Subjects as Objects

Peter Gorniak

May 15, 2003

I and I and It

We tend to put our notion of the soul, the self, the "I" in a different category than other objects we think about. This is not only a common attitude in the general population, but it is also reflected in many examples of philosophical discourse. Philosophers differ, however, in their opinions of how and why the self attains its special status. The two traditional extreme points of view, which are still commonly taken as viable positions at present, are those of Descartes and Hume [3, 7]. To caricature: in the Cartesian view the "I" is the only undoubtably existing object of thought, whereas Hume denies that the "I" is an object of thought at all. Under the influence of these two views, or perhaps sometimes independently arriving at similar ones, philosophers have tried to defend some version of the special status of the "I", or alternatively denied it its special status or its very existence.

In this paper, I will discuss some particular lines of argument that follow the views of Descartes and Hume, namely those of Wittgenstein and Kripke [9, 18]. In the process, I will argue that the two original views and their followers commit what Smith calls an *inscription error* [15] . By assuming a world in which objects are given to us as objects, they assume a difference between thinking about everyday objects and thinking about the "I" that is an illusory difference. I will suggest we should treat both everyday objects and the self as painstakingly constructed scaffoldings instead of cleanly provided entities. Under this treatment the similarities between the self and other objects are highlighted, and we can begin to intelligibly talk about differences as far as they exist. I will then proceed to apply this treatment to the problem of other minds, and conclude with an examination of its relation to Descartes' views. I am throughout using only a few representative examples of inscription errors of the Humean and Cartesian kind, but I believe with Smith that they are rather pervasive throughout Philosophy.

Against Given Objects

In the Humean view of the self, the problem of the self as an object of thought arises because this object is not "given in experience". Both Wittgenstein and Kripke seem to entertain related notions. For example, Wittgenstein asks "Where *in* the world is a metaphysical subject to be found?", and Kripke agrees that Wittgenstein is "under the influence of characteristically Humean ideas". Contrasting this with the Cartesian view, Kripke gives the example that "where Descartes would have said that I am certain that "I have a tickle", the only thing Hume is aware of is the tickle itself. The self - the Cartesian ego - is an entity which is wholly mysterious".

At the basis of the Humean stance towards the self lies a conflation of the process of perception and the process of constructing objects from perception. The process of perception is the physiologically better understood of the two processes. Cognitive Scientists have a relatively detailed story to tell about the process by which, for example, visual information is processed by the first few layers of neurons behind the eye. Such research shows that there the brain employs several sophisticated visual processing stages to locate features in the visual field. Similar stories can be told for other modalities. Furthermore, researchers have given plausible accounts of some aspects of cross-modal integration that explain how different perceptive modalities as well as actions by the subject might be correlated to establish modality-independent mental models [6, 14]. Such accounts might explain, for example, how one might locate perceptive features such as sights and sounds, as well as actions such as grasping behaviours and self-motion in space relative to one's own body.

The crucial element of such stories that relate perception of features to the perception of objects is that much complicated machinery is needed to construct the notion of an object by correlating modalities, constructing representations and bundling features. Whether evolution supplies human beings with this machinery itself or only with the learning mechanisms to create this machinery during their lifetimes is not at issue. Like Millikan indicates in her thoughts about substance concepts, such an evolutionary story needs to be told, but not here [11]. However, the list of modelling mechanisms necessary to create full human concepts of objects does not stop here. Prinz lists more crucial elements of such a list, including the ability to perform a causal analysis of constructed models [13]. It is perhaps in such causal analysis that the difference between perception of features, model and underlying structure becomes the clearest. Let us take Pearl's concrete algorithms for building causal models from data as an example [12]. They work from sets of observations, provided in the human case by the moderately reliable perceptual mechanisms hinted at above. Using correlations in these data, they hypothesize a model of a causal structure producing these observations. This structure goes beyond the observations themselves, as it uses the data to determine the direction of causality where possible, and proposes possible hidden causes at specific points in the structure. These algorithms work under the assumption that there is a causal structure in the world, underlying the observations, and we can thus model it based on the observations it produces.

A similar model construction account can be given of models that are closer to the perceptive front end, such as visual models. A visual model must assume that there are features the visual system can more or less reliably detect, and that certain relations between features hold in the world (e.g. that lines delimit regions, or that regions tend to move together). Using these assumptions, it can build its own model of visual regions, proposing a feature bundle, a visual proto-object that allows for visual similarity judgements and tracking. Note that just like a hidden variable in the causal model case, such a proto-object is purely a mentally hypothesized construct that is only 'given' to us in the sense that most human beings are born with or develop machinery that makes such constructive endeavours seem effortless and instantaneous.

Structured models of phenomena like Pearl's causal models or visual tracking models explicitly or implicitly posit hidden causes. That is, in bundling features for visual proto-object tracking, and in suggesting the existence of hidden causes, the modelling machinery makes the assumption that the underlying structure of the world is such that thinking about it as containing objects is useful. At a high level, a progression of models is exhibited by children's claims about essences of animals. They go from thinking that outwards appearance determines the species of animal, to thinking that inner material (bones, blood, etc.) constitutes species, to thinking that origin (species of parents) constitutes species [8]. Similarly, I would argue, do we posit hidden essences to think about all objects we construct from the world. How do I know that there is a ripe tomato before me? My feature detectors are responding in certain ways that are best predicted by my mental model of what it takes to be a tomato, thus I place an object with a hidden tomato essence in my model of the world.

Necessary Subjects

My revision of the story of how objects are given to us on the surface seems largely compatible with the Humean complaint about the absence of the "I" in perception. This compatibility, however, is an illusion. Modelling the structure of the world with a structured model requires placing oneself as an object in the model. In essence, it requires modelling the self. As always, such models of the self can be more or less complete, diverse and detailed. Furthermore, they can be implicit or explicit. The implicit version of a self model seems to be what WIttgenstein is referring to when he says that the self is the limit of the world. Early visual processing, for example, only implicitly models the self in that it relies on specific types of outputs from earlier layers (e.g. cones in the eye), makes assumptions about orientation of the visual field (and thus the visual self) and which types of features might be important to the self (vertical lines being emphasized, indicating perhaps an easy way to find beings similar to ourselves in the visual field at later stages of processing). Other models employed by the mind, however, explicitly model the self. For example, Evans, and following him in more detail Grush, point out that locating objects in space requires representing ego- and allocentric space, at least the latter locating the self as an object in space [4, 6]. Prinz's Proxytype theory presents one way in which spatial, causal and other mental models may relate to perception and fit together into a full account of mental representation [13].

Wittgenstein's analogy between the self and perception, and the eye and visual perception is a good one. However, it is not true that visual perception does not let us construct a model of the eye. It is not hard to imagine that experiments one can perform on one's visual system, such as mapping the visual field, determining focusing behaviours, depth of field, colour biases would let one propose the existence of an eye together with a rather detailed model of possible mechanisms and designs for an eye without requiring that one ever visually perceives the eye itself. In fact, I would argue that without mirrors and other animals we would still be likely to establish at least a very coarse model of the eye as the apex of our visual cone, probably without filling in much of the detail that a more thorough analysis would provide. Modelling the eye in any such way is what it means for the eye to be 'given' to us as an object, more or less accurately and fully modelled.

Finally, following Dennett and McDermott, I contend that only models that include explicit models of the self give rise to conscious perception of objects [2, 10]. Being aware of objects as human beings are aware of objects requires modelling oneself as an object standing in an awareness relation to another object. The degree of awareness thus achieved depends on the modelling capacities of the self. A frog might not represent its own location in space when tracking a fly, but rather have a hardwired link from size, shape and movement of a patch on its retina to a fly-directed catching behaviour. A dog probably has a far more sophisticated model of itself that might include a simple model of its own location in space, and possibly rudimentary models of the self's affective relation to other objects, leading to simple guilt reactions. Human beings, on the other hand, currently have the fullest model of a self as a physical, mental and social object known to us. When we see a ripe tomato, we use our model representing visual, tactile, functional, causal and other expectations of tomatoes. The important point is that these models include a model of our selves that lets us locate the tomato in space relative to our own body, that let's us phrase its functional properties in terms of our needs and, most importantly, that lets us think about the relation between the tomato and us. I claim that this necessary model of the self can be constructed using the same mechanisms we use to model objects other than our selves. Grush gives a relatively detailed account of how such a model might be constructed for a spatial representation [6], whereas Gopnik e.a., Steyvers e.a. and Tenenbaum and Griffiths demonstrate the applicability of computational causal models to human causal reasoning and hint at how these techniques may be used by the subject to establish a causal notion of the self and other objects [5, 16, 17].

Is it paining, then?

Do I have a pain? Does it? Or is it paining? We are happy to ascribe features to objects if our mental models and perceptions allow us to fixate them as objects using observable features. Thus we are happy to think of tomatoes as red and rain as wet, but are hard pressed to locate an object for the feature *rain*. Candidates are either too general (the world) or to specific (clouds) to be adequately causally connected to the feature to serve as its essence. In an important sense, then, we assign features to mentally constructed objects. The question of whether we can justifiably assign a pain to an object ("it feels pain") and can call this object "I" ("I feel pain") is therefore the question of whether we can construct a mental model that allows us to coherently and usefully model an object with sensations and mental states, and whether it makes sense to call this object the self.

A pain sensation at the basic level stems from neural receptors that evolution has equipped us with, similarly to those that fire in response to a certain wavelength of light or a certain frequency of sound. In this basic signal, there is no "I", as Hume, Wittgenstein and Kripke so rightly worry. However, there is no object of thought at all in either the pain signal or the visual signal. It is only the use of this signal as a feature in the construction of higher level mental models that gives rise to objects. As I have argued, such models allow us to assign features to objects, because we use them to construct the objects. Furthermore, a human level mental object model must include the object's relation to the human's mental model of the self.

There is therefore a difference between an unaware reaction to a pain (like the human reflex case) and an awareness of a pain, which implies a mental model which likely includes the pain's location in one's spatial model of the own body, the pain's influence on one's thinking processes, and perhaps a causal model that includes the pain's relation to an external object (the table I stubbed my toe against). These models together constitute the object to which we assign the feature of pain, and we construct these models using features exactly like this pain sensation. In the pain case, the object thus modelled happens to be the self, and we call the model constructed from such sensations and perceptions that relate it to other objects the self or "I".

Is this ascription a justified one? I believe it is as justified as it can be. As far as we are willing to trust the modelling capacities we are born with and acquire, this ascription is no different than the ascription of any feature to any object. Of course, there can be more or less support for such models and subsequent feature ascriptions from observations. Support can be problematic due to noisy or sparse observations, misclassification of features, faulty models and even blatant yet evolutionary useful deceptions built into our feature detectors such as the well known visual focus and colour perception illusions. The advantage of having structured models, however, is that it allows one to analyse the degree to which observations and actions yield information about the underlying structure. In fact, they also let one design experiments to achieve a maximal increase in useful information about the relation between the model and the underlying structure, and it has been shown that human beings do design such experiments [12, 16, 5]. All in all, an ascription is only justified with respect to the ascribing subject's mental models and modelling abilities, which in turn depend on the subject's evolutionary design and lifetime learning process. One should note, however, that different phenomena do provide different levels of support in a subject's lifetime. The reason that colours are modelled as features of objects is that they are frequently useful to our detection mechanisms in most cases of object identification and provide reliable indications for object models. Along the same lines I would argue that pain is a frequently occurring and extremely useful feature of our model of self, and thus may have a more secure foundation for ascription than a summary feature like "envy".

Is there a difference between ascribing mental features and non-mental features to objects? Not in principle. Consider a feature like "nourishing" when ascribed to a ripe tomato or "racoon essence" when ascribed to an animal by a child. When only viewing the object before us, both features are hidden from us, and we only ascribe them because we have prior models of tomatoes and animals that we use to conclude the existence of hidden features. Our models lead us to beliefs about how we might test for the existence of these hidden features, leading us to confirm our detection of the object, revise it, or perhaps revise the model itself given enough evidence. Mental states when ascribed to oneself range along the same spectrum from pain, which is relatively directly perceivable, to a notion of envy, which we apply to ourselves based on a prior social and physical model of ourselves, supported by detectable features such as bodily sensations, behaviours of others and our coarse second order access to first order neural processes.

Are you in pain, now?

Applied to the problem of other minds, the argument so far leads to a claim similar to Burge's, but in a different framework and argument structure: A model of the self is necessary to build a model of other minds [1]. One of the problems causing Wittgenstein and Kripke to examine the problem of self-ascription is that of ascribing mental states to other people. In the Humean line of argument addressed above, the problem is exaggerated by not only questioning the ascription of mental states to others, but also questioning the ascription of mental states to any object at all, and thus to an "I". Above, I have argued that this ascription of mental properties to the self as an object is as legitimate as any ascription of a hidden property to any object, because all objects are mental constructs derived from observed features. The difference between ascribing mental states to oneself and ascribing mental states to external objects is that more mental features must be considered hidden for other objects than for oneself. There is a clear way in which I cannot feel someone else's pain - my brain is simply not connected to their pain receptors. This means that many of the features which are observable in the first person case are not observable in the case of an external object.

The fact that many of the features that let one build second order models of one's own mental processes are unobservable in the case of external objects makes it difficult to provide a plausible story as to how one might build a sophisticated model of other's mental processes. Yet only such a model, according to the line of argument pursued here, would allow one to attribute mental states to external objects. Note that this is not the case with the ascription of other features - in fact, it is easier to build a good model of someone else's visual appearance than of one's own, due to the fact that usually one only rarely has access to an external view of oneself, and that such external views (mirrors, cameras) are often limited in representative power due to angle, size, resolution and other constraints.

However, there is no reason not to merge the model construction argument, which has the flavour of an argument concluding the existence of other minds from the fact that they are the best explanation of observations, with an argument from analogy. The analogy is constituted by the fact that the structure of the model of mental processes one builds of one's own case can be applied to the case of external object by treating more features as hidden from observation. As every active model builder knows, finding out the structure of models from observations is hard compared to setting model parameters once the structure is known. However, deriving structure is always easier the more features are observable, so it only makes sense to use the first person model as a template for models of other objects. In fact, this strategy explains the common error of over-anthropomorphisation, where we apply our models of human mental states to objects whose observable behaviours provide worse fits (dogs) and even awful fits (falling rocks). That is not to say that to some degree we may be justified in claiming that dogs have mental states, but by starting from a model of ourselves we are likely to make mistakes as to their nature and similarity to ours. I should point out that the asymmetry between the model of the self and the model of other minds in terms of observable features provides good reason to give the model of the self a special name, "I". This is also supported by the fact that they play a different role in thinking, so that "I am in pain" and not "it is in pain" when speaking of the self.

It is also true that beyond structural derivation problems, estimation of the value of hidden features based on observable features is a hard problem, especially when observations are noisy and sparse and when most features of interest are unobservable. All three properties hold when trying to conclude others' mental features from their observable behaviours, and thus the ascription of specific mental features is well known to be error prone. However, the feature of having mental states like one's own in a more abstract sense is given much support over one's lifetime, and concluding that the correct model to use for other human beings is one that includes such features is by no means a leap of faith.

Descartes and I

So far, I have mainly addressed Humean concerns about the existence of the self. But what about Descartes' bold assertion of the thinking self as the only undoubtable object? At first glance, it seems like the object model construction story laid out here runs against the notion of assuming an "I", because it claims such an object must be constructed from observable features during evolution or life-time learning, just like any other object. However, there is a valid sense in which the primacy of the self survives in the account of objects presented here.

As was pointed out earlier, structured mental models of objects allow for object detection, tracking and behaviour towards these objects, but they do not by themselves allow the modeller to represent the object as an object. To achieve a representation of objects closer to the human one, the modeller must not only model the object, but also itself as an object as well as the relation between itself and the object. Only then can it engage in conscious thoughts about objects, representing them as causally connected to the self in certain predictive ways, and as located in space relative to one's own physical presence. In this very important sense Descartes is correct in claiming that without awareness of oneself, i.e. a mental model of oneself, a full model of objects as objects cannot be achieved.

Where Descartes make an inscription error, however, is in thinking that the “I” is a given object, just like Hume makes one in assuming that external objects are given objects. Subjects and objects must be constructed together through a second order modelling process. ‘Second order’ here refers to the fact that a first order modelling process, such as that of a frog reacting to its primitive visual model of a fly, only implicitly locates subject and object in the world. Only a second order model, such as the human observer of the frog and fly pair, or the human observer of the self-object relationship explicitly gives rise to the notion of a self and the other under discussion in both Descartes and Hume.

Summary

I have argued that Hume, and following him Wittgenstein and Kripke, make an inscription error in claiming that there is no “I” in perception, whereas they think that a story can be told about how we perceive objects other than ourselves. In contrast, following Smith I have proposed a story in which much modelling work must be performed by the subject to create a full notion of objects from perceived features. Adding to Smith’s theory, I have pointed out that even such modelling only gives rise to a very limited version of subject and object until the subject-object relation itself is modelled. In the process, the subject must build a second order model of itself as an object while at the same time building models of external objects and relating them to the self.

This model construction theory of objects eliminates the problems caused by the Humean inscription error - neither objects nor self are given in perception, both must be mentally constructed. Constructing one is no more of a problem than constructing the other, and the process gives rise to a rich and diverse notion of the "I" as a physical, mental and social being.

Applying this theory to the problem of other minds, I have claimed that there is good reason to believe that we re-use our models of self to bootstrap our models of others as mental beings, because more features are unobservable in the case of other minds, making direct structure derivation much harder. This can be contrasted with the case of visual appearances, where building models of others is much easier than building models of the self for the same reasons as in the case of mental features.

Finally, I have argued that Descartes make an inverse Humean inscription error by assuming the "I" to be given as an object, but external objects to stand in need of construction. I claim that neither self nor external objects are given, but that they must be jointly constructed to yield full human objective thinking.

References

[1] T. Burge. Reason and the first person. In Wright, Smith, and MacDonald, editors, *Knowing One's Own Mind*. Blackwell, 1998.

[2] D. Dennett. *Consciousness Explained*. Little, Brown and Company, 1992.

[3] R. Descartes. *Discourse on Method and Meditations on First Philosophy*. Hackett, 1641/1999.

[4] G. Evans. *Varieties of Reference*. Oxford University Press, Oxford, UK, 1982.

[5] A. Gopnik, C. Glymour, D. Sobel, L. Schulz, T. Kushnir, and D. Danks. A theory of causal learning in children: Causal maps and bayes nets. In *PSA Workshops*. PSA, 2002.

[6] R. Grush. Self, world and space: The meaning and mechanisms of ego- and allocentric spatial representation. *Brain and Mind*, 1:59–92, 2000.

[7] D. Hume. *A Treatise on Human Nature*. Oxford University Press, 1739-40/2000.

[8] F. Keil. *Concepts, Kinds and Cognitive Development*. Bradford Books/MIT Press, 1989.

[9] S. Kripke. *Wittgenstein on Rules and Private Language*. Harvard University Press, 1984.

[10] D. McDermott. *Mind and Mechanism.* MIT Press, Cambridge, MA, USA, 2001.

[11] R. Millikan. *On Clear and Confused Ideas: An Essay on Substance Concepts.* Cambridge University Press, Cambridge, UK, 2000.

[12] J. Pearl. *Causality.* Cambridge University Press, Cambridge, UK, 2000.

[13] J. Prinz. *Furnishing the Mind: Concepts and their Perceptual Basis.* MIT Press, Cambridge, MA, USA, 2002.

[14] D. Roy and A. Pentland. Learning words from sights and sounds: A computational model. *Cognitive Science*, 26(1):113–146, 2002.

[15] B. C. Smith. *On the Origin of Objects.* MIT Press, Cambridge, MA, USA, 1996.

[16] M. Steyvers, J. B. Tenenbaum, E. J. Wagenmakers, and B. Blum. Inferring causal networks from observations and interventions. *under revision*, 2002.

[17] J. B. Tenenbaum and T. L. Griffiths. Structure learning in human causal induction. In *Neural Information Processing Systems 14*, 2002.

[18] L. Wittgenstein. *Philosophical Investigations.* American International, 1953/1998.