sample of behaviour, with special sensitivity in identifying those who have unusual personalities compared with those who have average personalities. What are the intuitions of ordinary people about our capacity to form accurate first impressions of personality?

Inspired by the ideas of Milgram (1963) in an investigation of people's intuitive perception in the phenomenon of obedience to authority, we developed a survey based on the findings of the eight studies examining ordinary people's commonsense views about first impressions of others' personalities. The data suggest that ordinary people believe that (1) it is very common and very important to form a first impression of personality in our social life, (2) people are not accurate in forming a first impression of personality, whether it is at an overall level or based on brief different types of information (e.g., video, picture, and audio), but personality inferences could be made from the behavior rather than just the appearance of another person, (3) people are capable of distinguishing persons with extreme personalities from the average person, and (4) people tend to feel highly confident of being able to make an accurate judgment of another's personalities. Apart from underestimating the abilities to infer personalities on first meeting, most of their commonsense views captured by the survey are basically consistent with evidence of actual performance (reported in the previous chapters).

8.2 How Well Can People Guess Personalities From a Brief Sample of Behavior?

8.2.1 Overall Accuracy of Personality Inferences

Overall, the eight studies have demonstrated perceivers' capacity to form accurate impressions of personalities on first meeting. People can correctly guess empathic traits after viewing a short video (either with or without sound), after observing still pictures capturing an activity, or after merely hearing a soundtrack for several seconds. They are able to draw inferences of many personality dimensions of the Big Five traits after briefly observing the target's behaviour or hearing the target speaking, either across different scenarios or only in the Conversation Scenario; they are also capable of inferring the trait O after watching the video without sound and making correct judgments of the dimensions of E and O while hearing sound from the video in the Conversation Scenario.

In Comparison with previous research on accuracy in first impressions of personality, the research reported in this thesis offers new information in answer to the question, "how well can people draw inferences of personalities?" First of all, in spite of adopting self-other agreement as a criterion of accuracy, unlike previous studies, the current research utilized signal detection method instead of correlation analysis to calculate self-other agreement of personality judgments; that is, the target's actual personality (measurable with the relevant questionnaires) served as a "signal", and only when perceivers correctly identified that "signal" could self-other agreement be counted. This method allows an assessment of accuracy that is independent of underlying base rates of targets' actual scales of personality traits (low, average, & high) and perceivers' response bias in a particular scale of each trait. Correlational measures, in contrast, can only tell us about the extent to which the

perceiver's ratings are associated with the target's self-ratings but not about accuracy based on the amount of elevation (i.e., positive or negative bias). Hence, signal detection is more useful in telling us about perceivers' capacity for personality inferences.

In addition, instead of using artificial stimuli, such as crafted trait-related scenarios and deliberate actions, the present research involved a large sample of stimuli collected from several naturalistic social settings where targets engaged in various trivial activities like having a casual conversation, telling a joke, being kept waiting and so forth, which could happen to anyone in everyday life. Besides, the context in which perceivers made personality judgments of the targets resembled the way in which people usually judge what another person is like in the real world. Taking consideration of these factors, we assume that the current findings of overall accuracy in personality inferences should have high external validity.

Moreover, the present findings have added a body of knowledge to the existing accuracy research of personality judgments. First of all, we conducted the initial investigation to examine the capacity for inferences of empathic traits from thin slices of behaviour, and provided compelling evidence indicating that people are able to form an accurate first-impression judgment of empathic capacity. Secondly, we examined people's abilities to infer personality traits based on a battery of observable behaviour samples in more than one situation, and found that people perform equally well in drawing personality inferences in different situations experienced by the same sample of targets. Thirdly, we explored influences of different types of behavioral cues (visual and sound, visual only, and sound only) on accuracy in personality judgments, conditions which have received little attention in extant research. Results indicated that specific kinds of information have varying

effects on judgmental accuracy depending on which traits are being inferred. For example, merely visual (either in animated or still form) or auditory information is sufficient for perceivers to make an overall accurate judgment of empathic traits though performance is slightly better after observing visual than auditory information. However, while having access to only visual or sound information, perceivers are only able to infer one or two personality dimensions of the Big Five traits, while they can infer the majority of the Big Five traits while having access to both visual and sound information.

8.2.2 The U-shaped Pattern in Accuracy of Personality Judgments

Most importantly, the research presented in the current thesis explored people's fine-grained abilities to infer different levels of particular traits, and revealed a striking U-shaped pattern of accuracy, in which perceivers can make systemically accurate judgments of targets who have extreme traits but are unable to make accurate judgments of targets who are average. The U-shaped trend emerged consistently across different scenarios and different types of target behaviour when perceivers were inferring empathic traits. Perceivers were equally effective in identifying targets who had either low or high empathic capacity after watching a video of the target (whether the video was audible or not) and after viewing one or three still photographs of the target engaging in some kind of behaviour; they were also good at identifying the targets with high empathy while listening to a snippet of sound of the target. In contrast, the strength of U shape for inferences of the Big Five traits was affected by several factors such as scenarios, personality dimensions, and different types of behaviour information. In general, the U-shaped trend was strong in the Conversation Scenario and in the Video and Audio Condition, demonstrating the same kind of U-shaped pattern for accuracy in inferring personality dimensions.

In the Video Only Condition, perceivers were effective only in making judgments of low O and high E, and in the Audio Only Condition, they were only good at guessing the cases of low and high E, high N, low O and high A.

In summary, it seems that people have a great talent for making an accurate inference of extreme personality traits after observing a brief sample of behaviour. How could this happen? According to the lens model, proposed by Brunswick (1955), social perception of uncertain (distal) events is dependent upon a process of inference from immediately observable (proximal) cues. Given the lack of one-to-one connection between the two, perceivers must consider the probabilities that proximal cues signal distal events. That is, judgmental accuracy depends on valid information being available and it depends on the perceiver having the ability to interpret this information.

In the lens model for personality judgments (see Fig. 8.1), the distal psychological event is the actual personality that may manifest in a variety of behavioral cues; these cues may or may not be properly interpreted by perceivers. Only when perceivers notice the valid information and appropriately perceive it can they make an accurate judgment of personality. A person who has some unusual personality traits might stand out from average people by emitting clearer and more noticeable indicators of behaviour (proximal cues), such as an expressive face, peculiar body movements, unique vocal characteristics, and the like. As such, perceivers are more likely to stand a chance of detecting the clues and attribute them to signs of having an extreme personality.

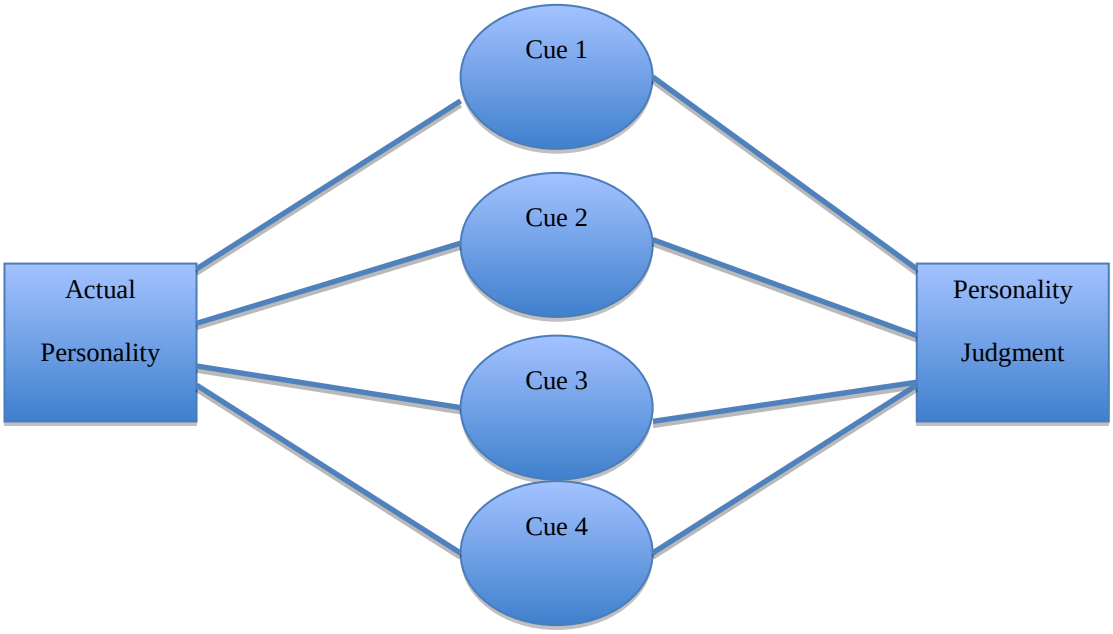


Fig. 8.1. Lens model of personality judgments and behaviour cues.

As a social participant, we have numerous practices to observe personality perception in daily life, and during this time, we might monitor if we have formed an accurate impression of another's personality. Perhaps this experience allows us to develop expertise in identifying those who have an unusual personality. For example, if we perceive someone as being very sociable on first meeting, we would expect the person to act in accordance with such a trait, and we could tell whether our impression is accurate or inaccurate while interacting with the person in future. Through such experiences we might thus refine our ability to identify those behavioral cues which offer information about personality, especially extreme personality.

8.3 How Do People Engage in Guessing Personalities of Others?

8.3.1 Accurate Personality Judgments Based on Behaviour

A series of studies on empathic trait judgments have demonstrated that perceivers made an accurate inference of empathic capacity through information

about behaviour rather than the mere appearance of the target. When observing a still photograph where the target was at the very first moment reading joke, perceivers failed to correctly guess the EQ, either at overall or fine-grained levels. In comparison, when viewing photographs showing the moment as the target delivered the punch line of the joke (such as a smile), perceivers were generally good at inferring the EQ, and performance formed a U-shaped pattern, indicating more sensitivity to targets who were extreme than the targets who were average.

Although Studies 7 and 8 did not directly compare perceivers' performance in viewing animated videos and in viewing still photographs, there was still evidence suggesting that accuracy in judgments of the Big Five traits was based on information about the target's behaviour but not facial appearance. If information about the target's appearance were sufficient for perceivers to draw accurate inferences of the Big Five traits, perceivers would perform equally well after watching a video whether it was with sound or silent. The results reveal that perceivers were much more effective in inferring the five personality dimensions in the Video and Audio Condition than in the Video Only Condition. That is, on the basis of behaviour rather than appearance, it seems perceivers can make an accurate judgment of the Big Five traits.

As discussed in Chapter 6, owing to properties of different personality traits, manifestations of personality connect with behaviour cues varying in quantity (e.g., the duration of a "thin slice") and the quality (the types of behavioral information channels). In general, it seems that some personality traits, such as empathy and extraversion, are relevant to a diversity of activities and events people perform in everyday life; these traits thus become more visible and observable compared with other personality dimensions like neuroticism, which are perhaps more inward and

private such that perceivers do not have easy access to sufficient observable cues for making an accurate judgment (e.g., Funder & Sneed, 1993; Carney et al., 2007).

Accordingly, different types of behavioral information could have different effects on accuracy of personality inferences depending upon the personality trait. Overall, it seems that visual and audio information in combination provides more informative cues than audio only for making accurate intuitive personality judgments. For some traits, such as empathy, visual only information and visual and audio information might provide the same level of informative cues for perceivers to form an accurate first impression; yet, for other traits, such as some dimensions of the Big Five, visual only information might involve behavioral indicators that could be interpreted in ambiguous or even conflicting ways, which leads to poor performance in guessing some of the Big Five traits based on visual only information.

In summary, perceivers can make accurate inferences of personality by observing thin slices of behaviour. The behaviour can be demonstrated in a video with sound, a video without sound, or a soundtrack only; the behaviour can be presented in one or three photographs that sampled different moments while targets were conducting a certain activity. And different types of behaviour information have different impacts on the process of being able to infer empathic traits and the Big Five traits.

8.3.2 Assumed Similarity and Personality Judgments

"Assumed similarity" generally refers to the tendency to view one's own psychological characteristics in others (e.g., Cronbach, 1955; Human & Biesanz, 2012). Both studies reported in Chapter 3 reveal that perceivers assumed similarity while engaging in guessing empathic traits of another person who they were unacquainted with. That is, when asked to make a judgment of a target' EQ after

observing thin slices of behaviour, perceivers tended to project their own EQ onto the target EQ. In Study 1, perceivers' actual EQs (that were fed back before they rated the target's EQ) were significantly associated with their average ratings of the targets' EQs. In Study 2, perceivers did not receive feedback of their actual EQs but were asked to guess their own EQ instead; in this case, perceivers' ratings of the targets' EQs did not correlate with perceivers' actual EQs (measured by their self-reports on the EQ questionnaire) but with the EQ that they guessed about themselves. Even after controlling for their actual EQs, this correlation still survived. Therefore, the results of both studies converge to suggest that on first meeting someone else, those who believe they have low empathy tend to judge others having low empathic capacity while those who believe they have high ability in empathising are more likely to perceive others as having high empathy.

Previous research has reported an assumed similarity effect on personality judgments, which was indicated as the correlation between perceivers' actual personality and perceivers' assessments of another's personality (e.g., Beer & Watson, 2008; Watson, Hubbard, & Wiese, 2000). However, as discussed above, the present data suggest that perceivers' actual personality does not necessarily impact on personality judgments. This inconsistency might be due to the way by which assumed similarity is determined. In previous studies, perceivers' actual personality and the personality they rated about the targets were usually assessed using the same or similar personality questionnaires (e.g., the 20 adjective bipolar scales of the Big Five traits or the first-person and the third-person version of the NEO-FFI questionnaire). In this case, the correlation between perceivers' self-reported personality and their estimation of the personality of others might reflect perceivers' consistent patterns in response to the items in the questionnaires rather than

indicating that perceivers projected their own personality onto other people. In asking perceivers to directly guess the results of the targets' EQ test instead of evaluating the targets' EQs by completing the EQ questionnaire, the present procedure has the advantage of eliminating this potential risk of inflating the effect of assumed similarity, and thus is more useful in examining assumed similarity in personality judgments on first meeting.

Why would perceivers assume similarity when judging other people? According to the explanation of self-based heuristic, perceivers use the information about the self to "fill in the gap" in the absence of sufficient information about another person when making an intuitive personality judgment (Ready et al., 2000; Human & Biesanz, 2012). Evidence indicates that less visible traits (such as N) are more likely to correlate with stronger assumed similarity while more visible traits (such as E) are perceived with less assumed similarity (Watson et al., 2000). In Studies 1 and 2, perceivers had to form a first impression of the target's empathic traits after watching a video only spanning from several to thirty seconds; because this information is so sparse perceivers might have drawn upon knowledge of themselves by default in making a judgment on how empathising another person is.

8.3.3 Expressivity and Empathic Trait Judgments

Studies 5 and 6 suggest that there is a significant correlation between estimates of targets' expressivity and ratings of targets' EQs. Those who were perceived as having high expressivity were also seen as having high empathic capacity while those who were judged as having low expressivity were regarded as being low in the trait of empathy.

As discussed in Chapter 5, the trait of empathy is more or less expressive, which could manifest in a diversity of behaviour, such as facial expressions, vocal

cues, prosocial actions, and so on. Even so, owing to its behavioral manifestations, expressivity might be more visible than the trait of empathy (the EQ). In a mundane scenario (e.g., reading a standard text) that involves little other-oriented emotions, people might not leak out obvious empathy-oriented reactions to others' feelings; by contrast, the extent to which a person is expressive could be more or less revealed in some mannerisms: For example, we might be likely to perceive a person as being expressive if the person is talkative and demonstrative; conversely, if a person is usually quiet and reserved, we might see the person as having low expressivity. While observing the target reading a joke, perceivers might have a better chance to observe behavioral cues in relation to expressivity than in detecting empathising responses to other-oriented emotions. Hence, by linking the visible (expressivity) with the less visible (the EQ), perceivers engage in guessing another's empathic traits on the basis of evaluating how expressive the person would be, though this does not necessarily mean that judgments of expressivity could predict accuracy in guesses of empathic traits.

8.4 Confidence and Personality Judgments

As shown in Study 2, despite finding that perceivers generally believed that they had correctly guessed the EQs of the targets, those perceivers who were more confident of their ability for guessing EQ were not actually any better at guessing EQ, and perceivers who had less confidence in their EQ judgments did not perform worse in identifying who had high and low EQ. Moreover, perceivers seemed to be rather overconfident in their ability to infer empathic traits of another person: Even though they were unacquainted with targets and observed them only in a short video, they were inclined to be confident of being able to identify who had low or high EQ.

Does this mean that people consider it to be easy to form an accurate first impression of personality based on thin slices of behaviour? According to the survey presented in Chapter 7, ordinary people are skeptical about forming an accurate first impression of personality. Then why would any people be confident that their judgment of EQ was accurate in Study 2?

In accordance with the self-verification theory, people actively seek to behave in ways that confirm what they believe to be true about themselves (Swan & Read, 1981). In this case, even if it was difficult to guess the EQ based on a video lasting merely several seconds, once the judgment was made, perceivers might want to be believed they are capable of guessing the EQ of another person and duly express this belief by rating high confidence of their EQ judgments.

Previous research on personality judgments (e.g., Carlson et al., 2010; Ames, et al., 2010) has also suggested high levels of confidence in ratings of first impressions of the Big Five traits, though the evidence of a relationship between judgmental accuracy of personality and judgmental confidence has been inconsistent (see Chapter 3 for detailed discussion). In addition, the survey suggests that people usually believe that they have formed an accurate first impression of others but they also hold negative views about whether they can form an accurate first impression. Taken together, these data raise a question for the future research: Would the phenomenon of being overconfident only happen in first-impression formation of personality or is it more generally the case that once a judgment is made, whether it is about physical events or psychological phenomena, people tend to be confident?

8.5 People's Commonsense Views About First Impressions of Personality

It is very common to form an intuitive impression of another's personality on first meeting in everyday social life. With this experience, people naturally develop many commonsense views about first impressions of personalities, some of which, as reported in Chapter 7, are accurate to some extent, whereas some are inaccurate.

During social interaction, some commonsense views are more likely to be verified in comparison with others. For example, if a person has a distinguishing characteristic, we would be more likely to remember the person and thus might have more chances to check whether our first impression of the person is accurate. Consequently, our first impressions of those who have unusual personalities might be subject to verification through our social experiences. Thus, when asked whether we are more accurate in judging another person with an extreme personality than a person with an average personality, we are probably qualified to give a well-informed response. In contrast, when asked, for example, whether people are generally able to form an accurate first impression of others, we might lack sufficient experience to give a well-informed response and default to a negative position, which happens to be wrong.

8.6 Future Research

8.6.1 Person-Situation Debate and Personality Judgments

The person-situation debate concerns whether aspects of the person or the situation are more powerful in shaping behaviour (e.g., Kenrck & Funder, 1988; Malle, 2011). Given that the effects of many situations that people encounter vary depending on the traits of the person, consensus has developed in suggesting the

interplay of persons and situations in the determination of actions (e.g., Shinner, 2009; Funder, 2006). However, there remains a lack of experimental data directly examining the interaction of person-situation in naturalistic circumstances; instead, there seems to be over-use of hypothetical persons and situations described in sentences. To address this matter, Funder (2006) argues that behaviour of a sample of subjects (targets) must be directly measured in more than one situation.

In Study 2 (see Chapter 3), we used a large corpus of behaviour stimuli, in which the same group of 47 targets were filmed while experiencing three different scenarios (The Conversation, the Joke, & the Screen Test). Perceivers were asked to make judgments of the EQs of the targets after watching videos of either the Conversation, the Joke, or the Screen Test Scenario. The results indicated that perceivers performed equally well in making accurate judgments of the targets' EQs across the scenarios. Likewise, utilizing the same stimuli, Study 7 (see Chapter 6) also found no main effect of scenario when perceivers had to make inferences of the Big Five traits after viewing a video in one of the three scenarios. Hence, the findings from Study 2 and Study 7 converge in suggesting that overall there is cross-situational consistency in the accuracy of first impressions that perceivers formed of the targets' personality traits. In other words, targets have fairly stable traits that allow perceivers to make the same accurate judgments across the scenarios encountered by the targets.

On the other hand, Study 2 demonstrated that perceivers were better at identifying targets who had high empathic capacity in the Conversation and the Joke Scenario than in the Screen Test Scenario. In a similar vein, Study 7 shows that the U-shaped pattern for accuracy in judgments of the Big Five traits was affected by the scenarios – There was an interaction among Trait, Scale and Scenario, and the

interaction between Trait and Scale was also influenced by the scenario (e.g., the U-shaped trend of the trait E in the Screen Test Scenario was quite different from the other scenarios). These findings imply that situational factors could also play a part in determining the targets' behaviour; as a result, the impressions that perceivers formed on the targets might be affected by different scenarios.

In brief, Studies 2 and 7 provided preliminary evidence for the interplay of persons and situations of the long-standing person-situation debate by investigating people's ability to infer personalities from observable behaviour samples in more than one situation. Moreover, the lines of research offered an operational approach allowing examination of the person-situation controversy with high ecological validity. That is, as with the procedure in Studies 2 and 7, perceivers make personality judgments of the same sample of targets who are observed in several real situations – If personality played a key role in deciding behaviour, then the targets would be consistently perceived as who they are across situations; if situations were sufficiently powerful to determine behaviour, then there would be a main effect associated with situation; if both personal and situational factors contributed to the determination of behaviour, there would be an interaction between personality judgments and situations.

8.6.2 Further Research on the U-shaped Trend of Personality Judgments

The present research has revealed the U-shaped trend in personality judgments, where perceivers are more effective in recognizing targets who have one or more extreme traits compared with the targets who have average personalities. As with signal detection, in which a certain signal (such as a light or a sound) is more likely to be detected accurately and quickly when the signal is stronger, can we argue that people who have unusual personalities present a stronger "signal" than those with

average personality? If so, perhaps there are other ways of measuring this stronger signal and in particular perceivers might be able to detect unusual traits more quickly as well as more accurately. Based on the methodology developed in this thesis, future research can shed light on this question by examining response time along with judgmental accuracy.

Using eye-tracking methodology, previous studies on mentalising have found that people flexibly deploy different visual strategies for making retrodictive mentalising inferences about events happening in the world (Pillai et al., 2012) and inferring the gift another person received (Cassidy et al., in press). For instance, while watching a person in a video who was experiencing one of four given scenarios, perceivers who spent more time looking at the eye region of the target were less successful at detecting the scenarios of the Compliment, the Story and the Waiting but not the Joke; in contrast, looking at the mouth region did not relate to accuracy in identifying any of the four scenarios (Pillai et al., 2012). Would variations in gaze patterns correlate with accuracy in identifying targets with different intensities of the same trait? That is, would the way in which people scrutinize those who have extreme personality be different from the way in which they scrutinize the average person? Furthermore, would eye movement strategies vary depending on the particular trait that the perceiver is asked to judge? Integrating eye-tracking methodology into the procedure of trait inferences will provide an answer to these questions.

Finally, in order to investigate the neural bases of empathic processing, Zaki et al. (2009) adopt a social neuroscientific methodology, whereby perceivers' brain activity was observed using fMRI while they watched videos of targets and made judgments on how they were feeling. Results demonstrated that an accurate

understanding of another's emotions depends on structures within the human mirror neuron system thought to be involved in shared sensorimotor representations and regions implicated in mental state attribution (the superior temporal sulcus and medial prefrontal cortex). What is the neural basis of making trait inferences? Moreover, when observing a target with extreme traits, is this represented as more intense activity in relevant substrate or does a different substrate become active compared with when we observe an average person? An approach combining the procedure developed in this thesis with neuroscienfic methods (e.g. fMRI & EEG) will help to answer these questions.

8.6.3 A Broad Framework of Research on Mentalising

As summarized in Chapter 1, research on mentalising began with an effort to chart when and how young children acquire a theory of mind using the false-belief paradigm. More recent research has embraced a wider age range of participants, including infants, older children and adults, and has developed ecologically appropriate tasks in researching people's abilities to infer mental states based on a brief sample of behaviour.

By contrast, person perception in the area of social psychology has traditionally studied adults' understanding of other people in regard to three categories (Brunner & Tagiuri, 1954): the recognition of emotions in others, the accuracy of personality judgments, and the processes by which personality impressions are formed. Research on interpersonal perception accordingly focuses on the question of how well people can perceive each other and the processes by which people come to perceive each other.

Despite adopting different theoretical frameworks and different paradigms, research on mentalising and person perception, in a broad sense, deals with the same

issues of inter-personal understanding. Both areas of research concentrate on the issues of how well people understand other minds. Furthermore, as discussed in Chapter 1, if the main purpose of mentalising is to interpret (retrodict) behaviour and to predict behaviour, then surely it is essential to consider not only mental states but personality traits as well given their roles in explanation and prediction of behaviour.

Why are perceivers better at detecting targets who have extreme traits than targets who have average traits? Is it adaptive to be especially good at identifying people with an unusual personality and if so, why? As depicted in Fig. 8.2, when predicting how a target will behave or when retrodicting what happened to the target, we need to take account of the target's state (situation) and the target's trait (personality). In some cases, when situational norms are strong, it might not be necessary to give much attention to traits. For example, there is a strong norm to the effect that we stop at a red traffic light: In predicting that a target will stop we only need to consider the situation and there is not normally any need to consider the target's personality. However, if we thought that the target had an unusual personality, we might predict that he or she would not stop at a red light, even though most other people do so. Hence, in the particular case, if we are to make a prediction that varies from the default (people stop at red lights), it will probably be because we think the target is peculiar. In short, being aware of extreme traits is probably adaptive in making accurate predictions of behaviour; being aware of average traits might not help our prediction to be any more accurate than if we only knew about the situation. Accordingly, it might thus be adaptive to be particularly attuned to extreme traits when making predictions of behaviour.

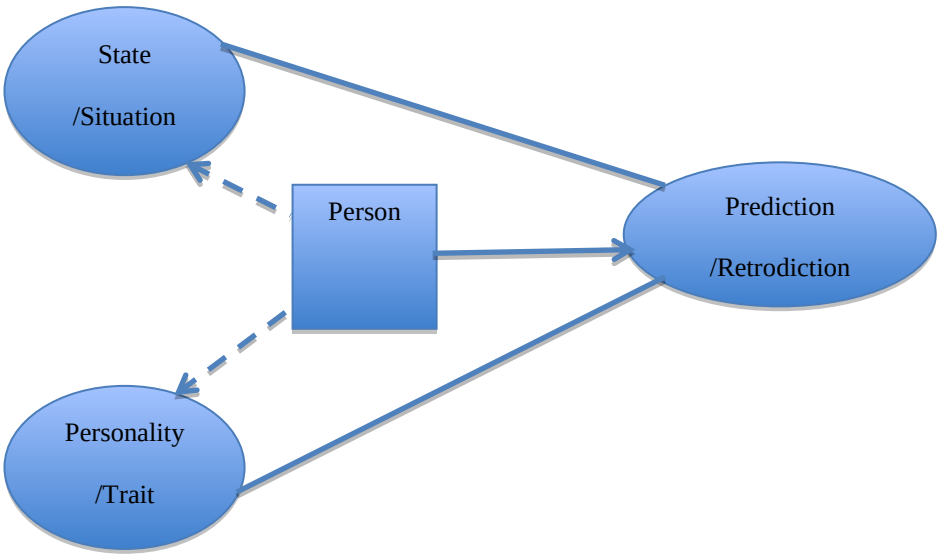


Fig. 8.2. A broad framework of mentalising. The dotted lines represent relationships that cannot be directly observed while the solid lines represent relationships that can be directly observed; the arrows represent the direction of the relationship.

A similar argument applies to retrodiction. If a target displays a strong reaction, this can probably be explained by saying that he/she experienced something striking; but if the target were peculiarly expressive, it might be that the target experienced something rather mundane but he/she is nevertheless the kind of person who reacts strongly. Hence, it will help to interpret the reaction if we can detect that the target is unusual but by default we would probably assume that the strong reaction was caused by a striking event. This example thus illustrates that it might be adaptive to be attuned to extreme traits when making retrodictions.

8.7 Overall Conclusion

In conclusion, a series of studies presented in this thesis offer the following results. (1) Overall, perceivers can form an accurate first impression of an empathic trait and some dimensions of the Big Five traits from a brief sample of behaviour. (2) The fine-grained abilities to infer personality form a U-shaped pattern, in which

perceivers are effective in guessing who have low or high empathic capacity and who is extreme in one or more of the Big Five dimensions. (3) Perceivers make systemically accurate personality judgments based on the targets' behaviour rather than appearance, and the behaviour can be presented either in the form of audio video, silent video, audio only, or photograph. Different types of behavioral information impact on judgmental accuracy depending upon different types of personality traits. (4) There is no difference in the accuracy in guessing empathic traits and the Big Five traits across three scenarios experienced by each target. (5) Perceivers use the strategy of assumed similarity in making an initial judgment of EQ, by which perceivers who believe they are high in EQ tend to guess the target also has high EQ while perceivers who judge themselves to have low EQ are more likely to rate low EQ in the target. (6) Perceivers are generally confident that they have correctly guessed empathic traits, either about themselves or the targets, though the confidence in self-perception is higher than in other-perception; notwithstanding, there is no correlation between judgmental confidence and judgmental accuracy. (7) Perceivers' ratings of the targets' expressivity significantly correlated with their assessments of the targets' EQ.

In addition, the survey reveals that people usually underestimate the ability to form an accurate first impression of others based on various behavioral information. With this exception, most of their commonsense views about first-impression personality judgments are consistent with the aforementioned empirical findings.

References

Ames, D. R., Kammrath, L. K., Suppes, A., & Bolger, N (2010). Dissociation between accuracy and confidence in thin-slice impressions. *Personality and Social Psychology Bulletin*, 36, 264–277.

Allport, G. W. (1937). *Personality: A psychological interpretation.* New York: Holt, Rinehart & Winston.

Albright, L., Kenny, D. A., & Malloy, T. E. (1988). Consensus in personality judgments at zero acquaintance. *Journal of Personality and Social Psychology*, *55*, 387–395.

Ambady, N. (2010). The perils of pondering: Intuition and thin slice judgments. *Psychological Inquiry*, *21 (4)*, 271–278.

Ambady, N., LaPlante, D., & Johnson, E. (2001). Thin slice judgments as a measure of interpersonal sensitivity. In J. A. Hall & F. J. Bernieri (Eds.), *Interpersonal Sensitivity: Theory and Measurement* (pp. 89–101), Mahwah, NJ: Erlbaum.

Ambady, N., Bernieri, F., & Richeson, J. (2000). Towards a histology of social behavior: Judgmental accuracy from thin slices of behavior. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (pp. 201–272). New York, NY: Academic Press.

Ambady, N., & Rosenthal, R. (1993). Half a minute: Predicting teacher evaluations from thin slices of behavior and physical attractiveness. *Journal of Personality and Social Psychology*, *64*, 431–441.

Ambady, N., & Rosenthal, R. (1992). Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis. *Psychological Bulletin*, *111*, 256–274.

Apperly, I. A. (2013). Can theory of mind grow up? Mindreading in adults, and its

implications for the neuroscience and development of mindreading**.** In Baron-Cohen, S., Tager-Flusberg, H. & Lombardo, M. (Eds.) *Understanding Other Minds* (3rd Ed.).

Apperly, I. A., Warren, F., Andrews, B. J., Grant, J. & Todd, S. (2011). Error patterns in the belief-desire reasoning of 3- to 5-year-olds recur in reaction times from 6 years to adulthood: evidence for developmental continuity in theory of mind. *Child Development, 82(5),* 1691–1703.

Apperly, I. A., Carroll, D. J., Samson, D., Qureshi, A., Humphreys, G. W., & Moffatt, G. (2010). Why are there limits on theory of mind use? Evidence from adults' ability to follow instructions from an ignorant speaker. *Quarterly Journal of Experimental Psychology*, *63(6),* 1201–1217.

Apperly, I. A., Back, E., Samson, D., & France, L. (2008). The cost of thinking about false beliefs: Evidence from adults' performance on a non-inferential theory of mind task. *Cognition*, *106*, 1093–1108.

Apperly, I. A. (2008). Beyond Simulation-Theory and Theory-Theory: Why social cognitive neuroscience should use its own concepts to study "theory of mind". *Cognition*, *107(1)*, 266–283.

Apperly, I. A., Riggs, K. J., Simpson, A., Chiavarino, C., & Samson, D. (2006). Is belief reasoning automatic? *Psychological Science*, *17(10*), 841–844.

Apperly, I. A., Samson, D., & Humphreys, G. W. (2005). Domain-specificity and theory of mind: evaluating evidence from neuropsychology. *Trends in Cognitive Sciences*, *9(12)*, 572–577.

Asch, S. E. (1946). Forming impressions of personality. *Journal of Abnormal and Social Psychology*, *41*, 1230–1240.

Austin, E. J. (2004). An investigation of the relationship between trait emotional intelligence and emotional task performance. *Personality and Individual Differences*, *36*, 1855–1864.

Back, E., & Apperly, I. A. (2010). Two sources of evidence on the non-automaticity of true and false belief ascription. *Cognition*, *115(1)*, 54–70.

Back, M. D., Stopfer, J. M., Vazire, S., Gaddis, S., Schmukle, S. C., Egloff, B., & Gosling, S. D. (2010). Facebook profiles reflect actual personality not self-idealization. *Psychological Science*, *21*, 372-374.

Baillargeon, R., Scott, R. M., & He, Z. (2010). False-belief understanding in infants. *Trends in Cognitive Sciences*, *14(3)*, 110–118.

Baron-Cohen, S., (2012). *Zero Degrees of Empathy: A New Theory of Human Cruelty and Kindness*. London: Penguin Books.

Baron-Cohen, S., & Wheelwright, S. (2004). The empathy quotient: An investigation of adults with Asperger Syndrome or High Functioning Autism, and normal sex differences. *Journal of Autism and Developmental Disorders*, *34*, 163–175.

Baron-Cohen, S., Jolliffe, T., Mortimore, C., & Robertson, M. (1997). Another advanced test of theory of mind: evidence from very high functioning adults with autism or Asperger Syndrome. *Journal of Child Psychology and Psychiatry*, *38*, 813–822.

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind?" *Cognition*, *21*, 37– 46.

Batson, C.D., Lishner, D., Carpenter, A., Dulin, L., Harjusola-Wevv, S., Stocks, E.

L., Gale, S., Hassan, O et al. (2003). "…As you would have them do unto you": Does imagining yourself in the other's place stimulate moral action?. *Personality and Social Psychology Bulletin*. *29,* 1190–1201.

Batson, C. D. (1991). *The altruism question: Toward a social psychological answer.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Batson, C. D., O'Quin, K., Fultz, J., Vanderplas, M., & Isen, A. M. (1983). Influence of self-reported distress and empathy on egoistic versus altruistic motivation to help. *Journal of Personality and Social Psychology, 45,* 706−718.

Baumeister, R. F., Vohs, K. D., & Funder, D. C. (2007). Psychology as the science of self-reports and finger movements: Whatever happened to actual behavior? *Perspectives on Psychological Science, 2,* 396–403.

Beer, A., & Watson, D. (2008). Personality judgment at zero acquaintance. Agreement, assumed similarity, and implicit simplicity. *Journal of Personality Assessment, 90,* 250–260.

Bernieri, F. J., Zuckerman, M., Koestner, R., & Rosenthal, R. (1994). Measuring person perception accuracy: Another look at self-other agreement. *Personality and Social Psychology Bulletin, 20,* 367–378.

Besel, L.D.S., Yuille, J. C. (2010). Individual differences in empathy: The role of facial expression recognition. *Personality and Individual Differences, 49,* 107–112.

Birch, S. A. J., & Bloom, P. (2007). The curse of knowledge in reasoning about false beliefs. *Psychological Science, 18(5),* 382–386.

Birch, S. A. J., & Bloom, P. (2004). Understanding children's and adults' limitations in mental state reasoning. *Trends in Cognitive Sciences, 8,* 255–260.

Bloom, P., & German, T. P. (2000). Two reasons to abandon the false-belief task as a

test of theory of mind. *Cognition, 77*, B25–B31.

Borkenau, P., Mauer, N., Riemann, R., Spinath, F., Angleitner, A. (2004). Thin slices of behavior as cues of personality and intelligence. *Journal of Personality and Social Psychology, 86*, 599–614.

Borkenau, P., & Ostendorf, F. (1998). The Big Five as states: How useful is the Five-Factor Model to describe intraindividual variations over time?. *Journal of research in personality, 32*, 202–221.

Borkenau, P., & Liebler, A. (1993). Convergence of stranger ratings of personality and intelligence with self-ratings, partner ratings, and measured intelligence. *Journal of Personality and Social Psychology, 65*, 546–553.

Borkenau, P., & Liebler, A. (1992a). The cross-modal consistency of personality: Inferring strangers' traits from visual or acoustic information. *Journal of Research in Personality, 26*, 183–204.

Borkenau, P., & Liebler, A. (1992b). Trait inferences: Sources of validity at zero acquaintance. *Journal of Personality and Social Psychology, 62*, 645–657.

Boven, L. V., & Loewenstein, G. (2003). Social projection of transient drive states. *Personality and Social Psychology* Bulletin, *29*, 1159–1168. DOI: 10.1177/0146167203254597.

Bruner, J. S., & Tagjuri, R. (1954). The perception of people. In G. Lindzey (Ed.), *The handbook of social psychology* (1st ed., Vol. 2, pp. 634–654). Reading, MA: Addison-Wesley.

Brunswik, E. (1955). Representative design and probabilistic theory in functional Psychology. *Psychological Review, 62*, 193–217.

Buttelmann, D., Carpenter, M., & Tomasello, M. (2009). 18-month-old infants show false-belief understanding in an active helping paradigm. *Cognition, 112,*

337–342.

Carlson, S. M., & Moses, L. J. (2001). Individual differences in inhibitory control and children's theory of mind. *Child Development, 72*, 1032–1053.

Carlson, S. M, Moses, L. J., & Hix, H. R. (1998). The role of inhibitory processes in young children's difficulties with deception and false belief. *Child Development, 69*, 672–691.

Carlson, E. N., Furr, R. M., & Vazire, S. (2010). Do we know the first impressions we make? Evidence for idiographic meta-accuracy and calibration of first impressions. *Social Psychological and Personality Science, 1*, 94–98.

Carney, D. R., Colvin, C. R., & Hall, J. A. (2007). A thin slice perspective on the accuracy of first impressions. *Journal of Research in Personality, 41*, 1054–1072.

Cassidy, S., Mitchell, P., Aquah, D., Ropar, D., & Chapman, P. (in press). Can adults with autism successfully infer what happened to someone from their reactions?

Chandler, M. J., Fritz, A. *S.*, & Hala, S. (1989). Small-scale deceit: Deception as a marker of two-, three-, and four-year-olds' theories of mind. *Child Development, 60*, 1263–1277.

Costa, P. T., & McCrae, R. R. (2011). *NEO™ Inventories: NEO™ Personality Inventory-3 (NEO™-PI-3) Manual*. Psychological Assessment Resources, Inc.

Costa, P. T., Jr., & McCrae, R. R. (1992). *NEO PI-R and NEO-FFI professional manual.* Odessa, FL: Psychological Assessment Resources.

Cronbach, L. J. (1955). Processes affecting scores on "understanding of others" and "assumed similarity". *Psychological Bulletin, 52*, 177–193.

Davis, M. H. (1980). A multidimensional approach to individual differences in empathy. *JSAS Catalog of Selected Documents in Psychology, 10*, 85.

Decety, J., & Jackson, P. L. (2004). The functional architecture of human empathy. *Behavioral and Cognitive Neuroscience Reviews, 3*, 71−100.

Dennett, D. C. (1978). "Beliefs about Beliefs" (commentary on Premack, et al.). *Behavioral and Brain Sciences, 1*, 568–570.

Doyle, A. C. (1893). The Musgrave Ritual. In The Memoirs of Sherlock Holmes.

Eisenberg, N., & Fabes, R. A. (1990). Empathy: Conceptualization, measurement, and relation to prosocial behavior. *Motivation and Emotion, 14*, 131–149.

Estes, S. G. (1938). Judging personality from expressive behavior. *Journal of Abnormal and Social Psychology, 33*, 217–236.

Ferguson, F., & Austin, E. J. (2010). Associations of trait and ability emotional intelligence with performance on Theory of Mind tasks in an adult sample. *Personality and Individual Differences, 49*, 414–418.

Flavell, J. H. (2004). Theory-of-Mind development: retrospect and prospect. *Merrill-Palmer Quarterly, 50(3)*, 274–290.

Flavell, J. H. (2000). Development of children's knowledge about the mental world. *International Journal of Behavioral Development, 24*, 15–23.

Flavell, J. H., Miller, P. H., & Miller, S. A. (1993). Cognitive development (3rd ed.). Englewood Cliffs, NJ: Prentice-Hall.

Fleeson, W. (2001). Toward a structure- and process-integrated view of personality: traits as density-distributions of states. *Journal of personality and social psychology, 80*, 1011–1027.

Fletcher, P. C., Happé, F., Frith, U., Baker, S. C., Dolan, R. J., Frackowiak, R. S. J., et al. (1995). Other minds in the brain: A functional imaging study of "theory

of mind" in story comprehension. *Cognition*, *57*, 109–128.

Fletcher, G. J. O. (1984). Psychology and common sense. *American Psychologist*. *39*, 203–213.

Flury, J. M., Ickes, W., Schweinle, W. (2008). The borderline empathy effect: Do high BPD individuals have greater empathic ability? Or are they just more difficult to "read"?. *Journal of Research in Personality*, *42*, 312–332.

Funder, D. C. (2012). Accurate personality judgment. *Current Directions in Psychological Science*, *21*, 177–182.

Funder, D. C., & Fast, L. A. (2010). Personality in social psychology. In D. Gilbert & S. Fiske (Eds.), Handbook of social psychology (5th ed., pp. 668-697.) New York: Wiley.

Funder, D. C. (2007). *The personality puzzle* (4th ed.). New York: Norton.

Funder, D. C. (2006). Towards a resolution of the personality triad: Persons, situations, and behaviors. *Journal of Research in Personality*, *40(1)*, 21–34.

Funder, D. C. (2001). Personality. *Annual Review of Psychology, 52*, 197–221.

Funder, D. C. (1995). On the accuracy of personality judgment: A realistic approach. *Psychological Review*, *102*, 652–670.

Funder, D. C., Kolar, D. C., & Blackman, M. C. (1995). Agreement among judges of personality: Interpersonal relations, similarity, and acquaintanceship. *Journal of Personality and Social Psychology*, *69*, 656–672.

Funder, D. C., & Sneed, C. D. (1993). Behavioral manifestations of personality: An ecological approach to judgmental accuracy. *Journal of Personality and Social Psychology*, *64*, 479–490.

Funder, D. C. (1991). Global traits: A Neo-Apportian approach to personality. *Psychological Science*, *2(1)*, 31–39.

Funder, D. C., & Colvin, C. R. (1988). Friends and strangers: Acquaintanceship, agreement, and the accuracy of personality judgment. *Journal of Personality and Social Psychology, 55*, 149–158.

Funder, D. C. (1987). Errors and mistakes: Evaluating the accuracy of social judgment. *Psychological Bulletin*, *101*, 75–91.

Gage, N. L., & Cronbach, L. J. (1955). Conceptual and methodological problems in interpersonal perception. *Psychological Review*, *62*, 411–422.

Gallese V, & Goldman, A. (1998) Mirror neurons and the simulation theory of mind-reading. *Cognition, 2(12)*, 493–501.

German, T. P., & Hehman, J. A. (2006). Representational and executive selection resources in "theory of mind": Evidence from compromised belief-desire reasoning in old age. *Cognition*, *101*, 129–152.

Gery, I., Miljkovitch, R., Berthoz, S., & Soussignan, R. (2009). Empathy and recognition of facial expressions of emotion in sex offenders, non-sex offenders and normal controls. *Psychiatry Research*, *165*, 252–262.

Goldman, A. I. (1989) Interpretation psychologized. *Mind and Language, 4*, 161–185.

Gopnik, A., & Astington, J. W. (1988). Children's understanding of representational change and its relation to the understanding of false belief and the appearance-reality distinction. *Child Development*, *59*, 26–37.

Gordon, R. (1986) Folk psychology as simulation. *Mind and Language, 1*, 158–170.

Gosling, S. D., Ko, S. J., Mannarelli, T., & Morris, M. E. (2002). A room with a cue: Judgments of personality based on offices and bedrooms. *Journal of Personality and Social Psychology*, *82*, 379–398.

Hall, J. A., Andrzejewski, S. A., Murphy, N. A., Mast, M. C., & Feinstein, B. A. (2008). Accuracy of judging others' traits and states: Comparing mean levels across tests. *Journal of Research in Personality*, *42*, 1476–1489.

Hall, J. A., & Mast, M. S. (2007). Sources of accuracy in the empathic accurate paradigm. *Emotion*, *7(2)*, 438–446.

Heider, F. (1958). The psychology of interpersonal relations. New York: Wiley.

Heller, D., Watson, D., & Ilies, R. (2004). The role of person versus situation in life satisfaction: A critical examination. *Psychological Bulletin, 130*, 574–600.

Heyman, G. D., & Gelman, S. A. (1998). Young children use motive information to make trait inferences. *Developmental Psychology*, *34*, 310–321.

Hoffman, M. 2000. *Empathy and Moral Development*, Cambridge: Cambridge University Press.

Hoffman, M. L (1984). Interaction of affect and cognition in empathy. In C. E. Izard, J. Kagan, & R. B. Zajone (Eds.), *Emotions, cognitions, and behavior* (pp. 103–131). Cambridge, UK: Cambridge University Press.

Holleran, S. E., Mehl, M. R., & Levitt, S. (2009). Eavesdropping on social life: The accuracy of stranger ratings of daily behavior from thin slices of natural conversations. *Journal of Research in Personality, 43*, 660–672.

Houston, J. P. (1985). Psychology: Untutored lay knowledge of the principles of psychology: Do we know anything they don't?. *Psychological Reports*, *57*, 567–570.

Houston, J. P. (1983). Psychology: A closed system of self-evident information? *Psychological Reports*, *52*, 203–208.

Human, L. J., & Biesanz, J. C. (2012). Accuracy and assumed similarity in first impressions of personality: Differing associations at different levels of analysis. *Journal of Research in Personality, 46,* 106–110.

Ickes, W. (2003). *Everyday mind reading: Understanding what other people think and feel*, Amherst, NY: Prometheus Books.

Ickes, W., Buysse, A., Pham, H., Rivers, K., Erickson, J. R., Hancock, M., Kelleher, J., & Gesn, P. R. (2000). On the difficulty of distinguishing "good" and "poor" perceivers: A social relations analysis of empathic accuracy data. *Personal Relationships, 7*, 219–234.

Ickes, W. (1993). Empathic accuracy. *Journal of Personality, 61*, 587–610.

Ickes, W., Stinson, L., Bissonnette, V., & Garcia, S. (1990). Naturalistic social cognition: Empathic accuracy in mixed-sex dyads. *Journal of Personality and Social Psychology*, 59, 730–742.

Jones, E. E, Davis, K. E. (1965). From acts to dispositions: The attribution process in person perception. *Advances in Experimental Social Psychology, 2*, 219–266.

Jussim, L. (1991). Social perception and social reality: A reflection-construction model. *Psychological Review, 98*, 54–73.

Kelley, H. H. (1992). Common-sense psychology and scientific psychology. *Annual Review of Psychology, 43*, 1–23.

Kenny, D. A., Albright, L., Malloy, T. E., & Kashy, D. A. (1994). Consensus ininterpersonal perception: Acquaintance and the big five. *Psychological Bulletin, 116,* 245–258.

Kenny, D. A., Horner, C., Kashy, D. A., & Chu, L. (1992). Consensus at zero acquaintance: Replication, behavioral cues, and stability. *Journal of Personality and Social Psychology, 62*, 88–97.

Kenny, D. A. (1991). A general model of consensus and accuracy in interpersonal perception. *Psychological Review*, *98*, 155–163.

Kenny, D. A., & Albright, L. (1987). Accuracy in interpersonal perception: A social relations analysis. *Psychological Bulletin*, *102*, 390–402.

Kenrick, D.T., & Funder, D.C. (1988). Profiting from controversy: Lessons from the person-situation debate. *American Psychologist*, *43*, 23–34.

Keysar, B., Barr, D. J., Balin, J. A., & Brauner, J. S. (2000). Taking perspective in conversation: The role of mutual knowledge in comprehension. *Psychological Sciences*, *11*, 32–38.

Keysar, B., Lin, S. H., & Barr, D. J. (2003). Limits on theory of mind use in adults. *Cognition*, *89(1)*, 25–41.

Kilianski, S. E. (2008). What do you think I think I am? Accuracy in perceptions of others' self-esteem. *Journal of Research in Personality*, *42*, 386–398.

Kinderman, P., Dunbar, R., & Bentall, R. P. (1998). Theory-of-mind deficits and causal attributions. *British Journal of Psychology*, *89*, 191–204.

Kressel , L. , & Uleman , J. S . (2010). Personality traits function as causal concepts. *Journal of Experimental Social Psychology*, *46*, 213–216.

Lawrence, E. J., Shaw, P., Baker, D., Baron-Cohen, S., & David, A. S. (2004). Measuring empathy: reliability and validity of the empathy quotient. *Psychological Medicine*, *34*, 911–924.

Leslie, A. M., & Thaiss, L. (1992). Domain specificity in conceptual development neuropsychological evidence from autism. *Cognition*, *43*, 225–251.

Lewis, D. (1972). Psychological and theoretical identifications. *Australasian Journal of Philosophy*, *50*, 249–258.

Malle, B. F. (2011). Time to give up the dogmas of attribution: An alternative theory of behavior explanation. In J. M. Olson and M. P. Zanna, *Advances of Experimental Social Psychology* (Vol. 44, pp. 297–352). Burlington: Academic Press.

Malle, B. F. (2001). Attribution processes. In N. J. Smelser and P. B. Baltes (Eds.), *International encyclopedia of the social and behavioral sciences* (Vol. 14, Developmental, social, personality, and motivational psychology; section editor N. Eisenberg, pp. 913–917). Amsterdam: Pergamon/Elsevier.

Martin, R. A., Berry, G. E., Dobranski, T., & van Horne, M. (1996). Emotion perception threshold: Individual differences in emotional sensitivity. *Journal of Research in Personality, 38,* 290–305.

McArthur, L. Z., & Baron, R. M. (1983). Toward an ecological theory of social perception. *Psychological Review, 90,* 215–238.

McCarthy, R. J., & Skowronski, J. J. (2011). What will Phil do next? Spontaneously inferred traits influence predictions of behavior. *Journal of Experimental Social Psychology, 47,* 321–332.

McCrae, R. R., Terracciano, A., et al. (2005). Personality profiles of cultures: Aggregate personality traits. *Journal of Personality and Social Psychology, 89(3),* 407–425.

McCrae, R. R., & Costa, P. T. (2004). A contemplated revision of the NEO Five-Factor Inventory. *Personality and Individual Differences, 36 (3),* 587–596.

McCrae, R. R., & John, O. P. (1992). An introduction to the five-factor model and its applications. *Journal of Personality, 60,* 175–215.

McKinnon, M. C., & Moscovitch, M. (2007). Domain-general contributions to social

reasoning: Theory of mind and deontic reasoning re-explored. *Cognition*, *102 (2)*, 179–218.

McLarney-Vesotski, A. R., Bernieri, F., Rempala, D. (2006). Personality perception: A developmental study. *Journal of Research in Personality*, *40*, 652–674.

Mehl, M. R., Gosling, S. D., & Pennebaker, J. W. (2006). Personality in its natural habitat: Manifestations and implicit folk theories of personality in daily life. *Journal of Personality and Social Psychology*, *90*, 862–877.

Milgram, S. (1963). Behavioral study of obedience. *Journal of Abnormal and Social Psychology*, *67*, 371–378.

Mischel , W . (1968). *Personality and assessment*. New York : Wiley.

Nisbett, R. E., & Ross, L. (1980). Alternatives to representation or alternative representations. Comments on the ecological approach. *Contemporary Social Psychology*, *14*, 163–174.

Norman, W. T., & Goldberg, L. R. (1966). Raters, ratees, and randomness in personality structure. *Journal of Personality and Social Psychology, 4*, 681–691.

North, M. S., Todorov, A., & Osherson, D. N. (2012). Accuracy of inferring self- and other-preferences from spontaneous facial expressions. *Journal of Nonverbal Behavior*, DOI 10.1007/s10919-012-0137-6.

North, M. S., Todorov, A., & Osherson, D. N. (2010). Inferring the preferences of others from spontaneous, low-emotional facial expressions. *Journal of Experimental Social Psychology, 46(6),* 1109–1113.

Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science*, *308*, 255–258.

Ozer, D. J., & Benet-Martinez, V. (2006). Personality and the prediction of consequential outcomes. *Annual Review of Psychology*, *57*, 401–421.

Park , B . (1986). A method for studying the development of impressions of real people . *Journal of Personality and Social Psychology*, *51*, 907–917.

Paulhus, D. L., & Bruce, M. N. (1992). The effect of acquaintanceship on the validity of personality impressions: A longitudinal study. *Journal of Personality and Social Psychology*, *63*, 816–824.

Peirce, J. W. (2007) PsychoPy - Psychophysics software in Python. *Journal of Neuroscience Methods*, *162 (1-2)*, 8–13.

Perner, J., & Lang, B. (1999). Development of theory of mind and executive control. *Trends in Cognitive Sciences*, *3*, 337–344.

Peterson, D. M, & Riggs, K. J. (1999). Adaptive modelling and mindreading. *Minds & Language*, *14 (1)*, 80–112.

Petrides, K. V., & Furnham, A. (2003). Trait emotional intelligence. Behavioural validation in two studies of emotion recognition and reactivity to mood induction. *European Journal of Personality*, *17*, 39–57.

Pillai, D., Sheppard, E., Mitchell, P. (2012). Can people guess what happened to others from their reactions? *PLoS ONE, 7(11)*: e49859. doi:10.1371/journal.pone.0049859.

Premack, D. & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *The Behavioral and Brain Sciences*, *4*, 515–526.

Ready, R. E., Clark, L. A., Watson, D., & Westerhouse, K. (2000). Self- and peer-reported personality: Agreement, trait rateability, and the ‘‘self-based heuristic’’. *Journal of Research in Personality*, *34*, 208–224.

Realo, A., Allik, J., Nolvka, A., Valk, R., Ruus, T., Schmidt, M., et al. (2003). Mind reading ability: Beliefs and performance. *Journal of Research in Personality, 37,* 420–445.

Rentfrow, P. J., & Gosling, S. D. (2003). The Do Re Mi's of everyday life: The structure and personality correlates of music preferences. *Journal of Personality and Social Psychology, 84,* 1236–1256.

Rim , S., Uleman , J. S., & Trope, Y . (2009). Spontaneous trait inference and construal level theory: Psychological distance increases non-conscious trait thinking. *Journal of Experimental Social Psychology, 45,* 1088–1097.

Roberts, B. W., Kuncel, N. R., Shiner, R., Caspi, A., & Goldberg, L. R. (2007). The power of personality: The comparative validity of personality traits, socioeconomic status, and cognitive ability for predicting important life outcomes. *Perspectives on Psychological Science, 2,* 313–345.

Rosenberg, J., "Wilfrid Sellars", The Stanford Encyclopedia of Philosophy (Summer 2011 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/sum2011/entries/sellars/>.

Rumble, A. C., Van Lange, P. A. M., & Parks, C. D. (2010). The benefits of empathy: When empathy may sustain cooperation in social dilemmas. *European Journal of Social Psychology, 40,* 856−866.

Rutherford, M. D. (2004). The effect of social role on theory of mind reasoning. *British Journal of Psychology, 95,* 91–103.

Sabbagh, M. A., Moses, L. J., & Shiverick, S. (2006). Executive functioning and preschoolers' understanding of false beliefs, false photographs and false signs. *Child Development, 77,* 1034–1049.

Schmitt, D. P., Allik, J., McCrae, R. R., Benet-Mart ńez, V. (2007). The geographic distribution of Big Five personality traits: Patterns and profiles of human self-description across 56 nations. *Journal of Cross-cultural Psychology*, *38*, 173–212.

Sellars, W. (1956). Empiricism and the philosophy of mind. In H. Feigl & M. Scriven (eds.), *The Foundations of Science and Concepts of Psychology and Psychoanalysis: Minnesota Studies in the Philosophy of Science*, Vol. 1. Minneapolis: University of Minnesota Press. 253–329.

Shiner, R. L. (2009). Persons and situations act together to shape the course of human lives. *Journal of Research in Personality*, *43*, 270–271.

Snodgrass, S. E. (1985). Women's intuition: The effect of subordinate role on interpersonal sensitivity. *Journal of Personality and Social Psychology*, *49*, 146–155.

Song, H., Onishi, K. H., Baillargeon, R., & Fisher, C. (2008). Can an actor's false belief be corrected through an appropriate communication? Psychological reasoning in 18.5-month-old infants. *Cognition*, *109*, 295–315.

Southgate, V., Senju, A., & Csibra, G. (2007). Action anticipation through attribution of false belief by two-year-olds. *Psychological Science*, *18*, 587–592.

Stich, S., & Nichols, S. (2003). Folk psychology. In S. Stich & T. A. Warfield (eds.), The Blackwell Guide to Philosophy of Mind, Oxford: Basil Blackwell, pp. 235–255.

Stone, V. E., Baron-Cohen, S., & Knight, R. T. (1998). Frontal lobe contributions to theory of mind. *Journal of Cognitive Neuroscience*, *10*, 640–656.

Stueber, K., "Empathy", *The Stanford Encyclopedia of Philosophy* (Summer 2013 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/sum2013/entries/empathy/>.

Surian, L., Caldi, S., & Sperber, D. (2007). Attribution of beliefs to 13-month-old infants. *Psychological Science*, *18*, 580–586.

Swann, W. B. (1984). Quest for accuracy in person perception: A matter of pragmatics. *Psychological Review*, *91*, 457–477.

Swann, W. B., & Read, S. J. (1981). Self-verification processes: How we sustain our self-conceptions. *Journal of Experimental Social Psychology*, *17*, 351–372.

Taft, R. (1955). The ability to judge people. *Psychological Bulletin, 52,* 1–23.

Tagiuri, R. (1958). Introduction. In R. Tagiuri & L. Petrullo (Eds.), *Person perception and interpersonal behavior*. Stanford, CA: Stanford Univ. Press.

Titchener, E. B. 1909: *Lectures on the Experimental Psychology of Thought-Processes*. New York: Macmillan.

Thoresen, J. C., Vuong, Q. C., & Atkinson, A. P. (2012). First impressions: Gait cues drive reliable trait judgments. *Cognition*, *124*, 261–271.

Todorov , A. , & Uleman , J. S . (2004). The person reference process in spontaneous trait inferences. *Journal of Personality and Social Psychology, 87,* 482 – 493.

Todorov, A., & Uleman, J. S. (2003). The efficiency of binding spontaneous trait inferences to actors' faces. *Journal of Experimental Social Psychology, 39,* 549–562.

Todorov, A., & Uleman, J. S. (2002). Spontaneous trait inferences are bound to actors' faces: Evidence from a false recognition paradigm. *Journal of Personality and Social Psychology*, *83*, 1051–1065.

Uleman, J. S., & Saribay, S. A. (2012). Initial impressions of others. In K. Deaux & M. Snyder (Eds.), *Oxford handbook of personality and social psychology* (pp. 337-366). New York: Oxford University Press.

Uleman , J. S . (2005). On the inherent ambiguity of traits and other mental concepts . In B. F. Malle & S. D. Hodges (Eds.), *Other minds: How humans bridge the divide between self and others* (pp. 253–266). New York: Guilford.

Uleman, J. S., Newman, L. S., & Moskowitz, G. B. (1996). People as flexible interpreters: Evidence and issues from spontaneous trait inference. In M. P. Zanna (Ed.). *Advances in experimental social psychology* (Vol. 28, pp. 211–279). San Diego, CA: Academic Press.

Uleman, J. S., Hon, A., Roman, R. J., & Moskowitz, G. B. (1996). On-line evidence for spontaneous trait inferences at encoding. *Personality and Social Psychology Bulletin*, *22*, 377–394.

Vazire, S., & Mehl, M. (2008). Knowing me, knowing you: The accuracy and unique predictive validity of self-ratings and other-ratings of daily behavior. *Journal of Personality and Social Psychology*, *95*, 1202–1216.

Watson, D., Hubbard, B., & Wiese, D. (2000). Self-other agreement in personality and affectivity: The role of acquaintanceship, trait visibility, and assumed similarity. *Journal of Personality and Social Psychology*, *78*, 546–558.

Watson, D. (1989). Strangers' ratings of the five robust personality factors: Evidence of a surprising convergence with self-reports. *Journal of Personality and Social Psychology*, *52*, 120–128.

Waxer, P. (1976). Nonverbal cues for depth of depression: Set versus no set. *Journal of Consulting and Clinical Psychology*, *44*, 493.

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory of mind development: The truth about false-belief. *Child Development, 72,* 655 – 584.

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition, 13,* 103–128.

Winter, L., & Uleman, J. S. (1984). When are social judgments made? Evidence for the spontaneousness of trait inferences. *Journal of Personality and Social Psychology, 47,* 237–252.

Zaki, J., Weber, J., Bolger, N., & Ochsner, K. (2009). The neural bases of empathic accuracy. *Proceedings of the National Academy of Sciences of the United States of America, 106,* 11382–11387.

Zaki, J., Bolger, N., & Ochsner, K. (2009). Unpacking the informational bases of empathic accuracy. *Emotion, 9,* 478–487.

Zaki, J., Bolger, N., & Ochsner, K. (2008). It takes two: The interpersonal nature of empathic accuracy. *Psychological Science, 19,* 399–404.

Zhou, Q., Valiente, C., & Eisenberg, N. (2003). "Empathy and Its Measurement." Ed. S.J. Lopez and C.R. Snyder, *Positive Psychological Assessment: A Handbook of Models and Measures,* 269–284. Washington, DC: American Psychological Association.

Zelazo, P. D., Carter, A., Reznik, S., & Frye, D., (1997). Early development of executive function: a problem-solving framework. *Review of General Psychology, 1,* 1-29.