

Explaining the Illusion of Asymmetric Insight

Asbjørn Steglich-Petersen & Mattias Skipper

Penultimate draft. Final version forthcoming in *Review of Philosophy and Psychology*.

Abstract: People tend to think that they know others better than others know them (Pronin et al. 2001). This phenomenon is known as the “illusion of asymmetric insight.” While the illusion has been well documented by a series of recent experiments, less has been done to explain it. In this paper, we argue that extant explanations are inadequate because they either get the explanatory direction wrong or fail to accommodate the experimental results in a sufficiently nuanced way. Instead, we propose a new explanation that does not face these problems. The explanation is based on two other well-documented psychological phenomena: the tendency to accommodate ambiguous evidence in a biased way, and the tendency to overestimate how much better we know ourselves than we know others.

Keywords Illusion of asymmetric insight · Interpersonal knowledge · Biased evidence assimilation · Cognitive bias

1. Introduction

A series of recent psychological experiments suggest that people tend to think that they know others better than others know them (Pronin et al. 2001). If Alice and Ben are as people are most, Alice will tend to think that she knows Ben better than Ben knows her, and Ben will tend to think that he knows Alice better than Alice knows him. This phenomenon is known as the “illusion of asymmetric insight.”

If people really do have a tendency to think that they know others better than others know them, it must indeed reflect an illusion, at least for many of us. Although it may be the case that *some* people have superior interpersonal insight, this clearly cannot be the case for the majority of us. Just like Alice and Ben cannot *both* be right in the example above, many of us must be wrong in taking ourselves to have superior insight into others.

The aim of this paper is to investigate what *explains* the illusion of asymmetric insight. This means searching for a psychological mechanism or pattern that can be held responsible for the illusion. As it is, this explanatory challenge stands unmet. Extant

explanations, we argue, are inadequate because they either get the explanatory direction wrong or fail to accommodate the experimental results in a sufficiently nuanced way. Instead, we will propose a new explanation that does not face these problems.

The explanation consists of two parts. The first part posits a tendency to *overestimate* our own insight into others. This tendency, we argue, is predicted by another well-documented psychological phenomenon, namely the tendency to accommodate ambiguous evidence in a biased manner. Since third-personal evidence about other people's inner feelings and motives is often highly ambiguous, we will tend to take such evidence to confirm our prior beliefs about the personal characteristics of other people, thereby making us prone to overestimate our own interpersonal insight. The second part of our explanation posits a tendency to *underestimate* other people's insight into us. This tendency, we argue, is predicted by yet another well-documented psychological phenomenon, namely the tendency to overestimate how much better we know ourselves than we know others. This might seem surprising: how could the relatively innocuous idea that we know ourselves better than we know others have the effect of making us prone to underestimate how well other people know us? We show that, when coupled with a few other well-supported auxiliary hypotheses, it can in fact have that effect. Together these two tendencies—the tendency to overestimate our own interpersonal insight, and the tendency to underestimate that of others—provide a simple and well-founded explanation of the illusion of asymmetric insight that can predict important details in the experimental findings.

Why is it important to understand the psychological underpinnings of the illusion of asymmetric insight? One set of reasons has to do with social interaction. If we think that we know others better than others know us, it is natural to suspect that we will be inclined to invest more effort in correcting other people's perception of *us* than in correcting our own perception of *others*. We will tend to talk when we should listen, and will become impatient or defensive when others suggest that we have misinterpreted their statements or motives. This might lead to socially undesirable outcomes, such as entrenched disagreement and conflict (Pronin et al. 2001, pp. 652-53; Kennedy and Pronin 2008). But

if we understand the cognitive mechanisms that are responsible for these impulses, we might be in a better position to correct or alleviate them.

A second set of reasons concerns how the illusion of asymmetric insight relates to various other psychological phenomena. If our explanation is correct, understanding the illusion of asymmetric insight may be of even greater importance than previously thought. On the prevalent explanation, the illusion of asymmetric insight is seen as an effect of another illusory conviction about the relative importance of introspection for knowing ourselves versus other people: we tend to think that knowing *us* requires access to inner thoughts and feelings, whereas we can know *others* quite well merely on the basis of observable behavior (Pronin et al. 2004 and 2008; Pronin 2009). This so-called “introspection illusion” is relevant for understanding a wide range of psychological phenomena, including people’s tendency to make dispositional inferences about others, but not about themselves (Jones and Nisbett 1972), and their tendency to see others as being more susceptible to cognitive and motivational biases than they themselves are (Pronin et al. 2002 and 2004). But if we are right, the prevalent explanation gets the direction of explanation wrong: it is the illusion of asymmetric insight that gives rise to the introspection illusion, rather than *vice versa*. As such, our explanation of the illusion of asymmetric insight stands to explain a much wider set of psychological phenomena related to the introspection illusion as well.

We do not present new empirical findings in this article. Rather, we base our explanation on its ability to account for the available data, and on the significant amount of independent support that has been given for the auxiliary hypotheses that we invoke. However, we also highlight some interesting novel predictions of our explanation to be tested in future experimental work.

The agenda for the rest of paper is as follows. After reviewing the relevant experimental findings by Pronin et al. (§2), we argue that existing explanations of the illusion of asymmetric insight are inadequate (§3). We then present our new explanation of the illusion (§4), argue that it does not suffer from the same shortcomings as the existing ones (§5), and show that it can explain a two further psychological effect unearthed by

Pronin et al. (§6). Finally, we describe a hitherto untested third-personal version of the illusion of asymmetric insight, which may serve to test our explanation in further studies (§7).

2. The Illusion of Asymmetric Insight

Pronin et al. (2001) report a group of five studies designed to test the hypothesis that people tend to think that they know others better than others know them. Below we briefly review the setup and results of each study.

First study: Participants were asked to think of a close friend and rate (on a scale from 1-11) how well they felt they knew or understood their friend, and how well they felt their friend knew or understood them. The study revealed a tendency among the participants to rate their knowledge of their friend higher than their friend's knowledge of them.

Second study: Pairs of same-sex college roommates were asked a number of questions about how well they felt they knew their roommate, and how well they felt their roommate knew them. For example, the participants were asked to rate (on a scale from 1-11) how well they felt they knew how shy their roommate was, and how well they felt their roommate knew how shy they were. In line with the first study, the participants tended to rate their knowledge of their roommate higher than their roommate's knowledge of them. The study also sought to determine whether the effect would be more pronounced for judgments involving knowledge about private and unobservable psychological qualities rather than public and observable ones, and more pronounced for judgments involving relatively negative qualities rather than positive or neutral ones. For both contrasts, a significantly increased effect was found. The private and unobservable qualities included in this part were: "real" feelings, motives for action, doing things for the purpose of fitting in, and doing things just for the purpose of pleasing others. The study also gave support to an *intrapersonal* version of the illusion of asymmetric insight, whereby people tend to think that they know themselves better than others know themselves. Again, this tendency was found to be more pronounced for judgments involving knowledge about private qualities, and negative qualities.

Third study: Participants were given a questionnaire in which they were asked to describe when they were “most like themselves”, and when their close friend was “most like himself or herself.” The participants tended to give more weight to privately accessible thoughts and feelings when describing themselves, while giving more weight to publicly available information about performance and behavior when describing their friends. Pronin et al. take this to reflect a tendency to think that it is harder to gain insight into oneself than it is to gain insight into others.

Fourth study: Pairs of previously unacquainted college students engaged in brief face-to-face interaction, and were subsequently asked how much they felt they had learned about their partner, and how much they felt their partner had learned about them. In line with the third study, the reports revealed a tendency among the students to think that they had learned more about their partner during the conversation than their partner had learned about them.

Fifth study: Participants took a projective test in which they were asked to complete a series of incomplete word fragments, such as “g _ _ l” and “c _ r _ _ r.” Afterwards, the participants were asked how revealing they felt the projective test had been of them, and how revealing they felt it had been of the other participants. In line with the third and fourth study, the participants generally found their own word fragment completions to be less revealing than those of other participants.

3. Existing Explanations

Although the primary concern of Pronin et al. is to establish *that* people suffer from an illusion of asymmetric insight, they also consider three possible explanations of *why* people might suffer from this illusion. Two of the explanations are based on the so-called “self-enhancement bias” and the phenomenon of “naïve realism.” However, Pronin et al. reject these explanations because that they are unable to account for certain details in the experimental findings. Instead, they propose a third explanation, which is based on an illusory conviction about how *knowable* we are compared to others. Below we review

Pronin et al.'s discussion of all three explanations, and explain why we find their favored explanation wanting.

3.1 Explanations from Self-Enhancement Bias and Naïve Realism

One line of explanation discussed and ultimately rejected by Pronin et al. (2001) sees the illusion of asymmetric insight as a special instance of the so-called “self-enhancement bias.” Numerous studies have shown that the majority of people take themselves to be above-average on a wide range of desirable qualities and abilities, such as intelligence, driving skills, and the like.¹ Perhaps the claim to superior interpersonal insight should be understood as just another instance of this more general tendency.

However, Pronin et al. (2001, p. 652) rightly observe that if we simply subsume the illusion of asymmetric insight under the self-enhancement bias, we risk overlooking important details in the experimental findings. For example, as noted above, the tendency to claim superiority in interpersonal knowledge was more pronounced for judgments involving knowledge about private and unobservable psychological qualities rather than public and observable ones, and for judgments involving negative qualities rather than positive or neutral ones (Pronin et al. 2001, p. 652). While this finding is *consistent* with the view that the illusion of asymmetric insight is an instance of the self-enhancement bias, it is not *predicted* by it. Thus, even if the illusion of asymmetric insight may be seen as an instance of the self-enhancement bias, this would still leave important nuances in the experimental findings unexplained. Another problem pointed out by Pronin et al. (2001, p. 652) is that it is unclear from the experimental findings whether people primarily *overestimate* their own interpersonal insight, *underestimate* other people's interpersonal insight, or *both*. Thus, given that the self-enhancement bias only covers tendencies to overestimate oneself, the illusion of asymmetric insight might not be a clear-cut instance of the self-enhancement bias.

Pronin et al. also consider the possibility of understanding the illusion of asymmetric insight as a manifestation of *naïve realism*: the tendency to assume that one's perception of

¹ See Schlenker and Miller (1977), Greenwald (1980), Riess et al. (1981), Greenberg et al. (1982), and Mezulis et al. (2004) for studies and reviews of the self-enhancement bias.

the world accurately reflects what the world is really like.² A notable consequence of naïve realism is that when people make conflicting judgments on some matter, each party will tend to think that the other party's judgment is biased or otherwise flawed. One might naturally take this phenomenon to underlie the illusion of asymmetric insight: when people make conflicting assessments of the personal characteristics of themselves or others, each party will tend to think that the other party is wrong, and hence deem themselves more successful at judging the personal characteristics of the other party than *vice versa*.

However, this explanation is limited in much the same way as the explanation from the self-enhancement bias. Even if the illusion of asymmetric insight may be appropriately categorized as a manifestation of naïve realism, the explanatory value of such a categorization will be limited in the absence of more detailed predictions of the experimental results, such as the increased effect for judgments involving knowledge about private or negative qualities.

3.2 Explanations from Knowability and Introspection Illusions

Consider next the explanation favoured by Proning et al. Several of the experiments reviewed in §2 indicate that people tend to think that they are less knowable or “harder to access” than others. For example, the participants in the first study tended to deem their own “true self” less “visible” or more “hidden beneath the surface” than that of their close friend. Similarly, the participants in the third study provided self-descriptions that gave more weight to inner unobservable events and feelings than did their descriptions of their close friend.³ Pronin et al. (2001, p. 652) take these findings to suggest that the reason why people take themselves to have superior interpersonal insight is that they take themselves to be less knowable than others.

While we agree that the experimental findings indicate that people deem themselves less knowable than others, we have doubts about the significance of this “knowability illusion” in the context of explaining the illusion of asymmetric insight. The trouble is that

² See Ichheiser (1949), Ross and Ward (1996), and Gilovich and Ross (2015) for studies and reviews of naïve realism.

³ See also Goffman (1959), Markus (1983), and Andersen and Ross (1984) for studies and discussions of people's convictions about their own and other people's knowability.

we risk getting the *explanatory direction* wrong. It would seem more natural to explain the knowability illusion in terms of the illusion of asymmetric insight than *vice versa*. Here is why: if people really do suffer from an illusory conviction to the effect that they themselves are less knowable than others, this illusion is itself in need of an explanation. One plausible explanation is that people deem themselves to be more successful at gaining insight into others than *vice versa*, and, on this basis, infer that they are generally less knowable than others. For example, people might feel misunderstood more often than they take themselves to misunderstand others, and hence eventually come to think of themselves as being harder to access than others. Thus, even if people's own explanation of their perceived superior interpersonal insight is based on a belief to the effect that they themselves are less knowable than others, the causal psychological explanation plausibly goes the other way around: the tendency to think of oneself as having superior interpersonal insight is what moves people to think that they are less knowable than other people.

In subsequent work, Pronin and colleagues have refined the explanation from knowability by relating it to another well-documented illusion known as the "introspection illusion:" the tendency to think that observable behavior is more revealing of others than of oneself, and that access to private thoughts and feelings is more critical when it is oneself, rather than someone else, who is being interpreted (Pronin et al. 2004 and 2008; Pronin and Kugler 2007; Pronin 2009). It has been suggested that the introspection illusion may lead people to think that they are in a better position to gain insight into others than *vice versa*, thus giving rise to an illusion of asymmetric insight. For example, Pronin et al. (2004) submit that the illusion of asymmetric insight is rooted in the conviction that "knowing *us* demands that one enjoy access to our private thoughts, feelings, motives, intentions, and so forth," whereas "we can know *others* quite well solely from paying attention to their behaviors, gestures, verbal responses, and other observable manifestations" (Pronin et al. 2004, p. 794). A similar point is made by Pronin et al. (2008) who note that the illusion of asymmetric insight "has been traced to people's tendency to view their own spontaneous or off-the-cuff responses to others' questions as relatively

unrevealing even though they view others' similar responses as meaningful" (Pronin et al. 2008, p. 797).⁴

However, by using the introspection illusion to explain the illusion of asymmetric insight, we once again run a risk of getting the explanatory direction wrong. If people indeed suffer from an introspection illusion, this is itself in need of an explanation. A plausible explanation is that people deem themselves to be more successful at gaining insight into others than *vice versa*, and hence eventually come to conclude that the third-personal evidence that they have about others must be comparatively more revealing than the third-personal evidence that other people have about them. In other words, the reason why people are moved to think that third-personal evidence is less revealing of them than of others is that they take themselves to be more successful at gaining insight into others than *vice versa*.

If this is the correct way of understanding the direction of explanation, any explanation of the illusion of asymmetric insight that does not rely on the knowability and introspection illusions will thereby help explain these latter two illusions as well. Thus, since the explanation of the illusion of asymmetric insight that we propose in the following does not rely on the knowability and introspection illusions, it promises to provide a unified account of all three illusions.

Another potential weakness of the explanations from the knowability and introspection illusions is that it is difficult to see how these explanations could be extended to the *intrapersonal* version of the illusion of asymmetric insight. Pronin et al. discuss the interpersonal and intrapersonal versions of the illusion of asymmetric insight under the same heading, and they seem to assume that they are equally well explained by the knowability and introspection illusions. However, it is anything but clear that they *are*. If we think of ourselves as being less knowable than others, why should this lead us to think that we have superior intrapersonal insight? It seems odd to reason that since I am harder

⁴ On another possible reading of Pronin et al.'s discussion of the introspection illusions, they do not seek to explain the illusion of asymmetric insight in terms of the introspection illusion, but rather seek to explain both of these illusions in terms of the knowability illusion. However, even if this is the correct interpretation, our new explanation still provides a unified account of all three illusions, by providing an independent explanation of the illusion of asymmetric insight, which in turn explains the other two.

to know than others, I must know myself better than others know themselves. Likewise, it seems odd to reason that since introspection is more important for knowing me than for knowing others, I must know myself better than others know themselves. While our main focus will be on the interpersonal version of the illusion of asymmetric insight, we will also show, in §5, how our explanation provides resources for understanding the intrapersonal version.

4. A New Explanation

We now move on to present a new explanation of the illusion of asymmetric insight. As previously noted, it is unclear from the experimental findings whether people primarily tend to *overestimate* their own interpersonal insight, *underestimate* other people's interpersonal insight, or *both*. But regardless of the relative significance of these tendencies, it is clear that neither tendency on its own is enough to generate an illusion of asymmetric insight. If I overestimate my own interpersonal insight, I won't suffer from an illusion of asymmetric insight if I overestimate other people's interpersonal insight to the same degree. And if I underestimate other people's interpersonal insight, I won't suffer from an illusion of asymmetric insight if I underestimate my own interpersonal insight to the same degree. Thus, at a minimum, an explanation of the illusion of asymmetric insight must explain a *differential* assessment of how well people take themselves to know others as compared to how well others know them. We will meet this requirement by showing how a few well-founded psychological phenomena jointly predict a double tendency to *overestimate* our own interpersonal insight and *underestimate* the interpersonal insight of others.

4.1 *Overestimating One's Own Interpersonal Insight*

What might explain a tendency to overestimate one's own ability to know others? We submit that this tendency is predicted by the familiar, well-supported tendency to interpret ambiguous evidence in a biased manner:

Biased Assimilation of Evidence (BAE): People tend to interpret ambiguous evidence as supportive of their prior opinions.

This tendency has been documented by a number of experimental findings across a wide range of evidential domains. It is usually understood as an aspect of the more general confirmation bias, which is the tendency to search for, interpret, and remember evidence and information in ways that confirm one's prior beliefs, and impede the possibility of falsifying them. While early support for this general tendency came from Wason (1960), the classic study documenting the specific bias in assimilating ambiguous evidence is Lord et al. (1979). In their study, a group of participants, who were initially either for or against capital punishment, were presented with two sets of data on the effectiveness of capital punishment as a deterrent; one set of data speaking in favor of such an effect, and another speaking against it. The studies revealed a tendency among the participants to have a more favorable opinion of the data that supported their initial opinion, while being skeptical of the data speaking against it.

As Kunda (1990) notes, ambiguous evidence is particularly susceptible to biased assimilation, since reality itself to some extent constrains what one is able to conclude from a given body of evidence. One's evidence cannot speak too compellingly against one's prior opinion if it is to allow for a biased interpretation. Since mixed or ambiguous bodies of evidence do not do so, biased assimilation of such evidence becomes more likely. More generally, we might say that the degree of susceptibility for biased assimilation of evidence depends positively on the degree of ambiguity in the evidence, other things being equal.⁵

We take it as relatively obvious that third-personal psychological evidence is often quite ambiguous and open to interpretation, since it often leaves us to draw relatively subtle abductive inferences about the inner feelings and motives of other people. Given this background assumption, it is not hard to see how BAE predicts a tendency to overestimate one's own reliability in judging the personal characteristics of others. The ambiguity of third-personal psychological evidence makes us prone to interpret the evidence in a way that supports our initial judgments of the personal characteristics of others. As a result, more false judgments are likely to be made and retained than we are

⁵ For useful overviews, see Gilovich (1991, ch. 3) and Oswald and Grosjean (2004).

aware of, thereby making us prone to overestimate our own ability to judge the personal characteristics of other people.

To illustrate, suppose that Alice has come to regard Ben as being shy (say, based on some relatively limited experience with him in social situations). Even if he is not in fact shy, third-personal evidence of shyness is ambiguous, which means that Alice is likely to interpret evidence gained in subsequent interaction with Ben in a way that supports her initial judgment. As such, she stands to retain her initial judgment, even if a more balanced assessment of the evidence would speak against it. Being unaware of this mistake, she takes herself to know Ben better than she in fact does.

Importantly, this explanation can account for the details in the experimental findings that the explanations from the self-enhancement bias and naïve realism failed to predict, namely that our tendency to think that we have superior interpersonal knowledge is particularly pronounced in cases where the knowledge concerns private and unobservable qualities, rather than public and observable ones (Pronin et al. 2001, p. 643). Our explanation predicts this effect, since evidence about private and unobservable qualities is likely to be more ambiguous than evidence about public and observable qualities, thus leaving more room for biased assimilation, which in turn will strengthen the overestimation of one's own ability to know others.

4.2 *Underestimating the Interpersonal Insight of Others*

Given that people overestimate their *own* interpersonal knowledge, one might naturally expect them to overestimate *other* people's interpersonal knowledge as well, since this would mean that they ascribed to others the same general capacities for interpersonal knowledge that they find in themselves. However, if this were the case, we would not see an illusion of asymmetric insight. What is needed is thus an explanation of why people do *not* overestimate other people's interpersonal insight, despite overestimating their own.

At the root of our proposal lies a claim about people's comparative assessments of people's capacity for *self-knowledge* versus knowledge of *others*:

Overestimation of Self-Other Asymmetry (OSA): People tend to overestimate the degree to which people's self-knowledge is superior to their knowledge of others.

This tendency differs in two important respects from the knowability and introspection illusions. First, OSA states that people overestimate how much better *people in general* are at knowing themselves compared to knowing others. By contrast, the knowability and introspection illusions are beliefs that people harbour about *themselves in particular*: that I am less knowable than others, and that introspection is comparatively more important for knowing *me*. Second, OSA claims that we overestimate how much better people know *themselves* than *others*, whereas the knowability and introspection illusions consist of the belief that we are in a better position to know *others* than others are to know *us*. While this latter belief seems rather immodest (after all, why should we believe ourselves to be special?), the overestimation described by OSA is much more subtle. There is nothing obviously remarkable in thinking that we are all better placed to know ourselves than we are to know others, and it is easy to see how one could come to overestimate the extent of this difference (more on this below). It is thus also a conviction that reasonable and reflective people could easily hold. Note finally the relative weakness of OSA. The thesis does not require, for example, that people believe themselves to be *perfectly* reliable in knowing themselves. It suffices that they regard themselves as *better* at knowing themselves than at knowing others, as long as they overestimate this difference in reliability.

What reason is there to think that OSA is true? That we know ourselves better than we know others, at least in important respects, seems like a core commitment of folk-psychology. Peter Carruthers (2008) goes as far as to maintain that a belief in what he calls the 'self-transparency' of key aspects of our minds is species-universal and innate. Carruthers relies mainly on theoretical considerations, including that an innate belief in self-transparency would make good sense from an evolutionary perspective, but he also notes some further explanatory virtues of this innateness hypothesis. For example, it would explain the 'near-ubiquity' of belief in self-transparency in Western philosophy prior to the writings of Sigmund Freud, and the surprised reactions of ordinary people to scientific

results and theorizing about the mind which undermine the belief in self-transparency. Indeed, when researchers in cognitive science present results indicating aspects of our mental lives to which the subject does *not* enjoy privileged access, these results are often presented as surprising, thus revealing that the researchers take for granted that people tend to believe in self-transparency.

Carruthers (2008) notes a lack of experimental work on the degree to which people actually presume introspective access to their own minds. However, a later study by Kozuch and Nichols (2011) sought to remedy this. They distinguished two dimensions of the putative folk belief in introspective access, namely the *power* of the access, i.e. how reliable people believe their introspective access to be, and the *domain* of access, i.e. what types of mental events people believe they are able to introspect. Although they didn't find that people presume universal introspective access, they found that people do assume a high degree of introspective access, especially in certain domains (we return to these domains in a moment).

Given that people think that they have a kind of access to their own minds that they do not have to the minds of others, what reason is there to suppose that they *overestimate* the superiority of this access? Here, we can distinguish two relevant sets of considerations, one to do with the *nature* of the access, the other to do with the *reliability* of the access.

In a classic study, Alison Gopnik (1993) argued that while we as adults believe that our access to our own psychological states is different in nature to our access to the psychological states of others, developmental evidence suggests that it is not. We tend to believe that first-person knowledge is gained through direct perception, whereas third-person knowledge must be inferred from behavior. But developmental evidence suggests, she maintains, that both kinds of knowledge operate according to the same interpretive principles. In another discussion, Carruthers (2010) provides further reason to doubt the existence of a special form of introspective access to our own psychology, at least if introspection is understood as a kind of cognitive pathway that is distinct from the way in which we come to know psychological states of others. Carruthers advances a number of arguments for this claim, including a lack of anatomical evidence, and the unexplained

evolutionary extravagance that would be involved in developing a separate pathway for understanding ourselves. These claims of Gopnik and Carruthers are of course compatible with people having an accurate estimation of their relative abilities for knowing themselves and others. After all, simply by experiencing one's own behavior on a greater scale, subjects plausibly have more and better psychological evidence about themselves than they have about others. But a commitment to a false theory of self-knowledge as a matter of direct perception or introspection does seem to go hand in hand with an exaggerated sense of *how much* better we are at knowing ourselves over knowing others, which is what our thesis claims.

Moving on to the second set of considerations, this conclusion is in line with the very significant body of research over the last several decades, demonstrating significant limitations and exceptions to our reliability in knowing ourselves. Numerous studies suggest that there are certain classes of attitudes, tendencies, processes, and personality traits that we are systematically bad at detecting in ourselves. Following Wilson and Dunn (2003), we can distinguish two broad categories of limitations to our self-knowledge, which they label "motivational" and "non-motivational." Motivational limitations extend to e.g. uncomfortable memories, repressed "Freudian" beliefs, biases, undesirable personality traits, and similar characteristics that people tend to have motivational reasons for assessing in a skewed manner.⁶ Non-motivational limitations apply especially to much of the unconscious mental processing underlying reasoning, perception, motor learning, attitude formation, decision-making, and more.⁷ As such, there might in fact be certain types of psychological states and characteristics that are more reliably detectable by third-person evidence than by first-person evidence. In this connection, it is worth remarking that in the aforementioned study by Kozuch and Nichols (2011), people were particularly prone to overestimating the amount of self-access they had when it came to the motivational basis for their own decisions and behavior.

With this motivation of OSA in place, we can return to our main question. How does OSA lead us to underestimate other people's ability to gain insight into us? The basic

⁶ See Vazire (2010) for further discussion of motivational limitations to self-knowledge.

⁷ See, e.g., Nisbett and Wilson (1977) for a classic study.

mechanism that we propose is this: sometimes we become aware of what personal or psychological traits other people ascribe to us. These ascriptions may sometimes be explicitly communicated, as when someone tells me that she thinks I am shy or prone to anger. At other times, the ascriptions may be inferred from more subtle cues, as when I become aware that someone seems overly careful not to make me angry, or appears to imply that I would be uncomfortable in some particular social situation. We will not make any substantive assumptions about the exact prevalence of these cues, but we take it that we are at least sometimes in a position to make informed comparisons of our own self-assessments with other people's assessments of us. Naturally, these assessments sometimes come into conflict. I might infer from your calming tone that you regard me as anxious, while not regarding myself as anxious. Or I might take your bantering to indicate that you regard me as a resilient person, while regarding myself as rather fragile. We will again not make any substantive assumptions about the exact prevalence of such conflicts. But a number of psychological studies indicate that people quite commonly experience a feeling of being misunderstood by others.⁸

In cases of conflicting assessments, OSA predicts that we will tend to regard our own self-assessment as the one most likely to be correct on the grounds that, in general, people's self-knowledge tends to be superior to people's knowledge of others. However, since we tend to *overestimate* how much better we know ourselves than we know others, there will be more cases of conflicting assessments in which the other person is right about us than we think. Over time, we will thus come to regard other people as being less reliable in their judgments of our own personal characteristics than they really are. In other words, we will come to underestimate their interpersonal insight.

As with the mechanism proposed in §4.1, this mechanism also predicts the details found in Study 2, that the illusion of asymmetric insight was particularly pronounced in cases where participants felt that knowledge of them would require access to private thoughts and feelings, and in cases where the knowledge concerned negative or undesirable

⁸ See e.g. Lun et al. (2008) and Oishi et al. (2013) for recent studies and literature reviews on people's perception of being misunderstood by others, and the importance of this for their well-being and feeling of being socially connected.

qualities. As previously mentioned, we are particularly prone to overestimating our own self-knowledge when it comes to these exact qualities. In fact, when finding an increased effect with respect to qualities that require access to private thoughts and feelings, Pronin et al. focused in particular on grounds and motivations for decisions and behavior (e.g.: ‘motives for action, doing things just for the purpose of fitting in, and doing things just for the purpose of pleasing others’ (2001: 643)). And it is with respect to these very qualities that Kozuch and Nichols (2011) found that people were most prone to overestimate their self-knowledge. For these qualities we are therefore likely to be particularly dismissive of other people’s conflicting assessments of us.

Putting this effect together with the one described in §4.1, we have an explanation both of the tendency to overestimate our own interpersonal insight and of the tendency to underestimate that of others. The first effect is grounded in the general tendency to accommodate evidence in a biased way, which becomes strengthened by the ambiguous nature of third-personal psychological evidence. The second effect is grounded in the exaggerated conviction that people’s self-knowledge is superior to their knowledge of others, which leads us to dismiss other people’s assessments of ourselves more often than we should.

It is worth pointing out that our belief that we are in a better position to know ourselves than we are to know others need not be false or exaggerated for our explanation of the illusion of asymmetric insight to hold, although the effect will be weaker if the belief is accurate. If the belief is accurate, we will not underestimate the ability of others to know us, but rather come to a more realistic assessment of their ability. When we become aware of other people’s assessments of us, and they conflict with our own self-assessments, we will still dismiss their assessments because of our belief in the superiority of self-knowledge over knowledge of others. But if our self-knowledge really is as superior as we take it to be, our dismissals of other people’s assessments will be largely correct, and our assessment of their ability to know others will therefore also be more accurate. However, because of our tendency to overestimate our own ability to know others, we will still regard ourselves as better at knowing others, than others are at knowing us, thus retaining weaker version of

the illusion of asymmetric insight. That being said, we think that the research summarized above does give us strong reason to think that we have an exaggerated sense of our ability to know ourselves, as described by OSA.

5. Interactions of BAE and OSA

We have now seen how BAE and OSA can work in tandem to generate an illusion of asymmetric insight. However, the two can also interact with each other in a more direct manner. In fact, BAE and OSA can sometimes *counteract* each other. To illustrate, suppose that Alice judges Ben to be shy. How will Alice react if she learns that Ben does not consider himself to be shy? If we were to focus exclusively on OSA, we should expect Alice to trust Ben's judgment because she thinks that people in general know themselves better than others know them. But if we were to focus exclusively on BAE, we should expect Alice to place greater weight on her own judgment because she tends to assimilate third-personal evidence in a biased manner. Thus, in cases like these, OSA and BEA pull in opposite directions. What should we make of this?

First and most importantly, the fact that BAE and OSA can sometimes have opposite-directed effects does not undermine the explanation proposed in the previous section. Most obviously, the second part of the explanation remains intact because BAE does not generate any countereffect in those cases where OSA leads us to place too little weight on other people's judgments of us. If anything, we should expect BAE to *bolster* the effect of OSA in cases where your judgment of me conflicts with my own judgment of myself. Only a little less obviously, the first part of the explanation also remains intact because the occasional countereffect of OSA does not change the fact that BAE in general makes us prone to interpret third-personal evidence in a way that supports our prior opinions. The effect of BAE may well be *mitigated* to some extent by the interaction with OSA. But we should expect the overall pattern to be the same, namely that we are likely to make and retain more false judgments about other people's personal characteristics than we are aware of, thereby leading us to overestimate our own interpersonal insight. Hence, the

interaction between BAE and OSA does not undermine our explanation of the illusion of asymmetric insight.

On the contrary, we can use the interaction between BAE and OSA to explain two further psychological effects documented by Pronin et al. The first effect is the previously mentioned *intrapersonal* illusion of asymmetric insight: the tendency to think that we know ourselves better than other people know themselves. This tendency was demonstrated by the second of the studies reviewed in §2, but was found to be overall less pronounced than the interpersonal illusion. The interaction between BAE and OSA predicts exactly that. To see why, suppose again that Alice and Ben make conflicting judgments about whether Ben is shy. The effect of OSA will be to make Alice inclined to trust Ben's judgment. However, BAE will have the countereffect of sometimes leading Alice to think that her own judgment of Ben is more accurate than Ben's judgment of himself. As a result, Alice will tend to underestimate, at least to some extent, how well Ben knows himself. By contrast, if Alice and Ben make conflicting judgments about whether *Alice* is shy, BAE will obviously not have the effect of leading Alice to underestimate how well she knows herself. On the present picture, this is why we tend to think that we know ourselves better than others know themselves.

The second effect consists of a tendency to think that we know ourselves *much* better than others know us, whereas other people know themselves only *slightly* better than we know them. This tendency was also demonstrated by the second of the studies reviewed in §2, and was found to be highly sensitive to the level of external observability of the assessed trait. For example, Alice would take her knowledge of her own "shyness" (a low-observability trait) to be much better than Ben's knowledge of her shyness, whereas she would only take Ben's knowledge of his shyness to be slightly better than her knowledge of his shyness. By contrast, Alice would take her knowledge of her own "messiness" (a high-observability trait) to be only slightly better than Ben's knowledge of her messiness, whereas she would take Ben's knowledge of his messiness to no better (in fact, slightly *worse*) than her knowledge of his messiness.⁹ These effects are exactly what we

⁹ These effects are illustrated in Figure 1 (Pronin et al. 2001: 644).

should expect from the interaction between BAE and OSA. In general, OSA makes us inclined to think that people know themselves better than others know them. However, BAE will have the countereffect of sometimes leading us to think that we know others better than they know themselves, especially for traits that are high in external observability. By contrast, BAE does not have the effect of sometimes leading us to think that others know us better than we know ourselves. As a result, we will tend to underestimate how well others know themselves, without underestimating how well we know ourselves. On the present picture, that is why we tend to think that we know ourselves much better than others know us, whereas other people know themselves only slightly better than we know them.¹⁰

In sum, BAE and SOA not only jointly predict the interpersonal illusion of asymmetric insight, but allow us to explain in detail several related phenomena as well.

6. Comparisons with Existing Explanations

In this section, we compare our explanation from §4 with the ones discussed in §3. We begin by revisiting the shortcomings that led us to reject the existing explanations. In relation to the explanations from the illusions of knowability and introspection, we raised the worry that they got the explanatory direction wrong. Even if people really do tend to deem themselves less knowable than others, we argued that it is more natural to explain this conviction in terms of people's tendency to deem themselves superior at knowing others, rather than letting the explanation go the other way around. Our explanation does not raise a similar worry. Although our explanation is partly based on a general belief about the comparative epistemic positions of oneself and others, the content of this belief is that subjects in general enjoy a superior access to their own mental state, which clearly cannot be explained in terms of a tendency to regard oneself as superior at knowing others.

Furthermore, it is worth noting that believing oneself to enjoy superior access to one's own mental state is *prima facie* much more reasonable, and thus less in need of explanation, than believing oneself to be less knowable than others. Indeed, if Carruthers

¹⁰ We are grateful to an anonymous reviewer for bringing this effect to our attention.

(2008) is right, there is reason to think that an exaggerated belief in the superiority of self-knowledge is innate, or at least a very basic commitment of folk-psychology.

Finally, since our explanation of the illusion of asymmetric insight does not rely on the illusions of knowability and introspection, we plausibly have an explanation of the latter two illusions as well, as argued in §3. This promises to provide us with a highly unified picture of a wide range of otherwise disparate psychological phenomena. Obviously, we have not offered anything like a complete picture of how these phenomena hang together. In particular, it may require further experimental work to establish whether the knowability and introspection illusions can indeed be explained in terms of the illusion of asymmetric insight, as we have suggested, or whether they should rather be seen as arising directly from BAE and OSA, and hence as being “on a par” with the illusion of asymmetric insight at the level of explanation. But we take the foregoing considerations to suggest that BAE and OSA at least provide us with the basic ingredients for a unified understanding of the knowability and introspection illusions and the illusion of asymmetric insight.

In relation to the explanations from the self-enhancement bias and naïve realism, Pronin et al. pointed out that they are too general to account for specific details in the experimental findings. In particular, they failed to explain the heightened tendency to claim superior interpersonal knowledge about negative qualities and qualities that require access to private thoughts and feelings. By contrast, we have seen that our explanation predicts exactly this effect.

A further potential weakness of the explanation from self-enhancement bias was that it is unclear from the experimental findings whether the illusion of asymmetric insight is primarily a result of a tendency to *overestimate* one’s own interpersonal insight, *underestimate* other people’s interpersonal insight, or *both*. Thus, it remains unclear whether the illusion of asymmetric insight can indeed be categorized as an instance of the self-enhancement bias. By contrast, our account offers explanatory resources regardless of how this question is settled, since it relies both an explanation of our tendency to overestimate ourselves and of our tendency to underestimate others.

Finally, neither of the explanations reviewed in §3 are well-placed to account for the effects discussed in §5. Most obviously, the explanations from the self-enhancement bias and naïve realism are too general to make any detailed predictions about why the effects discussed in §5 should be sensitive to the level of external observability of the assessed trait. But the explanations from the illusions of knowability and introspection also fall short. As argued in §3.2, it is unclear how the knowability and introspection illusions would generate a tendency to think that we know ourselves better than other people know themselves. The same goes for the tendency to think that we know ourselves *much* better than other people know us, whereas other people know themselves only *slightly* better than we know them. It seems odd to reason that since I am harder to know than you, I must know myself much better than you know me, whereas you know yourself only slightly better than I know you. Likewise, it seems odd to reason that since introspection is more important for knowing me than for knowing you, I must know myself much better than you know me, whereas you know yourself only slightly better than I know you. Thus, the knowability and introspection illusions seem unable to explain the effects discussed in §5.

Taken together, these considerations make what seems to us an overall strong case in favor of our explanation from §4 and against the ones reviewed in §3.

7. A Third-Personal Illusion?

In this final section, we want to bring attention to a hypothesis that is related to, but importantly different from, the ones discussed in the foregoing. This is the hypothesis that people tend to think that they are better at knowing others than others are at knowing others.¹¹ To illustrate, if Alice and Ben consider how well they know a third person Claire, Alice will tend to think that she knows Claire better than Ben does, and Ben will likewise tend to think he knows Claire better than Alice does. Let us call this putative effect the *third-personal* illusion of asymmetric insight. To our knowledge, this illusion has not yet been tested by experiment.

¹¹ We thank an anonymous reviewer for urging us to consider this.

Apart from its independent interest, the third-personal illusion is interesting in the present context because it may provide further assistance in adjudicating between the competing explanations of the illusion of asymmetric insight discussed so far. The illusion is predicted by our own explanation because, on this account, we will tend to overestimate our own ability to know others, while at the same time underestimating the ability of others to know us. And since this underestimation is not based on us believing ourselves to be especially hard to know, but rather on a belief about the superiority of self-knowledge in general, the resulting underestimation of others will not be restricted to their knowledge of ourselves in particular. The explanation thus predicts that people will tend to think that they are better at knowing others than others are at knowing others.

The third-personal illusion is also predicted by the explanations from self-enhancement and naïve realism. If the illusion of asymmetric insight is primarily a manifestation of the general tendency to think that one is above-average on a wide range of desirable qualities and abilities, including the ability to know others, one will plausibly also manifest the third-person asymmetric insight illusion. Likewise, if the illusion of asymmetric insight is primarily a manifestation of naïve realism, which leads one to assume that those who disagree with oneself must be mistaken, one will also manifest the third-person asymmetric insight illusion. Note, however, that even if the third-personal illusion is eventually confirmed, we have already seen that our explanation remains superior to the explanations from self-enhancement and naïve realism.

Interestingly, the third-personal illusion is *not* predicted by the explanations based on the illusions of knowability and introspection. According to both of these explanations, the illusion of asymmetric insight arises as a result of believing *oneself* to be especially difficult to know. But such a belief does not have any obvious implications for how one will compare the abilities of oneself and others to know *other* people. One may well regard oneself as difficult to know, while regarding oneself and others as equally good at knowing others. If it turns out to be true, the third-personal illusion will thus speak in favor of our explanation, and speak against Pronin et al.'s explanations; and *vice versa* if the hypothesis turns out to be false.

8. Conclusion

We began this paper by reviewing a group of studies reported by Pronin et al. (2001), which suggest that people tend to think that they know others better than others know them (§2). We then argued that existing explanations of this “illusion of asymmetric insight” are wanting (§3), and went on to propose a new explanation of the illusion, based on two other well-documented psychological phenomena: the tendency to accommodate ambiguous evidence in a biased manner, and the tendency to overestimate how much better we know ourselves than we know others (§4). Subsequently, we showed that our explanation can also account for two further psychological effects documented by Pronin et al. (§5), and argued that it outperforms existing explanations on several different counts (§6). Finally, we described a hitherto untested “third-personal” illusion of asymmetric insight, predicted by our account, which we hope to test in future experimental work (§7).

References

- Andersen, S. M. and Ross, L. (1984): “Self Knowledge and Social Inference: I. The Impact of Cognitive/Affective and Behavioral Data.” In: *Journal of Personality and Social Psychology* 46: 280-93.
- Carruthers, P. (2008): “Cartesian Epistemology: Is the Theory of the Self-Transparent Mind Innate?” In: *Journal of Consciousness Studies* 15(4): 28-53.
- Carruthers, P. (2010): “Introspection: Divided and Partly Eliminated.” In: *Philosophy and Phenomenological Research* 80: 76-110.
- Gilovich, T. (1991): *How We Know What Isn't So*. New York: The Free Press.
- Gilovich, T. and L. Ross (2015): *The Wisest One in the Room: How You can Benefit from Social Psychology's Most Powerful Insights*. New York: The Free Press.
- Goffman, E. (1959): *The Presentation of Self in Everyday Life*. New York: Anchor Books.
- Gopnik, A. (1993): “How we Know Our Minds: The Illusion of First-Person Knowledge of Intentionality.” In: *Behavioral and Brain Sciences* 16(1): 1-14.

- Greenberg, J., T. Pyszczynski, and S. Sheldon (1982): "The Self-Serving Attributional Bias: Beyond Self-Presentation." In: *Journal of Experimental Social Psychology* 18(1): 56-67.
- Greenwald, A. (1980): "The Totalitarian Ego." In: *American Psychologist* 35: 603-18.
- Ichheiser, G. (1949): "Misunderstandings in Human Relations: A Study in False Perception". In: *American Journal of Sociology* 55(Suppl.).
- Jones, E. and R. Nisbett (1972). "The Actor and the Observer: Divergent Perceptions of the Cause of Behavior." In E. Jones, D. Kanouse, H. Kelley, R. Nisbett, S. Valins, and B. Weiner (Eds.), *Attribution: Perceiving the Causes of Behavior*, General Learning Press: 79-94.
- Kennedy, K. and E. Pronin (2008): "When Disagreement Gets Ugly: Perceptions of Bias and the Escalation of Conflict." In: *Personality and Social Psychology Bulletin* 34(6): 833-48.
- Kozuch, B. and S. Nichols (2011): "Awareness of Unawareness. Folk Psychology and Introspective Transparency." In: *Journal of Consciousness Studies* 18(11-12): 135-60.
- Kunda, Z. (1990): "The Case for Motivated Reasoning." In: *Psychological Bulletin* 108: 480-98.
- Lord, C., L. Ross, and M. Lepper (1979): "Biased Assimilation and Attitude Polarization: The Effects of Prior Theories on Subsequently Considered Evidence." In: *Journal of Personality and Social Psychology* 37: 2098-2109.
- Lun, J., S. Kesebir, and S. Oishi (2008): "On Feeling Understood and Feeling Well: The Role of Interdependence." In: *Journal of Personality* 42(6): 1623-28.
- Markus, H. (1983): "Self-Knowledge: An Expanded View." In: *Journal of Personality* 51: 543-65.
- Mezulis, A., L. Abramson, J. Hyde, and B. Hankin (2004): "Is There a Universal Positivity Bias in Attributions? A Meta-Analytic Review of Individual, Developmental, and Cultural Differences in the Self-Serving Attributional Bias". In: *Psychological Bulletin* 130(5): 711-47.

- Nisbett, R. and T. Wilson (1977): "Telling More Than We Can Know: Verbal Reports on Mental Processes." In: *Psychological Review* 84: 231-59.
- Oishi, S., S. Akimoto, J. Richards, and E. Suh (2013): "Feeling Understood as a Key to Cultural Differences in Life Satisfaction." In: *Journal of Research in Personality* 47(5): 488-91.
- Oswald, M. and S. Grosjean (2004): "Confirmation Bias." In: R. Pohl (ed.), *Cognitive Illusions. A Handbook on Fallacies and Biases in Thinking, Judgment and Memory*. New York: Psychology Press: 79-98.
- Pronin, E. (2009): "The Introspection Illusion." In: *Advances in Experimental Social Psychology* 41 (Ed. M. Zanna). Academic Press: 1-67.
- Pronin, E., J. Kurger, K. Savitsky, and L. Ross (2001): "You Don't Know Me, But I Know You: The Illusion of Asymmetric Insight." In: *Journal of Personality and Social Psychology* 81: 639-56.
- Pronin, E., D. Lin and L. Ross (2002): "The Bias Blind Spot: Perceptions of Bias in Self Versus Others." In: *Personality and Social Psychology Bulletin* 28(3): 369-81.
- Pronin, E., T. Gilovich, and L. Ross (2004): "Objectivity in the Eye of the Beholder: Divergent Perceptions of Bias in Self Versus Others." In: *Psychological Review* 111(3): 781-99.
- Pronin, E. and M. Kugler (2007): "Valuing Thoughts, Ignoring Behavior: The Introspection Illusion as a Source of The Bias Blind Spot." In: *Journal of Experimental Social Psychology* 43(4), pp. 565-78.
- Pronin, E., J. Fleming, and M. Steffel (2008): "Value Revelations: Disclosure is in the Eye of the Beholder." In: *Journal of Personality and Social Psychology* 96: 795-809.
- Riess, M., P. Rosenfeld, V. Melburg, and J. Tedeschi (1981): "Self-Serving Attributions: Biased Private Perceptions and Distorted Public Descriptions." In: *Journal of Personality and Social Psychology* 41(2): 224-31.
- Ross, L., and A. Ward (1996): "Naive Realism in Everyday Life: Implications for Social Conflict and Misunderstanding". In T. Brown, E. Reed, and E. Turiel (Eds.), *Values and knowledge*. Hillsdale, NJ Erlbaum: 103-35.

- Schlenker, B. and R. Miller (1977): "Egocentrism in Groups: Self-Serving Biases or Logical Information Processing?" In: *Journal of Personality and Social Psychology* 35(10): 755-64.
- Vazire, S. (2010): "Who Knows What About a Person? The Self-Other Knowledge Asymmetry (SOKA) Model." In: *Journal of Personality and Social Psychology* 98: 281-300.
- Wason, P. (1960): "On the Failure to Eliminate Hypotheses in a Conceptual Task." In: *Quarterly Journal of Experimental Psychology* 14: 129-40.
- Wilson, T. and E. Dunn (2003): "Self-Knowledge: Its Limits, Value, and Potential for Improvement." In: *Annual Review of Psychology* 55: 1-26.