Social cognition as causal inference: implications for common knowledge and autism

Jakob Hohwy and Colin Palmer
Cognition and Philosophy Lab
Monash University

[Forthcoming in *Social Objects*, John Michael and Mattia Gallotti (eds.). (This version is a draft and will undergo another round of revisions, please contact us if you wish to quote from it).]

Abstract: This chapter explores the idea that the need to establish *common knowledge* is one feature that makes social cognition stand apart in important ways from cognition in general. We develop this idea on the background of the claim that social cognition is nothing but a type of causal inference. We focus on autism as our test-case, and propose that a specific type of problem with common knowledge processing is implicated in challenges to social cognition in autism spectrum disorder (ASD). This problem has to do with the individual's assessment of the reliability of messages that are passed between people as common knowledge emerges. The proposal is developed on the background of our own empirical studies and outlines different ways common knowledge might be comprised. We discuss what these issues may tell us about ASD, about the relation between social and non-social cognition, about social objects, and about the dynamics of social networks.

*Acknowledgements. Thanks to the audience at the Aarhus-Paris Conference in Social Ontology and Social Cognition at Aarhus University, especially John Michael, Mattia Gallotti, Andreas Roepstorff and Chris and Uta Frith; thanks also to the staff and students at Tokyo University's Centre for Philosophy, in particular, Yukihiro Nobuhara, Iizuka Rie and Kataoka Masa.

Correspondence: jakob.hohwy@monash.edu

1. Introduction

Social cognition concerns the representation of states of affairs in the world that, in a wide variety of ways, involve other people's mental states and agency. It is tempting to try to understand the nature of social cognition by assuming that it is essentially different from non-social cognition, and, consequently, exploring and interpreting behavioural and neurological differences in the light of this assumption. One reason why this assumption is appealing is that creatures with social cognition, like us, seem so different from creatures without much recognisable social cognition. Another reason is that the perception of things like the intentions and beliefs of other people feels more intangible than, for instance, the perception of visual objects. A further reason is that some disorders, in particular ASD, seem to have specific differences in certain aspects of social cognition, suggesting that specialised, dissociable circuits in the brain take care of these functions.

Here, we begin with a different assumption: namely, that though social cognition is no doubt in part processed in domain specific areas of the brain it is *not* essentially different from non-social cognition. To substantiate this approach, we will examine how both social and non-social cognition are instances of causal, perceptual inference. We then propose that what makes some cognition recognizable as social is related to the emergence of *common knowledge*, and explain ways in which underlying problems with cognition in general could lead to profound problems in common knowledge in particular. We explore the consequences of this approach for both our understanding of ASD and social cognition in general.

ASD is an important testing ground for approaches to social cognition because this set of developmental disorders is first and foremost characterized by deficits and differences in social cognition. Individuals with ASD can be deeply socially disabled, with very severe language and communication deficits. Even when language is present there can be profound challenges in the ability to infer other people's mental states. ASD is also characterized, however, by more subtle and difficult to describe sensory and perceptual differences. In fact, it is an astounding characteristic of ASD that seemingly disparate social and non-social symptoms are found together. At times, these sensory differences present as islands of enhanced or superior performance, at other times, performance is diminished relative to the wider population. For example, on one hand individuals with ASD have been found to be less susceptible to some visual illusions than control groups, and on the other hand, they have been found to be less proficient in visual tasks involving the discrimination of coherence between perceptual elements (e.g., motion coherence; reviewed in Happé and Frith 2006). A key question is then whether and how these perceptual differences relate to the social deficits. One possibility is that these features of ASD are independent, another possibility is that the perceptual differences cause the social deficits, a third is that the social deficits cause the perceptual deficits (e.g., through problems with learning). A fourth possibility, which we pursue here, is that the perceptual differences and the social deficits in ASD are different effects of a common cause. We shall understand this common cause to be something afflicting causal inference, which is a process that manifests differently in the perceptual and social domains. The hope is, of course, that this approach will allow for a better understanding of this debilitating and heart-breaking disorder.

The plan of this chapter is to first, in Section 2, describe why social cognition is nothing but causal inference, and then, on the background of this commonality with the non-social perceptual domain, identify some notable characteristics of causal inference that occur when applied to the social domain. In Section 3 we then make the connection between social cognition, understood in this causal way, and common knowledge. Section 4 describes ways in which common knowledge can be challenged and compromised and how this would impact on social cognition. In Section 5, we explore how specific sensory differences hypothesised to occur in ASD could be continuous with compromised common knowledge, and how this may account for profound social deficits in this disorder. We exemplify this point with research performed in our own lab. The overall consequences for our conception of social cognition are then discussed in the final Section 6.

2. Social cognition as causal inference

The paradigm of social cognition that we consider here is *mentalising*, the act of representing other people's mental states. This faculty is invaluable for both predicting and understanding the behaviour of others. For example, if someone says, "the train leaves at three o'clock", we represent them as having the *belief* that the train leaves at three o'clock. Similarly, if someone says, "It is very hot today, isn't it? Do you know when the ice-cream shop opens", we represent them as having the *desire* for an ice-cream.

The representation that occurs in mentalising is entirely analogous to the representation that occurs in non-social contexts. For example, if you hear a particular rapid "tock-tock-tock" noise, then you may well represent the world as having a woodpecker nearby. Similarly, if you see smoke and hear a fire-engine, then you represent the world as having a fire nearby. In both mentalising and non-social representation of the world the process begins with some sensory input, which triggers an inference about what the causal origin of the input might be.

Mostly this inference is unconscious, namely when it concerns perceptual states – this is the unconscious perceptual inference made famous by Ibn al-Haytham and Helmholtz, and defended by Neisser, Gregory and more recently in machine learning and computational neuroscience by Mumford, Dayan, Hinton, Friston and others (Helmholtz 1867; Neisser 1967; Gregory 1980; Mumford 1992; Dayan, Hinton et al. 1995; Friston and Stephan 2007; al-Haytham ca. 1030; 1989). On occasion, such inference can of course also be conscious; for example, you could go through in your mind the various hypotheses about what may cause an individual's statement about the hot weather and the ice-cream shop, or try to imagine different common causes for both the visual input of the smoke and the auditory input from the fire engine. It isn't necessarily the case that the process of inference leading to mentalising is conscious, of course: a mental state attribution may pop into mind as automatically as a visual object does when we shift our gaze. In each case, the ease at which a new perception enters consciousness belies the non-trivial computational demand that an accurate causal inference from the sensory data requires.

If words, gestures, and additional behaviours that we pick up from other people are treated as just being characteristics of sensory input, and if the mental states of other people are treated as the causes of this input, then mentalising can be characterised as causal inference from sensory effects to worldly causes (Wolpert, Doya et al. 2003; Kilner, Friston et al. 2007). Mentalising is then nothing but the kind of causal inference that the brain is in any case consigned to engage in to make sense of the world. Of course, there are differences between social and non-social cognition, but viewed from the perspective of causal inference, we should not expect these differences to be more dramatic than the differences that exist between other kinds of cognition (for example, the difference between the processing of moving and stationary objects, or between 2D and 3D perception). In other words, the dissimilarities between these processes will only pertain to the kind of challenges in performing causal inference specific to a given class of worldly causes of sensory input.

If there are specific difficulties in the application of causal inference to social phenomena, then we are likely to find that they stem from uncertainty in the sensory input. This is because what makes causal inference difficult is the lack of unequivocal one-one relations between cause and effect. Evidence for one-one relations is made uncertain by the presence of noise, ambiguity and non-linear interactions in general. For example, when we see smoke and hear fire engines, there is ambiguity regarding whether the cause of this sensory input could be a real fire, a pretend-fire in a movie set, or harmless smoke from a chimney co-occurring with a fire engine out on a false alarm.

In order to engage in causal inference in spite of uncertainty, the brain can appeal to both the fit between the sensory input and the different hypotheses about its causes, and to prior beliefs about the probability of each hypothesis. For example, I might disregard the 'movie set' hypothesis because it is very unlikely that a movie would be set in my neighbourhood, and I might disregard the hypothesis that the smoke and the fire engine are independent causes of my input because the smoke disappeared very soon after the fire engine sound ceased.

All this is to say that we engage in *Bayesian* reasoning in order to infer the causes of our sensory input (Kersten, Mamassian et al. 2004). Such probabilistic inference is necessary precisely because the sensory input is riddled with ambiguity and uncertainty. The specific manner in which our brains engage in inference in a given context depends heavily on the place of the relevant worldly causes in the overall causal structure of the world. Some causes give rise to their effects in more highly non-linear, context-dependent ways than others and some causes are hidden deeper in the causal hierarchy than others (for example, the subprime mortgages that caused the global financial crisis are deeply hidden and there are numerous non-linearly working factors in the way they cause parts of our sensory input; in contrast, the redness of the apple in front of you is less deeply hidden, though it also depends on contextual factors such as lighting conditions).

It is crucial to add an active element to our understanding of Bayesian perceptual inference, namely in the way we actively test our hypotheses about the causes in the world. For instance, we may engage in more vigorous visual and auditory *search* in order to figure out whether the smoke and the fire engine sound is correlated, or we may check the emergency services on the net to see if a fire is mentioned. Similarly, in the ice cream case, we may *ask* the person whether they feel like an ice cream. This active element is clearly recognized in key treatments of causal inference, where causation is conceptualized in terms of invariant relations under (active) intervention (Pearl 2000; Woodward 2003).

Social cognition, we therefore propose, is nothing but causal, Bayesian inference from sensory input to mental states. To understand social cognition and how it may differ from other areas of cognition, the task is then to specify how uncertainty may arise in the inference from sensory input to mental causes.

Some sources of uncertainty in social causal inference spring quickly to mind. The mental states of other people are quintessentially hidden causes, so hidden in fact that their existence can be doubted on epistemological grounds, leading to skepticism about other minds. This is known as the *other minds problem*. John Stuart Mill famously proposed an inferential solution to this problem, via an argument from analogy with our own, known mental states (Mill 1865). Modern Bayesian accounts of social causal inference merely update Mill's idea. The key observation is that mental causes are deeply hidden, that is, they must be inferred on the basis of various causal links, including observed behaviour. One problem here is that observed behaviour has a rather volatile relation to mental states. Different contexts will make it considerably more or less likely that a particular piece of behaviour is caused by a particular mental state. Famously, this occurs in deception, pretending, and stage-acting, but the point generalizes such that a context can be found which makes any kind of mental state a cause of a certain behaviour (for example, we could assume you have rather bizarre beliefs about what aggressive ice cream shop owners do to force their products on consumers on hot days, and assume you fear such aggression, and therefore infer that your question about opening hours was motivated by a desire to be far away from ice-cream shops on hot days).

It is thus tempting to say that social cognition is special in the sense of being dependent on the context of our existing knowledge regarding the other person's more or less idiosyncratic sets of beliefs and desires. However, this does not seem to set social cognition especially apart from other types of cognition. Context-dependence is everywhere, and can entail many different degrees of difficulty. Already we have mentioned the example of subprime mortgages and the highly context dependent ways they cause other phenomena such as low interest rates and high unemployment. But everyday examples of perception are also highly context-dependent. For example, in the visual occlusion of a cat behind a fence there is a very intricate non-linear interaction between the context of the fence and the observer's movements relative to the fence and the cat, which makes the unconscious perceptual inference of the presence of a cat non-trivial in this specific context. In inference under context-dependence, it is crucial to rely on prior statistical expectations about what the cause and the context might be, as well as on an ability to predict how the flow of sensory input will change under various interventions. For example, we expect the world to be populated by many more whole cats than by curiously aligned, detached cat slices, and we expect things like fences to be stationary in the world as we walk past, seeing the whole cat behind it. Similarly, we rely on statistical regularities about the likely beliefs and desires of people around us as we try to infer their mental states. For example, I rarely consider the possibility that you may have somewhat paranoid beliefs about ice cream shop owners.

Hence, even though mentalising is riddled with context dependence, we should not expect this to be what sets it apart from non-social cognition.

Another candidate for what makes mentalising special has to do with the number of sources of evidence causal inference can appeal to in social contexts. In many instances of perceptual inference, the same cause can be accessed through its different effects on different senses. For example, if you see smoke, you may also expect to smell it, and to feel heat. Similarly, different types of witnesses might be able to provide evidence about the existence of subprime mortgages (mortgagees, lenders, economists). Having multiple (conditionally independent) sources of evidence can be a very efficient way to minimise uncertainty about a causal inference. This kind of case should not be confused with the case where different instruments pick up the same piece of evidence; this can also be useful but speaks more to the reliability of the instrument than the reliability of the evidence. In the courtroom analogy, this corresponds to the difference between on the one hand relying on two lawyers to interrogate the same witness; and on the other hand having one lawyer interrogate two different witnesses. When trying to infer other people's mental states we might of course rely on different senses (e.g., hearing words and seeing the mouth move) but this mostly seems to be akin to two lawyers interrogating the same one witness. This can be used to address uncertainty about the reliability of the senses, but not about the evidence itself. It is less clear how multiple, independent sources of evidence can be made available for the occurrence of a mental state. We can only assess mental states through the behaviour of other people, that is, mental states do not, in very clear ways, give rise to other effects that we could access. This is why there is focus on developing reliable lie-detectors, although such instruments are of course controversial (and are, indeed, not allowed in most courtrooms). One might operate with distinctions within behavioural evidence itself, and claim that this constitutes different sources of evidence (as when we can see the mouth is smiling but the eyes are lying, or hear the person denying they are smoking but smell that they are). We think it is unclear how to treat such cases. Our point is that, even if these cases are allowed, there is still a distinction between the range of sources of evidence that can be brought to bear on mental causal inference, and the wider range that can be brought to bear on most of the everyday objects and events around us that we make everyday inferences about.

We can call this aspect of social cognition *evidential insulation*: relatively few independent sources of evidence are available. Evidential insulation makes it particularly difficult to overcome doubts about the reliability of causal inference. This is thus an element that makes especially good sense once we understand social cognition as causal inference, which must proceed under conditions of uncertainty. Just as doubts about the reliability of a witness in court can be overcome by obtaining a second witness who gives the same testimony, we can overcome doubts about sensory evidence by obtaining further evidence from different sources. Without different sources it may be hard, or impossible, to gain sufficient confidence in one's inference to justify further action. This aspect of social cognition may therefore exacerbate uncertainty associated with the context-dependent nature of mental states, and the especially indirect nature of the evidence that we use to infer their existence (compared to basic visual perception, for example). Evidential insulation is not exclusive to mentalising, since it occurs, for a variety of reasons, in many other types of causal inference. For example, when you hear a sound in the dark you may not have other sources of evidence available (you're camping in the bush and the torch has run out of battery). However, we do think that it occurs *systematically* for mentalising and in this respect is a marker of social cognition.

The last contributor to uncertainty in social causal inference that we will consider is, in fact, specific to the social domain. In general, this factor has to do with the kind of uncertainty that stems from non-linear interactions between causes, and as such is of a piece with all other kinds of causal inference (for example, inferring the whole cat behind the fence). But in the social domain there is an intricate, special level of non-linear interaction: when we interpret other people we are often aware that they are also interpreting us and that their behaviour depends in a non-linear way on which mental states they interpret us as harbouring. For example, when, on a hot day, the kids ask for the opening hours of the ice cream shop, you are interpreting their verbal behaviour under a model of the world that includes their model of your mental states; it is crucial for the further negotiations to understand that they ask this question under the hypothesis that you might allow them to get an ice cream – you could therefore lie and say the ice cream shop is closed all day. This aspect of mentalising we might call *meta-mentalising*. It is a fascinating concept because the interaction of mental causes is so pervasive: it can even be necessary to model how other people model you modeling them, and so on. This comes about because other people are agents: that is, their intervention in the causal chain is contingent upon their model of the world, and how they intervene impacts both on what you experience and how you model them.

It is important to recognize that the need for meta-mentalising arises only because causal inference in general is challenged by non-linearly interacting causes. It is just that we happen to find ourselves in an environment with sensory input from worldly causes (i.e., other people) who can act, conditional on what our and their own mental states are. This does not mean mentalising is different from causal inference, but there does not seem to be any other area of causal inference where meta-modelling is required to overcome non-linearity.

In this section, we have argued that social cognition, in particular in the shape of mentalising, is nothing but causal inference on hidden causes of sensory input. We pointed out that, as such, social cognition can only be set apart from other areas of cognition by the way causal inference is challenged by sources of uncertainty and ambiguity. This lead to the suggestion of two factors in particular, which each contribute to uncertainty and ambiguity in social cognition. The first factor was systematic evidential insulation and the second factor was the need to engage in meta-mentalising.

This proposal goes somewhat against an assumption that lies behind much research on social cognition: namely, that there are domain specific elements in social cognition. The benefit of taking our approach is that the nature of social cognition, and the important challenges to social cognition in mental and developmental disorder, can be understood exclusively in terms of how causal inference occurs under uncertainty, which is a well-studied, standard problem-set in science. Because this approach makes social cognition continuous with all other areas of causal inference it holds potential for understanding how, for example, the social deficits in ASD are connected to the more poorly understood perceptual differences in this condition. In other words, social and non-social deficits may be different manifestations of an underlying issue with causal inference under uncertainty, where the apparent differences in these symptoms are driven by domain-specific factors creating different constellations of uncertainty. (Note that this is not to claim that there are no areas of the brain that are specifically engaged in mentalising, neuroscience evidence certainly suggests that there are such areas or modules; the claim is merely that such areas are engaged in causal inference too, just like areas engaged in other domains of perception; what makes it special are the constraints under which such inference proceeds, as we suggested above and continue to develop below).

3. Common knowledge in social cognition

Having argued that social cognition can be reduced to causal inference, we now proceed to characterise an important purpose of mentalising. Specifically, we focus on what people get out of representing mental states not just as simple causes in the world but as causes that themselves represent and meta-represent other people's mental states including our own. This is an important concept to consider in identifying what people may get out of engaging in causal inference about other people's mental states. With this focus on meta-representation we are able to speak specifically to a factor that we argue makes social cognition a particular kind of causal inference.

The idea we wish to pursue is that the main purpose of representing, and re-representing, other people's mental states, including their representation of our own mental states, is to enable common knowledge. Common knowledge is a technical notion, deriving from economics, semantics, and epistemology. We can introduce the idea with a famous example from one of the first treatments of this concept.

> When a man loses his wife in a department store without any prior understanding on where to meet if they get separated, the chances are good that they will find each other. It is likely that each will think of some obvious place to meet, so obvious that each will be sure that it is "obvious" to both of them. One does not simply predict where the other will go, since the other will go where he predicts the first to go, which is wherever the first predicts the second to predict the first to go, and so *ad infinitum.* Not "What would I do if I were she?" but "What would I do if I were she wondering what she would do if she were wondering what I would do if I were she … ?" *(Schelling 1960: 54).*

Schelling is describing a *coordination problem*, where the married couple needs to coordinate such that they both go to the same place (although in this case it doesn't matter exactly where that place is, just that they both get there). For this problem to be solved it is not enough to represent simply where that place might be but it is also necessary to represent the spouse's knowledge of what the place might be, and the spouse's knowledge of where the first spouse believes the meeting place is, and so on. The solution must involve, in Schelling's terms, that they "must 'mutually recognise' some unique signal that coordinates their expectations of each other" (*ibid*.).

This sets common knowledge apart from mere mutual knowledge. In mutual knowledge, people know the same thing: we may all know that the game will be shown in the park. But mutual knowledge can fall short of solving the coordination problem of deciding where to go tonight, because you may not know whether other people know that the game is on in the park, and this may matter to you because you don't want to end up in the park alone, or at home while everyone else goes to the park. So you need to also know that others know that the game is on in the park. But of course if you only know that others know that, then they might not go because they might not know that you and others know about the game in the park, or indeed that you know that they know that you know, and so on. In fact, to solve the coordination problem, an infinite hierarchy of knowledge about each others knowledge must be established. What establishes this hierarchy is not an actual infinite series of mental states in each person's brain but Schelling's unique signal that is mutually recognized. This signal can be very many different things. In the department store example it might be knowledge that the spouses would each find it amusing if they found each other in the wine store, in the park example it might be a particular tweet.

In other cases it might be something as simple as eye contact. In a well-rehearsed example, two friends enter a full bus, but end up sitting at opposite ends of it. At a stop halfway through the ride a third person, also a friend, calls out from the street to ask whether the two on the bus would like to come for a drink. This initiates a coordination problem for the two friends on the bus: both want to get off the bus together to get the drink, and if not that then they both want to stay on the bus, foregoing the drink; but neither wants to leave the other behind. The key here is whether each of the two friends on the bus knows that the other heard the third friend's invitation, and knows of each other that they heard this, and so on. This knowledge, and thus common knowledge, can be established if they both look up at the same time and their eyes meet, whereupon they are assured that the message was heard, and that they both know it, and know that they know it, and so on, and they therefore both alight the bus to get the drink. Thus, while the difficulty in consciously holding in mind multiple levels of the hierarchy of knowledge states required for common knowledge is an argument against the behavioural relevance of this concept, our brains may all the same be tuned to recognise cues that establish common knowledge efficiently.

In general, a proposition P is common knowledge for S and S' if and only if, S and S' know that P, S knows that S' knows that P, and knows that S' knows that S knows that P, and so on; and similarly for S'. There is a very sizable literature on common knowledge, and different formalisations, interpretations and applications of it (for a classic statement, apart from Schelling, see (Lewis 1969); for review see (Vanderschraaf and Sillari 2009)).

We want to make the observation now that common knowledge manifests pervasively, and that insults to the ability to establish common knowledge will have profound and variegated effects on one's communal function. Straight off, the kinds of cases where common knowledge is useful can seem rare and recherché. It does not seem central to the human endeavor to finesse with common knowledge our ability to find each other in department stores, to alleviate awkward situations when the waiter spills the soup, or to have a convention for who should call back (the caller or the called) when the connection goes (Lewis 1969). But of course common knowledge is everywhere, for social creatures like us who live in close quarters with each other and whose trajectories constantly cross. A good example is the convention to drive on the left (or the right as the case might be). We don't just have mutual knowledge that driving is on the left, we have common knowledge: I would not go on the roads if I didn't know that you know that driving is on the left, and you know that I know, and so on.

Driving is an example where there are two equilibria, namely where we all drive on the left or all on the right. We don't care which it is as long as we all do the same thing and we are confident that this is established as common knowledge. These cases are not rare but it is important to observe that there are cases as well where we do care which of several equilibria we end up deciding on. The stag hunt is one such case. In this classic example, two hunters can each hunt rabbit or stag. There are two equilibria, namely where we both hunt rabbit or we both hunt stag. Each hunter is not interested in the scenario where he or she goes stag hunting alone, because it is impossible to kill a stag without collaboration. Importantly, both hunters are more interested in sharing the stag than getting a rabbit each, because this way they individually get more to eat. Common knowledge helps with the navigation of this scenario because the hunters need to set up mutual expectations that they are going to do the same thing. Similarly, in the example with the two people on the bus, they both had a preference for getting off to get a drink, but only if they both get off.

So common knowledge plays a role in the great many endeavors where we jointly engage in some activity: particularly in situations where it matters that we do the same thing, that we together achieve an outcome that is optimal for each of us individually, and that we all know what others know, and so on. This even applies to simple, everyday matters such as cooking dinner. Even though the family members all know dinner is at seven, you will not be enthusiastic about cooking dinner for everyone unless you know that they know that dinner is at seven, and that they know you know that they know that dinner is at seven – if they don't know this then they will not expect dinner to be at seven after all. Moreover, even though there are many solutions to the coordination problem of all being at the dinner table when dinner is served, members of the family will all prefer the final decision to be that dinner is at seven because that's when they are hungry. Common knowledge is essential not only in cases where we need to establish awareness of a specific individual's intentions, but also for the function of shared rules and interpretations.

Michael Suk-Young Chwe (Chwe 1999; Chwe 2000; Chwe 2001) has developed a set of intriguing analyses of cases involving common knowledge. These analyses are important in part because they anchor common knowledge in a very wide set of social contexts. For example, Chwe analyses the decision to revolt in terms of common knowledge. He notes that people will have a threshold for when they will revolt, that is, they will revolt only if a certain number of other people also revolt. But of course it matters to your decision not only what your threshold is but also what other people's thresholds are. You might be prepared to revolt if 2 others do so, but everyone else might only want to revolt if there is a million on the street already; if you know their thresholds then you know that your low threshold is pretty immaterial. There will also be cases where meta-mentalising is crucial. If three people communicate their thresholds of three to each other, and they know that this has been communicated, then they know that they occupy a world where the three of them have a desired equilibrium – and so they can each revolt.

If four people each have a threshold of three then we should expect revolt to occur – but this in fact depends on the shape of their social group, and whether this shape is itself common knowledge. If their communication is organized in a *square* then the revolt will not happen, because in this case common knowledge is not engendered. In a square, Person 1 communicates her threshold with 1 and 4 but not 3; similarly, Person 2 communicates with 1 and 3 but not 4, and so on. This means that Person 1 cannot rule out that Person 3 has a threshold of five, and therefore she cannot rule out the possibility where Person 2 and 4 will not revolt, so she will not herself revolt. The key here is that the knowledge she misses is knowledge about what her neighbors know. Similar cases hold for all four people, so the revolution doesn't happen because they each do not have knowledge of what other people know.

If instead the group was organized as a *kite*, that is a triangle made up of Persons 1, 2, and 3, with person 4 dangling at the tail, then the revolt will happen, albeit with only three people. This is because now each of the three in the triangle know what each other's threshold is and know that they each know this, and so on. The fourth person is unable to revolt due to an inability to establish common knowledge.

Importantly, under this analysis the shape of the social network must itself be common knowledge such that the participants must know whether they are organized in a kite or a square. That is, they must know who communicates with whom and how. In other words, mentalising and meta-mentalising must proceeds under models of the wider social landscape, including models of whom the people you talk to talk to.

Chwe discusses a number of interesting elements to this kind of analysis. One element is the distinction between *strong* and *weak* links that can exist between participants in a social network (Chwe 1999). Strong links differ from weak links in how probable it is that the friends of your friends are your friends too. If the probability of this is low, then the network is more a network of acquaintances than of close friends. When an individual passes a message in a network of strong links, they know that the likelihood of others in the network receiving the message, and the likelihood that others know that the rest of the network has received the message, is increased due to the shared knowledge that the network is highly interconnected. Chwe's analysis shows that strong links are good for ensuring participation (in revolt, etc.) when thresholds are low, because strong links ensure good communication in small groups. On the other hand, weak links are better for participation when thresholds are high, because information traverses weakly linked networks more quickly. Common knowledge scenarios therefore depend on an interaction between thresholds and weak vs. strong links; conversely, the shape of social networks can be expected to reflect the common knowledge scenarios they focus on (small, strongly linked scenarios might involve cases like when to make and come for dinner, and larger, weakly linked scenarios might involve cases like fashion trends or, indeed, revolution).

A second element is the notion of *bandwagons* and their fragility (Chwe 1999). A bandwagon is, for example, a situation where Person 1 has a threshold of 1, so revolts, Person 2 has a threshold of 2, so revolts on knowing that Person 1 has a threshold of 1, and Person 3 has a threshold of 3 so revolts on knowing about the thresholds of the first two, and so on and so forth for the rest of the people in the group. Bandwagons are very dependent on the thresholds and reciprocality of the first few links. If Person 1 and 2 both have a threshold of 2, then nothing will happen across the whole group of people if communication is one-way only between Person 1 and Person 2. If communication is reciprocal in such a way that common knowledge is established, however, then the bandwagon can get going. Roughly put, this means that without reciprocality one will be less engaged in taking initiatives for social collaboration and will be left more to one's own devices.

A third element concerns the formation of *cliques* and the flow of information between cliques (Chwe 2000). For example, a leading clique might be a group of three people each with a threshold of three, organized in a triangle. This clique will revolt, and this will be known to a follower clique of two people each with threshold 5, who will revolt, knowing about the leading clique. Notice that here the follower clique needs to model the shape and common knowledge properties of distinct groups, and at the same time model their own group in relation to this. That is, they need to interpret their own local knowledge in a more global network of groups. Being too "myopically" focused on one's own group means that the behaviour of leading cliques will be missed and one's own group will fail to join the collaborative action.

In addition to discussing these elements of common knowledge networks, Chwe (2001) offers many examples where common knowledge is crucial for the way groups are organized and interact. Common knowledge thus becomes a key element in the understanding of ritual, advertising, and the organization of public fora. Ritual dancing, for example, is interpreted as a tool for ensuring joint attention on the common knowledge signal, and easy detection of those who fail to attend. This ensures that the participants know that everyone got the message, and that they know that everyone knows that they got the message, and so on.

There are methods other than ritual to ensure people's attention to a common knowledge signal. In general, creating a signal with much redundancy helps because then it is more likely that many people will notice it and also notice that many people notice it, and so on. With this in mind, one can look at important events that initiate common knowledge based processes. The revolts of the Arab Spring, for example, purportedly began with the tragic self-immolation of the Tunisian street vendor and protester Mohamed Bouazizi. Though there may have been many protesters before him, the act of self-immolation is a signal that carries immense redundancy and as such many people would see it and see that many people see it.

Interestingly, Chwe broadens the discussion of common knowledge to include objects too. That is, some objects exist in such a way that for most people they are represented in a manner that involves common knowledge. Chwe's main example is the marketing of the mouthwash product Listerine. Listerine was originally an antiseptic, and few would consider putting it in the mouth. But through blanket marketing that focuses on the medically-sounding term 'halitosis' for bad breath, the makers of Listerine made it common knowledge both that halitosis was widespread and that your friends will not tell if you suffer from it. The thought is that you will be more inclined to buy Listerine if you know that other people are likely to have halitosis, that they are likely to know about halitosis and its “treatment” through Listerine, that they are not likely to tell you about your halitosis, and that they know that you know about all this. Even though blanket marketing is in many ways characterized by redundancy, the redundancy helps create the common knowledge that sells the product. This means we can reasonably classify Listerine as a common knowledge object, or as a social object. Its representation is embedded in a functional role that involves what other people know, what other people know about what other people know, and so on. Chwe analyses Kotex, HIV tests, and Macintosh computers in a similar vein. Expanding the common knowledge conception to objects is important because it underscores the point that common knowledge is pervasive in our everyday lives.

The picture so far is then that common knowledge is a pervasive element of social cognition, and that social cognition is to be understood as causal inference. We mentioned that meta-mentalising is an element of what makes social causal inference stand apart, and it is clear that meta-mentalising is a crucial part of common knowledge endeavors. We do not want to claim that meta-cognition is needed for all and only common knowledge, but it is tempting to think that they are closely related nevertheless. That is, in some instances it may be useful to model other’s mental models of one’s own mental model, even if common knowledge is not in the offing. This may be the case in deception, for instance. But even in these cases perhaps it is useful to model precisely to check whether or not common knowledge can be established. For example, it may be that you stand to gain more by not joining collaborative action: perhaps you know that your fellows are poor stag hunters so you will gain more by knowing that they will be off hunting the stag while you scoop up the rabbits.

The proposal is then that the mentalising and meta-mentalising that comes with common knowledge processing is a pervasive and central part of the causal inference involved in social cognition. This includes all the different elements identified by Chwe, from the uptake of Listerine to the discerning of social network shapes like kites and squares. With this proposal in mind, we next turn to the ways in which common knowledge formation can be challenged and disrupted.

4. Challenges to common knowledge

Common knowledge requires, in Schelling's formulation, a unique signal that is mutually recognized such as to coordinate expectations. We have seen that this signal may take many forms: eye contact, communication about thresholds, blanket advertising, self-immolation, etc. The context in which this signal is delivered matters. If delivered during ritual dancing, attention may be ensured, and if delivered through mass events (like the NFL Super Bowl) uptake can be ensured. Also, the context of social groupings, such as cliques, will matter for how signals are processed, as will issues like communication reciprocity in relation to thresholds for bandwagons. We have also seen that there is a varied class of events and objects to which common knowledge signaling is relevant.

This means there is a rich tapestry of situations where common knowledge can be challenged and disrupted, where such situations will pertain to the processing of Schelling's unique signals.

A classic example of this is the *Byzantine generals problem*. Two generals, each situated on a separate hill, want to attack a city in the valley between them. They must attack together to succeed, or not attack at all, since a lone attack will be disastrous. The first general sends a messenger through the perilous valley with a message to attack at dawn but will of course only attack if receiving confirmation from the second general that he or she has received the first message. But the second general will not be happy to attack unless receiving a message confirming the first confirmation was well received. And so on. The consequence is that the attack never happens. The problem concerns the uncertainty about whether the messenger got through the enemy lines down in the valley. It would be solved if the signal could be made unmissable, for example, by agreeing to have a massive signal fire on each hill top – but in this scenario this would alert the city below too. This problem will arise whenever a signal isn't known to carry perfect information. Of course the stakes are not always as high as in the case of the generals, so often sending just a few messages will be deemed good enough. But the quality of the communication channel is a challenge to solving coordination problems.

Notice that in this case, it is the reliability of the communication channel that matters – how well it carries information about the mental states of the sender. Across different kinds of cases, this would be a matter of degree. Some communication channels are more precise than others. This means the severity of the problem will differ from case to case. But similarly, the severity of the problem will depend on the participants' expectations about the reliability of the communications channels. If a participant expects the communication channel to be very unreliable, then little trust is placed on the incoming message, and the urge bigger to enter a new round of messaging. This is a simple point that corresponds to the urge to sample for longer when one expects variability. Mismatches are then bound to occur when the communicators have different expectations for the precision of the signal. In particular, if one party believes the signal is as clear as a beacon on a hilltop but the other party thinks that it is as unreliable as a messenger sent through volatile territory, then we should expect common knowledge to suffer.

The notion of expectations of precision is essential to causal inference. A given hypothesis about the cause of sensory input will have different strengths if the signal in one case can be trusted to be very reliable and in other cases not. Uncertainty in the signals we base our inferences on is state-dependent, that is, it may vary according to the context in which the signal occurs. This means that levels and regularities of uncertainty must be learned, and inform causal inference in the shape of expectations for uncertainty, or precision. This holds for all types of causal inference, including the kind that the Byzantines generals each engage in when trying to decipher the mental state of the other general from the context and the signals sent.

The occurrence and robustness of common knowledge also depends on the degree of *alignment* between the participants. Alignment should here be understood as the degree to which different individuals share their initial beliefs about the world and the present situation in particular, including the probabilities assigned to those beliefs. This matters because the more aligned participants are, and the surer they know this, the more they can be sure that a new message, when sent, will also be interpreted in the predicted way. Alignment is then a tool for reduction of uncertainty in message passing for the purposes of common knowledge. Perhaps we can add to Chwe's account of the role of ritual here: not only is ritual used to ensure attention to the signal, it also serves to shape the prior expectations of the participants, such that there is less uncertainty about whether they interpreted the message in the right kind of way.

This means that misalignment is a challenge to common knowledge. Knowledge regarding what other participants know (and what they know about others), on the basis of a unique signal, is undermined if we are not sure that the participants employed the right frame of reference to the signal.

The list of challenges to common knowledge also extends to the notion of bandwagons. We noted that they are sensitive to the thresholds and reciprocality of the first few links. In particular, having reciprocal communication and common knowledge at the first links can do much to ensure that a bandwagon gets started. This requires a level of sophistication in the causal inference at play. For example, when Person 2 receives a message from Person 1 then it may be fruitful to engage in turn-taking where a message is sent back. This requires inference of not only Person 1's threshold but also of that person's representation of Person 2's threshold. If this level of representation is challenged, then bandwagons may get stuck.

Finally, we make the general point that the causal inference required for common knowledge is context dependent and hierarchical. It is rare that signals are as unequivocal as beacons on the top of hills or self-immolations. Mostly signals are imbued with a degree of uncertainty and ambiguity. Confident inference then requires tools for reducing uncertainty and for resolving ambiguity. In order to do this the participant needs to appeal to prior knowledge and to active probing of the situation. That is, if you are not sure how to interpret a signal, then you can down-weight the character of the signal itself and instead begin to appeal to your prior conceptions about the situation and the likely way the message was meant. Of course, these prior conceptions will involve prior beliefs about the degree of alignment too. This process attempts to reduce uncertainty by taking the wider context into account, namely in the shape of longer term, learned regularities about what to expect. Similarly, a vague idea about what the best hypothesis might be can be tested actively, by predicting what the interlocutor would say in response to a particular question, were the hypothesis in fact correct ("if she said her threshold is three, then she will confirm she will revolt when I tell her that myself and my neighbor will revolt if she does"; this is neural hermeneutics, see (Wentzer and Frith forthcoming), or, more generally, a social version of active inference (Friston, Mattout et al. 2011).

Notice that both of these tools for disambiguation of the signal depend on the expectations for the precision of the signal, which we mentioned before. If one expects much precision in the signal, then one will sample it for longer before appealing to prior conceptions and longer-term regularities; vice versa for the case where imprecision is expected. Similarly, if one expects much precision then one will sample for longer before resorting to actively testing a given hypothesis.

For all these challenges to common knowledge, there is the prospect of interactions and cascading effects. For example, if there is a unilateral problem with trusting a signal, then common knowledge is not established and bandwagons may fail or misalignment result, which again results in more difficult common knowledge consumption. Similarly, if an individual is expecting more than normal precision in a situation loaded more than usually with uncertainty, then the distance between that person and others in terms of their ability to make sense of signals is going to be compounded.

There is therefore ample scope for challenges and disruptions to common knowledge. We have presented this in terms that lend themselves to both the social deficits seen in ASD, and theories about which sensory deficits may be present in this disorder; we now turn to this issue.

5. Causal inference differences as a common cause of sensory differences and common knowledge differences.

We first presented social cognition as a matter of causal inference, then we described common knowledge as a major ingredient in social cognition and outlined types of challenges to common knowledge. We now want to bring these elements together, using ASD as the key test-case.

There is a very direct way to relate ASD and common knowledge: begin by positing mindblindness (i.e., a local deficit to a specialized mentalising circuit in the brain), then observe that common knowledge requires mentalising, and predict widespread difficulties with common knowledge processing in ASD. We believe this explanatory strategy is uninformative because it misses important aspects of the nature of both mentalising and ASD. If mentalising is just another type of causal inference, then mentalising deficits should be associated with a problem with causal inference; this direct strategy is blind to such issues. Similarly, this strategy ignores the presence of a wide class of non-social sensory differences in ASD, which are now so well-recognised that the upcoming fifth edition of the Diagnostic and Statistical Manual of Mental Disorders will for the first time include sensory dysfunction as a diagnostic criterion for ASD (i.e., "hyper- or hypo-reactivity to sensory input or unusual interest in sensory aspects of environment," Huerta, Bishop et al. 2012). It is hard to understand why there should be differences in very basic sensory processing if ASD is just a domain-specific deficit in mentalising (or indeed why a social deficit could cause a difference in the ability to perceive, for example, visual illusions).

Above we advocated a common cause model of the sensory differences and social deficits in ASD such that the same underlying aspect of causal inference causes both. We believe this explanatory strategy is more promising. It has the potential to explain constellations of traits in the ASD spectrum in more detail than an account of a local mentalising deficit. It also has the potential to create a deeper understanding of the nature of mentalising and social cognition more broadly: we can present social cognition as a type of cognition in general, rather than a somehow specialized module, and we can present social cognition as the upshot of causal inference.

Notice that the explanatory strategy that we favour also differs from an approach that begins with the sensory differences and explains the social deficits as caused by them. On our account it is a deeper aspect of causal inference that underlies both. The challenge in adopting this strategy is to explain why the social deficits in ASD are so profound and the sensory differences comparatively more subtle. It is in order to discharge this explanatory burden that we appeal to the intricate causal inference involved in common knowledge processes. We think that there is a type of difference in causal inference that can explain both subtle sensory differences and profound social deficits.

The underlying factor in causal inference that we will focus on is what we mentioned in the previous section, concerning the expectations for the precision of sensory input. This is a key ingredient in the idea that the brain processes its sensory input by minimizing prediction error (or more generally, free energy; Friston and Stephan 2007; Feldman and Friston 2010; Friston 2010; Brown, Friston et al. 2011). It is of particular interest for mental illness and developmental disorders like ASD because differences in this factor have the potential to regulate the relative weighting in causal inference of top-down prior expectations and bottom-up sensory input. That is, when sensory precision is expected, top-down priors are weighted less relative to bottom-up signals, and when imprecision is expected they are weighted more. In general, having such a mechanism is crucial to causal inference because it ensures the reasonable principle that one should base one's causal inference on reliable evidence, or retreat to priors when reliability drops off.

This is then also related to people's tendency to sample either more or less in sense perception, to shift attention, attend to detail, and to be sensitive to overall context – all aspects that are implicated in ASD. Expected precision, as we have mentioned, is related to learning of state-dependent levels of noise and uncertainty. This means that different levels of expected precision for different people can be expected to manifest differently for different contexts, giving rise to the varied landscape of sensory differences and perhaps the heterogeneity of symptoms on the autism spectrum.

The specific proposal is that as individuals get higher and higher on the autism spectrum, they tend to expect more and more precision in their sensory input (this proposal is worked out in some detail in Hohwy (2013 (forthcoming)). Heightened expectations for precision can be beneficial for some tasks because it is related to heightened attention and increased sampling. But similarly, it can be detrimental in other tasks, when the signal in fact is deteriorating, and when the context should be used to squash uncertainty. Because the concept of expected precision is very basic to all kinds of causal inference, and with a potential to cascade into many kinds of inference, it is conceivable that a domain-general trait bias in expectations for precision, that is very different from the majority, will present clinically. Thus, while for typical individuals the degree of precision expected from sensory input during causal inference should vary across contexts, here we suggest that expectations for precision are consistently high in ASD.

From a statistical point of view, expectations for precisions are related to the confidence of causal inference. As such they are part of second-order statistics. This means that more drastic problems with optimizing one's expected precisions can be very hard to rectify. It is basically a type of inference that is itself meant to ensure the reliability of first-order inference, so ensuring its own reliability requires going to a third level of statistics, and so on. This comes with metabolic costs and danger of regress that we don't think the brain can comfortably encompass. This aspect of expected precision then speaks to the recalcitrance of mental, developmental disorder such as ASD (and schizophrenia, see (Hohwy 2013 (forthcoming)).

This proposal finds a natural partner in the *weak central coherence account* of autistic perception (Frith 1989; Happé and Frith 2006), which suggests a processing style focused on local perceptual features and a diminished tendency to integrate perceptual features into a coherent whole. The idea that we suggest is that differences in expected precisions is the mechanism behind weak central coherence, and that it is able to explain the varied landscape of enhanced and diminished ability in ASD, which the weak central coherence account cannot so easily accommodate. (The proposal is also related to ideas from (Mitchell and Ropar 2004), from (Qian and Lipkin 2011), and from (Brock 2012; Pellicano and Burr 2012; Friston, Lawson et al. 2013).

The proposal is new and evidence is needed to substantiate it. Our own research is providing data that is consistent with it, in the context of sensorimotor processing and multisensory integration. Our key model is the rubber hand illusion, which has all the required elements to trigger differences in expected precisions. The rubber hand illusion occurs when a visible rubber hand and one's own hidden real hand are touched in synchrony, giving rise to the startling experience that the touch one can feel is located on the rubber hand (Botvinick and Cohen 1998). We assess the varying effects of this illusion on proprioception (perceived arm position), and also introduce a reach-to-grasp task after experiencing the illusion, which must then be performed under the uncertainty-inducing context of the rubber hand being experienced as the locus of touch.

We find that patients with ASD differ from controls, and that individuals with ASD-like traits differ from those low on ASD-like traits. Specifically, participants with ASD and ASD-like traits have more accurate proprioception, suggesting they do not integrate under a more global model, which would pull their proprioceptive estimate towards the (illuded) visuotactile estimate. This is consistent with an upregulation of bottom-up sensory estimates regarding arm position due to higher expectations for precision in sensory input compared to the control groups. Moreover, people low on ASD-like traits reach with much tentativeness and uncertainty after experiencing the illusion, which is not seen in individuals with high ASD-like traits, suggesting that the latter group expect more precision in the proprioceptive and kinestestic input they will receive as movement unfolds (Paton, Hohwy et al. 2011; Palmer, Paton et al. 2013).

This idea is also worth pursuing as the variability of findings for ASD in the sensory domain may be better explained by appealing to differences in the presence and absence of uncertainty-inducing contexts in specific experimental set-ups.

The question we wish to address now is whether expectations for high precision of sensory input would cause the kinds of challenges we have outlined for common knowledge, and thereby on social cognition.

It seems clear that someone with expectations for high precision will present differently in scenarios that invoke versions of the Byzantine generals problem. Under conditions of uncertainty (i.e., not a beacon on a hill but a more subtle signal), people with higher expectations for precision should trust the signal more and sample the signal for longer in order to arrive at the expected precise estimate. People who expect less sensory precision should be quicker to appeal to prior expectations (e.g., rely on known alignment) to overcome uncertainty, and should not sample for as long. This should manifest such that those expecting precision will sometimes act on a misinterpreted signal because they trust it more than the context mandates they should (compare: reach less tentatively and more smoothly), and might fail to a larger degree to integrate the signal under a model of (aligned) mental states of the sender; alternatively, they may sample for longer than neurotypical collaborators and thus not act when everyone else is acting – missing the boat and failing to learn common knowledge truths.

At the outset we noted that social cognition is special because it involves meta-cognition, that is, representation of others mental representation of one's own and other's mental states. This occurs not only in one's attempt at representing other's mental states but also in extracting information about the shape of social networks and who is telling what to whom (e.g., "is this a *square* or a *kite*?"). This was noted to be a special kind of non-linear, causal interaction, and was identified as a requirement for common knowledge. Non-linearity is what introduces ambiguity in causal inference because it makes it difficult to match cause and effect in one-one relations. Multiple, nested levels of non-linearity is then especially difficult to deal with and requires especially well-honed balance of trusting the signal and relying on prior knowledge. In other words, we expect meta-mentalising to be especially challenged when expectations for precision are not optimized. Specifically, expecting too much precision means being more stuck in low-level signal processing and less inclined to fit represented causes in with more global models. This would predict that highly interacting causes are missed, in particular those that relate to meta-mentalising.

This overall picture of lessened representation of high-level, interacting causes would then cascade to other areas. For example, if meta-mentalising is less prevalent, then there will be less inclination to offer information about one's own threshold, which could feed into other's model of oneself. This impedes the reciprocality that we saw was often needed to take initiative and get a bandwagon rolling. Likewise, we can expect such problems to cascade into lessened alignment, and reduced concern about being aligned with others. The result of these mechanisms is that not only does the person with expectations for high precision in sensory estimation fail to represent other's mental states with much depth, they also will tend to fail to be able to learn, and they will be marginalized in common knowledge efforts.

It thus seems to us that the quite simple proposal that individuals with ASD have problems with optimizing their expected precisions quite quickly can cause profound and widespread problems in common knowledge, with wide ramifications for social cognition at large.

Compared to typically developing children, those with ASD tend to show developmental delays on tasks designed to test for the basic ability to attribute mental states (reviewed in Happé 1995). Many individuals with ASD, however, especially older children and adults, are able to pass the classic tests of this faculty, instead showing more subtle behavioural and neurophysiological differences in tasks that have been suggested to more specifically elicit automatic mental-state attribution, rather than allowing for inference via explicit reasoning or other strategies (Klin 2000; Castelli, Frith et al. 2002; Senju, Southgate et al. 2009). It has thus been proposed that a deficit in the automatic and intuitive ability to attribute mental states can be compensated for, just not to the extent that everyday social difficulties can be avoided (Happé 1995; Frith 2004). Understanding mentalising with respect to coordination problems and differences in expected precisions may therefore be useful in characterising the extent to which individuals can compensate, or fail to compensate, for deficits in automatic processes involved in mental state attribution.

We also noted at the beginning that mentalising seems like a special kind of causal inference because it concerns a domain that is relatively evidentially insulated. This also relates to expected precisions. In a domain that is highly influenced by non-linearly interacting causes it is crucial to have tools for compensating for uncertainty and ambiguity. One such tool is to recruit other, conditionally independent sources of evidence. This is, as we noted, seen in the courtroom analogy where a few additional witnesses can resolve uncertainty about a particular witness report. Without access to further witnesses the court must resolve to either increase sampling from the same, one witness (interrogating more), or rely more heavily upon prior conceptions (the witness is already known to be unreliable, for example). In the sensory, and social, case it is hard to see how conditionally independent sources of evidence could be recruited. Therefore the brain must resort to the latter two strategies, which of course both pertain to expected precision as we have explicated the notion above. Evidential insulation of mental states therefore compounds the problems that may arise for those with suboptimal expectations for precisions in the social areas where optimal expectations are most needed.

We will end this section by noting how the proposed differences in expected precisions could dynamically impact on social interaction. This stems from the trivial observation that communication is a ‘two-way street’ where the quality and quantity of an individual’s participation depends on what the interlocutors offer up. On our proposal, we predict restricted messaging *to* individuals with ASD from other people. Common knowledge depends on everyone knowing what messages were received by whom and how they were interpreted. If interlocutors can see that some participant is consistently not paying attention (e.g., literally out of step in ritual dance; or engulfed in increased sensory sampling), then it may not be worth sending messages to that person. Thus Schelling's unique signal may become less and less available to individuals with ASD because the rest of us are less inclined to include them. At the same time, there may be restricted messaging *from* individuals with ASD to other people. If such an individual does not engage in much meta-mentalising and does not represent social networks and reciprocal communication channels correctly, then they will be less inclined to divulge information about their thresholds in the right way in the right circumstances, and then they will be gradually dealt out of common knowledge generation (treated as the tail of kites, or as one-person cliques).

Once upon a time ASD was explained with the sexist and now entirely discredited cold mother hypothesis, namely that it was caused by emotionally cold mothering. With our proposal comes a different kind of social interaction model, where a simple deficit in expected precisions leads to a cold social network, where fewer messages are being communicated both ways, and where people with ASD are increasingly in danger of being marginalized.

6. Concluding remarks

The agenda in this chapter has been to throw light on the notion of social cognition by aligning it with causal inference in general and common knowledge in particular. We have used ASD as a test case to bring out how basic, simple differences in the optimization of expectations of the precision of sensory input could challenge common knowledge and thereby social cognition in ASD.

This differs from many other accounts of ASD because we do not think people with ASD have a specific inability to represent mental states of other people. The problem does not arise because those states of the world are mental. It arises instead because the causal inference required to extract these causes, from the sensory input one receives, is especially sensitive to exquisite optimization of expected precisions. This has to do with the requirement to meta-mentalise to engage in common knowledge exchanges and the fact that uncertainty in mentalising cannot be resolved easily by appeal to conditionally independent sources of evidence.

If this kind of challenge to causal inference in the social domain occurs early in developmental processes, then it is possible that a deep-seated and incorrigible deficit in mentalising ensues – a profound mindblindness that impedes language learning and many other social aspects of normal life. But in principle, people with ASD should be able to represent mental states, since they are just causes in the world, on a par with other, non-mental causes, which they are able to discern. One interesting possibility here is what happens if people with similar, skewed expectations for precisions communicate with each other (on internet forums, for example). We expect that mentalising will be more likely when people share expected precisions and also that it will be easier for common knowledge to arise because problems like the Byzantine generals problem, the bandwagon issues, and the cold social network issues we have discussed all to some degree depend on people having differences in such expectations and misaligned priors.

Lastly, there is a clear program here for further, empirical study on two fronts. Firstly, one could study whether it is the case that people with ASD and high on ASD-like traits do expect more precision in their sensory input, how this plays out in sensory contexts under differing levels of uncertainty and how this may impact on social cognition. Secondly, one could study whether common knowledge is a main contributor to social cognition, and whether it is especially challenged in ASD, in the ways that we have outlined. Such empirical study could look at, for example, the relative uptake of common knowledge objects like Listerine and Macintosh computers; studies could focus on the relative participation in common knowledge activities such as revolts, and compare this with pure preference based activities; and studies could focus on participation and performance in collaborative games such as the stag hunt (see Yoshida, Dziobek et al. 2010). Finally, studies could investigate whether intra-autistic communication in fact has improved social cognition, mentalising and common knowledge, but perhaps with a different timbre, scope and depth than that seen in the general population.

al-Haytham, I. A. (ca. 1030; 1989). The optics of Ibn al-Haytham.

Botvinick, M. and J. Cohen (1998). "Rubber hands `feel' touch that eyes see." Nature **391**(6669): 756-756.

Brock, J. (2012). "Alternative Bayesian accounts of autistic perception: comment on Pellicano and Burr." Trends Cogn Sci **16**(12): 573-574; author reply 574-575.

Brown, H., K. J. Friston, et al. (2011). "Active inference, attention and motor preparation." Frontiers in Psychology **2**.

Castelli, F., C. Frith, et al. (2002). "Autism, Asperger syndrome and brain mechanisms for the attribution of mental states to animated shapes." Brain **125**(Pt 8): 1839-1849.

Chwe, M. S.-Y. (2000). "Communication and Coordination in Social Networks." The Review of Economic Studies **67**(1): 1-16.

Chwe, M. S.-Y. (2001). Rational Ritual: Culture, Coordination, and Common Knowledge, Princeton University Press.

Chwe, M. S. Y. (1999). "Structure and Strategy in Collective Action." The American Journal of Sociology **105**(1): 128-156.

Dayan, P., G. E. Hinton, et al. (1995). "The Helmholtz machine." Neural Comput. **7**(5): 889-904.

Feldman, H. and K. Friston (2010). "Attention, uncertainty and free-energy." Frontiers in Human Neuroscience **4**(215).

Friston, K. (2010). "The free-energy principle: a unified brain theory?" Nat Rev Neurosci **11**: 127-138.

Friston, K., J. Mattout, et al. (2011). "Action understanding and active inference." Biological Cybernetics **104**(1): 137-160.

Friston, K. and K. Stephan (2007). "Free energy and the brain." Synthese **159**(3): 417-458.

Friston, K. J., R. Lawson, et al. (2013). "On hyperpriors and hypopriors: comment on Pellicano and Burr." Trends Cogn Sci **17**(1): 1.

Frith, U. (1989). Autism: Explaining the Enigma. Oxford, Blackwell.

Frith, U. (2004). "Emanuel Miller lecture: confusions and controversies about Asperger syndrome." J Child Psychol Psychiatry **45**(4): 672-686.

Gregory, R. L. (1980). "Perceptions as hypotheses." Phil. Trans. R. Soc. Lond., Series B, Biological Sciences **290**(1038): 181-197.

Happé, F. and U. Frith (2006). "The weak coherence account: detail-focused cognitive style in autism spectrum disorders." J Autism Dev Disord **36**(1): 5-25.

Happé, F. G. (1995). "The role of age and verbal ability in the theory of mind task performance of subjects with autism." Child Dev **66**(3): 843-855.

Helmholtz, H. v. (1867). Handbuch der Physiologishen Optik. Leipzig, Leopold Voss.

Hohwy, J. (2013 (forthcoming)). The Predictive Mind, Oxford University Press.

Huerta, M., S. L. Bishop, et al. (2012). "Application of DSM-5 criteria for autism spectrum disorder to three samples of children with DSM-IV diagnoses of pervasive developmental disorders." Am J Psychiatry **169**(10): 1056-1064.

Kersten, D., P. Mamassian, et al. (2004). "Object perception as bayesian inference." Annual Review of Psychology **55**(1): 271-304.

Kilner, J., K. Friston, et al. (2007). "Predictive coding: an account of the mirror neuron system." Cognitive Processing **8**(3): 159-166.

Klin, A. (2000). "Attributing social meaning to ambiguous visual stimuli in higher-functioning autism and Asperger syndrome: The Social Attribution Task." J Child Psychol Psychiatry **41**(7): 831-846.

Lewis, D. K. (1969). Convention: A philosophical study, Wiley-Blackwell.

Mill, J. S. (1865). An Examination of Sir William Hamilton's Philosophy. London, Longmans.

Mitchell, P. and D. Ropar (2004). "Visuo-spatial Abilities in Autism: A Review." Infant and Child Development **13**: 185-198.

Mumford, D. (1992). "On the computational architecture of the neocortex. II. The role of cortico-cortical loops." Biol Cybern **66**: 241-251.

Neisser, U. (1967). Cognitive psychology. New York, Appleton-Century-Crofts.

Palmer, C., B. Paton, et al. (2013). "Jerk differences on the autism spectrum as a sign of precision expectations."

Paton, B., J. Hohwy, et al. (2011). "The Rubber Hand Illusion Reveals Proprioceptive and Sensorimotor Differences in Autism Spectrum Disorders." Journal of Autism and Developmental Disorders: 1-14.

Pearl, J. (2000). Causality. Cambridge, Cambridge University Press.

Pellicano, E. and D. Burr (2012). "When the world becomes too real : a Bayesian explanation of autistic perception." Trends in Cognitive Sciences.

Qian, N. and R. M. Lipkin (2011). "A learning-style theory for understanding autistic behaviors." Frontiers in Human Neuroscience **5**.

Schelling, T. C. (1960). The strategy of conflict. Harvard., Mass., Harvard University Press.

Senju, A., V. Southgate, et al. (2009). "Mindblind eyes: an absence of spontaneous theory of mind in Asperger syndrome." Science **325**(5942): 883-885.

Vanderschraaf, P. and G. Sillari (2009). Common Knowledge. The Stanford Encyclopedia of Philosophy. E. N. Zalta.

Wentzer, T. and C. D. Frith (forthcoming). "Neural hermeneutics."

Wolpert, D. M., K. Doya, et al. (2003). "A unifying computational framework for motor control and social interaction." Philosophical Transactions of the Royal Society London B **358**: 593-602.

Woodward, J. (2003). Making Things Happen. New York, Oxford University Press.

Yoshida, W., I. Dziobek, et al. (2010). "Cooperation and heterogeneity of the autistic mind." J Neurosci **30**(26): 8815-8818.