Volume 10
2017

ISSN 1807-9792

abstracta

Linguagem, Mente & Ação

http://abstracta.oa.hhu.de

d|u|p

abstracta
Linguagem, Mente & Ação

ISSN 1807-9792

Volume 10
2017

abstracta
2004 – 2017

Contents

abstracta Linguagem, Mente & Ação ISSN 1807-9792 Volume 10, pp. 3–14, 2017

Causal exclusion without physical completeness and no overdetermination

Alexander Gebharter

DCLPS, Heinrich Heine University Düsseldorf
Universitätsstraße 1
40225 Düsseldorf
alexander.gebharter@gmail.com

Abstract: Hitchcock (2012) demonstrated that the validity of causal exclusion arguments as well as the plausibility of several of their premises hinges on the specific theory of causation endorsed. In this paper I show that the validity of causal exclusion arguments—if represented within the theory of causal Bayes nets the way Gebharter (2015) suggests—actually requires much weaker premises than the ones which are typically assumed. In particular, neither completeness of the physical domain nor the no overdetermination assumption are required.[1]

1 Introduction

Causal exclusion arguments (Kim, 2000, 2005) are typically used as arguments against nonreductive physicalism or as arguments for epiphenomenalism. They conclude from several premises that mental properties cannot be causally efficacious. The premises typically endorsed are the following (cf. Woodward 2015, sec. 2; Hitchcock 2012, pp. 42ff):

Distinctness: Mental properties cannot be reduced to physical properties; they are ontologically distinct.

[1]Acknowledegments: This work was supported by Deutsche Forschungsgemeinschaft (DFG), research unit Inductive Metaphysics (FOR 2495). I would like to thank Gerhard Schurz for important discussions. Thanks also to Christian J. Feldbacher-Escamilla and an anonymous referee for helpful comments on an earlier version of this paper.

Supervenience: Mental properties supervene on physical properties.[2]

Physical completeness: Every physical property has a sufficient physical cause.[3]

No overdetermination: No property has more than one sufficient cause.

In a nutshell, exclusion arguments run as follows: Let M be a mental property and let P be M's physical supervenience base. Now assume X to be a spatio-temporally distinct (mental or physical) property. Let us further assume that all three properties are instantiated. In case X is a mental property, X has a supervenience base Y which is also instantiated and fully determines X. In that case, X is instantiated because Y is instantiated and there is nothing left M could contribute to whether X occurs. In case X is a physical property, there is a sufficient physical cause Y of X. This sufficient physical cause Y is either P alone or P together with some other physical cause(s) of X. Also in that case X is instantiated because Y is instantiated; there is nothing left M could contribute to whether X occurs. Since M and X were arbitrarily chosen, the argument generalizes: There is no mental property M and no property X spatio-temporally distinct from M and its supervenience base such that M can contribute anything to whether X occurs. Hence, mental properties are causally inefficacious.

Hitchcock (2012) convincingly demonstrated that the validity of causal exclusion arguments as well as the plausibility of several of their premises hinges on the specific theory of causation endorsed. In particular, he showed that for three different theories of causation, *viz.* Laplacean causation, process causation, and difference-making causation, at least one of the premises mentioned above is not plausible. Gebharter (2015) provided a reconstruction of causal exclusion arguments within another theory of causation, *viz.* the theory of causal Bayes nets (CBNs), and proved their validity (given the reconstruction of supervenience relationships he suggested is correct). He did, however, not say anything about the status of the premises typically used in such arguments within the CBN framework. This is what I will do in this paper. After briefly introducing some basics of the theory of CBNs and presenting the reconstruction of causal exclusion arguments suggested in Gebharter (2015) (section 2), I argue that physical completeness as well as the no overdetermination assumption, which have some weak spots which could be atacked from friends of non-reductive physicalism, are not required for the argument to go through (section 3). One nicely gets the conclusion of causal exclusion arguments within a CBN framework by assuming instead the quite harmless principle that

[2]Supervenience is understood as strong supervenience here, meaning that every change in the supervening property is necessarily accompanied by a change in its supervenience base, while the supervenience base determines the supervening property (with probability 1).

[3]There are also weaker versions of the physical completeness principle which say that every physical effect has a sufficient physical cause. The difference between the two is, however, not that important for most of what I will do in this paper. Hence, I will most of the time stick to physical completeness as introduced here.

if mental properties are causally efficacious, then also their physical supervenience bases are. This result strenghtens exclusion arguments as arguments against non-reductive physicalism and as evidence for epiphenomenalism from the perspective of a CBN framework. I conclude in section 4.

2 Causal exclusion and causal Bayes nets

A CBN is a triple $\langle V, E, P\rangle$. V is a set of random variables, $G = \langle V, E\rangle$ is a directed acyclic graph, and P is a probability distribution over V. E is a set of directed edges ($\longrightarrow$) between variables in V. G's edges $X \longrightarrow Y$ are interpreted as direct causal relations w.r.t. V. The variables X at the ends of the arrows pointing at another variable Y in G are called Y's parents ($Par(Y)$). The variables Y which are connected to another variable X via a chain of arrows of the form $X \longrightarrow ... \longrightarrow Y$ are called X's descendants ($Des(X)$). CBNs are assumed to satisfy the causal Markov condition (CMC) (Spirtes 2000, p. 29):

Definition 2.1 (causal Markov condition)**.** $\langle V, E, P\rangle$ satisfies the causal Markov condition if and only if $Indep(X, V\backslash Des(X)|Par(X))$ holds for all $X \in V$.[4]

CMC generalizes the Reichenbachian insight that conditionalizing on all common causes renders two formerly correlated variables independent, while conditionalizing on a variable's direct causes renders it independent of its indirect causes (cf. Reichenbach 1956). It lies at the very heart of the theory of causal Bayes nets and establishes an intimate connection between unobservable (theoretical) causal structures and empirically accessible probability distributions. It plays an important role for formal causal reasoning, for formulating and testing of causal hypotheses, (together with other conditions) for causal discovery, and for computing the effects of interventions even if only non-experimental data is available (see, e.g., Spirtes 2000).

Whenever CMC is satisfied, our CBN's graph determines the following Markov factorization (cf. Pearl 2000, sec. 1.2.2):

$$P(X_1, ..., X_n) = \prod_{i=1}^{n} P(X_i|Par(X_i))$$

Basically all kinds of relations that produce the Markov factorization can be represented by the arrows of a CBN. Direct causation is only one of these relations. Gebharter (2015) argued that supervenience is another such relation. Whether this argumentation is correct is still debatable. For this paper, however, I will take it for granted that supervenience can be

[4] $Indep(X, Y|Z)$ stands for probabilistic independence of X on Y conditional on Z, which is defined as $P(x|y, z) = P(x|z) \vee P(y, z) = 0$ for all x, y, z. $Dep(X, Y|Z)$ stands short for dependence of X on Y given Z, which is defined as the negation of $Indep(X, Y|Z)$, i.e., as $P(x|y, z) \neq P(x|z) \wedge P(y, z) > 0$ for some x, y, z.

represented like direct causal connection within CBNs. Or in other words: The present paper investigates which typical premises of causal exclusion arguments are actually needed if the argumentation provided by Gebharter is correct. If it is correct, then direct causation as well as supervenience can be represented by the arrows of a CBN.[5] (Note that I do not want to claim that supervenience is a special form of causation; I prefer to stay neutral on this ontological question.) In the following, we will represent direct causal relations by means of single-tailed arrows, and relationships of supervenience by means of double-tailed arrows. Both kinds of arrows are assumed to technically work like ordinary single-tailed causal arrows in a CBN.

Gebharter (2015) reconstructs causal exclusion arguments with help of the CBN depicted in Figure 1. M_1, M_2 stand for mental properties, and P_1, P_2 stand for their respective physical supervenience bases. It is assumed that P_1 is P_2's sufficient physical cause. The question marks over the arrows $M_1 \longrightarrow M_2$ and $M_1 \longrightarrow P_2$ indicate that these two arrows are the ones which should be tested for causal effectiveness.

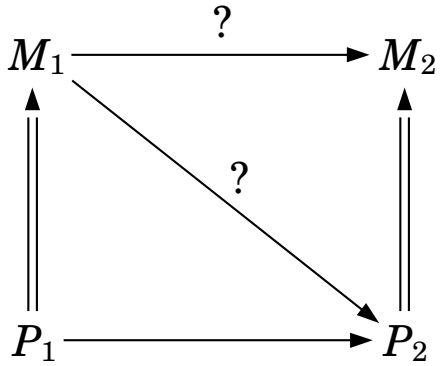

Figure 1

Note that the theory of CBNs comes with the following neat test for whether particular causal arrows can produce probabilistic dependence: To test for whether $X \longrightarrow Y$ is productive, check whether $Dep(Y, X|Par(Y)\backslash\{X\})$ holds (cf. Gebharter 2015; Schurz & Gebharter 2016). If yes, then $X \longrightarrow Y$ is productive. If no, then X cannot have a direct causal influence on Y. Informally speaking, we test for whether X can have an influence on its direct effect Y in any circumstances, i.e., in the light of any causal background story. When this test is applied

[5]Many philosophers seem to think that also another condition, *viz.* the faithfulness condition (see Spirtes 2000, p. 31), has to be satisfied. This is, however, not true. Faithfulness is a nice thing to have for many reasons, first and foremost it is essential for causal discovery. Faithfulness is, however, not necessary for representing a system's causal structure by means of a CBN. Everything needed for a CBN is that the Markov condition is satisfied.

to the causal exclusion CBN, it turns out that both arrows $M_1 \longrightarrow M_2$ and $M_1 \longrightarrow P_2$ are unproductive, meaning that M_1 is causally inefficacious w.r.t. both M_2 and P_2.

In particular, the argumentation for the unproductiveness of the arrow $M_1 \longrightarrow M_2$ runs as follows (Gebharter 2015, sec. 3): Let p_2 be an arbitrarily chosen P_2-value. Recall that M_2 supervenes on P_2. This implies that M_2's value is fully determined by P_2's value, i.e., that there is exactly one M_2-value m_2 for every P_2-value p_2 such that $P(m_2|p_2) = 1$ holds, while $P(m_2'|p_2) = 0$ holds for all $m_2' \neq m_2$. Now for every M_1-value m_1 there are two possible cases.

Case 1: m_1 and p_2 are compatible, meaning that $P(m_1, p_2) > 0$ holds. It is probabilistically valid that conditional probabilities of 1 and 0 cannot be changed when conditionalizing on compatible values of additional variables. Because of this, $P(m_2|m_1, p_2) = P(m_2|p_2) = 1$ and $P(m_2'|m_1, p_2) = P(m_2'|p_2) = 0$ will hold. Hence, no M_2-value depends on m_1 conditional on p_2.

Case 2: m_1 and p_2 are incompatible, meaning that $P(m_1, p_2) = 0$ holds. From this it follows by the definition of probabilistic independence that no M_2-value depends on m_1 conditional on p_2. Therefore, conditionalizing on p_2 renders M_2 probabilistically independent from m_1.

Recall that p_2 was arbitrarily chosen. Hence, the result obtained in both cases can be generalized: Conditionalizing on any P_2-value will render M_2 probabilistically independent from M_1, meaning that M_2 and M_1 are independent conditional on $Par(M_2)\backslash\{M_1\} = \{P_2\}$. It now follows directly from the definition of productivity that the arrow $M_1 \longrightarrow M_2$ is unproductive.

The argumentation for the unproductiveness of the arrow $M_1 \longrightarrow P_2$ runs as follows (Gebharter 2015, sec. 3): Let p_1 be an arbitrarily chosen P_1-value. Because P_1 is assumed to be P_2's sufficient cause, P_2's value is fully determined by P_1's value. Because of this for every p_1 there is exactly one p_2 such that $P(p_2|p_1) = 1$, while $P(p_2'|p_1) = 0$ for all $p_2' \neq p_2$. Now for every m_1 there are two possible cases.

Case 1: m_1 and p_1 are compatible, i.e., $P(m_1, p_1) > 0$. Since conditionalizing on compatible values of additional variables cannot have any influence on conditional probabilities of 1 and 0, conditionalizing on m_1 will not change the conditional probabilities of p_2 or p_2' given p_1, i.e., also $P(p_2|m_1, p_1) = P(p_2|p_1) = 1$ and $P(p_2'|m_1, p_1) = P(p_2'|p_1) = 0$ will hold, meaning that no P_2-value depends on m_1 conditional on p_1.

Case 2: m_1 and p_1 are incompatible, i.e., $P(m_1, p_1) = 0$. It then follows, again from the definition of probabilistic independence, that no P_2-value depends on m_1 conditional on p_1. It follows that conditionalizing on p_1 will render P_2 independent from m_1.

Since p_1 was arbitrarily chosen, the result obtained in the two cases can be generalized: Conditionalizing on any P_1-value p_1 will render P_2 independent from M_1, i.e., P_2 and M_1 are

independent conditional on $Par(P_2)\backslash\{M_1\} = \{P_1\}$. From our productivity test it follows then that the arrow $M_1 \longrightarrow P_2$ is unproductive.

3 Physical completeness and no overdetermination within the CBN framework

Gebharter's reconstruction of the exclusion argument seems to make use of all four premises introduced in section 1 (Gebharter 2015, sec. 3). Because of the distinctness premise, mental properties are represented by different variables (M_1, M_2) than the ones (P_1, P_2) used to represent their respective physical supervenience bases. The supervenience premise implies some constraints on the CBN's probability distribution, *viz.* that every change in M_i's value leads to a probability change of some P_i-value and that every P_i-value determines M_i to take a specific value with probability 1. The premise of the completeness of the physical domain implies that for every physical property represented by a variable P_i there is a sufficient physical cause, i.e., a variable P_j such that P_i is fully determined by P_j. The CBN reconstruction assumes P_1 to be such a sufficient physical cause of P_2. Finally, the no overdetermination assumption seems to be present in the productivity test applied to the causal arrows $M_1 \longrightarrow M_2$ and $M_1 \longrightarrow P_2$: M_1 is only accepted as causally efficacious if there is no systematic overdetermination, i.e., if M_1 has at least a slight influence on M_2's or on P_2's probability distribution when all parents of M_2 different from M_1 or all parents of P_2 different from M_1 are fixed to certain values.

The majority of philosophers and philosophically minded scientists seems to accept that mental properties supervene on physical properties. Every change of a decision, for example, is necessarily accompanied by changes in the brain and also fully determined (or constituted) by these changes. So the supervenience premise seems to be quite harmless and basically everyone wants to subscribe to it. Concerning the distinctness premise, I have neither any evidence for nor any intuition about whether it is true. However, if mental properties are not distinct from physical properties, then there seems to be little space for them to be autonomous in the sense the non-reductive physicalist would like them to be. And if mental properties are distinct from physical properties, then non-reductive physicalism seems to fall prey to the exclusion argument (at least within the theory of CBNs). Either way this is bad news for the supporter of non-reductive physicalism. To give non-reductive physicalism a chance, however, one has to assume distinctness. For the reasons mentioned I will leave the distinctness assumption and the supervenience premise untouched and will not discuss them in more detail in the remainder of this paper. I will rather focus on the more interesting premises which also clearly refer to causation: the physical completeness premise and the no overdetermination premise.

Let us start with a closer look at the assumption of the completeness of the physical domain. Though this premise is in principle compatible with the theory of CBNs, there are several possibilities for the non-reductive physicalist to attack it. One worry the non-reductive physicalist might have is, for example, that physical completeness is a quite strong metaphysical assumption. Why should we believe that really every physical property has a sufficient physical cause? The big bang, for example, might be an uncaused event. There are, however, weaker versions of the physical completeness premise available on the market which can avoid this worry. One might, for example, only assume that there is a sufficient physical cause for every caused physical event (cf. Esfeld 2007; Papineau 1993). This version of physical completeness would clearly allow for uncaused events like the big bang. And it would still be sufficient to run the exclusion argument within the CBN framework. If M_1 causes P_2, then P_1 is a sufficient cause of P_2 and there is no causal role left for M_1 to play.[6] But also this version as a premise seems to be quite strong. It excludes events which are only caused in a purely probabilistic way. An obvious example is the decay of uranium, which can only be probabilistically influenced. But if we have good reasons to doubt that every caused physical property has a sufficient physical cause, then Gebharter's argumentation for the unproductiveness of the arrow $M_1 \longrightarrow P_2$ does not go through (Gebharter 2015, sec. 3). If it cannot be guaranteed that P_1 fully determines P_2, then—so it seems—it might happen that P_2 still depends on M_1 when conditionalizing on P_1. In that case, the productivity test would tell us that M_1 can be causally efficacious w.r.t. P_2 and that non-reductive physicalism could—at least in principle—be saved.

I agree that Gebharter's original argument for the unproductiveness of the arrow $M_1 \longrightarrow P_2$ would be undermined if we are not allowed to assume that P_1 fully determines P_2 anymore (Gebharter 2015, sec. 3). However, there is a slightly different argument for the unproductiveness of this particular arrow that does not require P_1 to be a sufficient cause of P_2. It only requires the following as a premise instead:

> **No mental causation without physical causation:** If a mental property M is a cause of a physical property X, then also M's physical supervenience base P is a cause of X.

This assumption is weaker than the two versions of the assumption of the completeness of the physical domain mentioned above. The stronger one of the two versions of the physical completeness premise leads to infinitely many physical events in one's ontology once there is at least one such physical event: If there is a physical event e_1, then there is also e_1's sufficient physical cause e_2. But e_2's existence requires another sufficient physical cause e_3 and so on ad

[6] Note that the argumentation for the unproductiveness of the arrow $M_1 \longrightarrow P_2$ does not depend on a causal relation between P_1 and M_2 at all. For showing that this arrow is unproductive, physical completeness is, hence, not required.

infinitum. On the other hand, the no mental causation without physical causation principle stated above neither requires that all physical events are caused, nor that there are any sufficient physical causes at all. It just says that *if* there is a mental property that causes some physical property X, *then* also this mental property's supervenience base is causally relevant for X (in a deterministic or an indeterministic way). This seems to be a highly plausible assumption. It is clearly weaker than the stronger version of the premise of the completeness of the physical domain. From the pure existence of a physical event e_1 (alone) nothing follows according to the no mental causation without physical causation principle. The existence of other physical causes only follows if there are also mental causes of e_1. And even in that case these additional physical causes might be weak indeterterministic causes. Hence, the no mental causation without physical causation principle is also weaker than the weaker one of the two versions of the physical completeness premise, which only requires that caused physical events have sufficient physical causes.

Now one might think that the no mental causation without physical causation principle is, in truth, just a weaker version of the physical completeness premise. I think that the former is not just a weaker version of the latter. There is another crucial difference between the two assumptions. The mental causation without physical causation principle connects mental causation to physical causation. It says that certain phsical causal facts have to hold *if* certain mental causal facts hold. For the reductive physicalist, the principle is empty, simply because she believes that mental facts are nothing over and above physical facts. For her the principle just says that properties which have physical causes have physical causes. The physical completeness premise, on the other hand, is not empty for the reductive physicalist. For her the physical completeness premise still implies the existence of sufficient physical causes if there are any (physical) causes.

Before we go on, let me briefly illustrate the no mental causation without physical causation principle by means of Hitchcock's refrigerator example (Hitchcock 2012, p. 42): I decide to go to the refrigerator to grab something to drink. The decision is the mental event, certain changes in my brain form its physical supervenience base, and my body moving toward the refrigerator is the physical event I intend to bring about. Now let us assume that my decision causes my body to move toward the refrigerator (in a deterministic or indeterministic way). In that case—without much doubt—also the changes in my brain on which my decision supervenes will be causally relevant for my body moving toward the refrigerator. Note how weak the no mental causation without physical causation principle actually is: In case epiphenomenalism or reductionism is true, there are no mental causes (different from brain processes) and, hence, the principle keeps silent about the existence of any physical causes of my body's

moving toward the refrigerator different from mental properties. And even if there were mental causes—meaning that non-reductive physicalism were true—then the no mental causation without physical causation principle would only require that also these mental causes' physical supervenience bases are causes that make at least a slight probabilistic difference for my body's moving toward the refrigerator.

Now the assumption that there is no mental causation without physical causation is everything required to show that the arrow $M_1 \longrightarrow P_2$ is unproductive in the CBN depicted in Figure 1. In the original argument, the arrow $M_1 \longrightarrow P_2$ turned out as unproductive because P_2's parent P_1 was assumed to be a sufficient cause of P_2 and, hence, fully determined P_2's value. But if P_2's value is determined by P_1, then no change in M_1 can be associated with a change in P_2. Thus, we get the independence $Indep(P_2, M_1|P_1)$. But P_1 does not only determine P_2, but also M_1 (because M_1 supervenes on P_1). So we do not even need the arrow $P_1 \longrightarrow P_2$ to be deterministic, or, in other words: We do not even need P_1 to be a sufficient cause of P_2 to get the independence $Indep(P_2, M_1|P_1)$.

Here is the argument: Let p_1 be an arbitrarily chosen P_1-value. Due to the fact that M_1 supervenes on P_1, P_1 fully determines M_1. Hence, there is exactly one M_1-value m_1 for every P_1-value p_1 such that $P(m_1|p_1) = 1$ holds, while $P(m_1'|p_1) = 0$ holds for all $m_1' \neq m_1$. Now for every single P_2-value p_2 there are two possible cases.

Case 1: p_1 and p_2 are compatible, i.e., $P(p_1, p_2) > 0$. Because conditionalizing on compatible values of additional variables cannot have any influence on conditional probabilities of 1 and 0, also $P(m_1|p_1, p_2) = P(m_1|p_1) = 1$ and $P(m_1'|p_1, p_2) = P(m_1'|p_1) = 0$ will hold. Hence, no M_1-value depends on p_2 conditional on p_1.

Case 2: p_1 and p_2 are incompatible, meaning that $P(p_1, p_2) = 0$. From this it follows by the definition of probabilistic independence that no M_1-value depends on p_2 conditional on p_1. Therefore, conditionalizing on p_1 renders p_2 probabilistically independent from M_1.

Again, p_1 was arbitrarily chosen for both cases above. Hence, the result obtained in both cases can be generalized: Conditionalizing on any P_1-value will render M_1 probabilistically independent from P_2. This is equivalent with $Indep(P_2, M_1|Par(P_2)\backslash\{M_1\} = \{P_1\})$. From $Indep(P_2, M_1|Par(P_2)\backslash\{M_1\} = \{P_1\})$ and our productivity test it follows that M_1 cannot have any probabilistic influence on P_2 over the arrow $M_1 \longrightarrow P_2$.

As a last step, let us also take a brief look at the plausibility and the role of the no overdetermination assumption within the CBN framework. Within this framework, the no overdetermination assumption basically corresponds to assuming the causal minimality condition (cf. Spirtes 2000, p. 31), which is satisfied by a CBN if and only if every arrow of the CBN is productive (Gebharter & Schurz 2014, theorem 1). First of all, note that assuming minimal-

ity is perfectly rational from a methodological point of view: We only want to assume causal relations that are at least in principle identifyable by their empirical (probabilistic) footprints. Nevertheless, a supporter of non-reductive physicalism may, again, object that assuming no overdetermination (or minimality) for all kinds of systems is much too strong from a metaphysical point of view. I agree that this is a strong metaphysical claim and that it is—at least in principle—possible that there are causal relations out there in the world which are systematically overdetermined. Let us grant this to the non-reductive physicalist and see what it implies for the reconstruction of the exclusion argument by means of the CBN depicted in Figure 1.

The interesting thing we can learn from the CBN reconstruction is that causal efficacy and the presence of a causal relation are two slightly different things. Supporters of the causal exclusion argument may be perfectly happy with direct causal relations between M_1 and M_2 as well as P_2 as long as M_1 can be shown to be inefficacious, i.e., as long as it can be shown that these relations cannot propagate any probabilistic dependence. And this is exactly what the reconstruction suggested by Gebharter (2015) shows. It does not require the no overdetermination premise (or the assumtion of minimality) at all. The productivity test porposed can be applied to every single arrow and it can be shown that the arrows $M_1 \longrightarrow M_2$ and $M_1 \longrightarrow P_2$ are unproductive. Whether we believe in no overdetermination and take the results of our productivity test as evidence to remove the arrows or do not care about overdetermination at all and leave the arrows intact: In any case M_1 can be shown to have no direct (probabilistic) influence on M_2 or P_2 in any circumstances. In other words: Even if M_1 actually is a cause of M_2 or P_2, it is necessarily an inefficacious cause. I think that even epiphenomenalists would be happy with this particular kind of mental causation (if it deserves to be called mental causation at all).

4 Conclusion

Causal exclusion arguments typically rest on four premises which I labeled distinctness, supervenience, physical completeness, and no overdetermination in section 1. While it is uncontested that mental properties supervene on physical properties, the distinctness of mental properties and physical properties is questionable. However, for the kind of autonomy of the mental the non-reductive physicalist demands it is essential to assume the latter. In this paper I focused on the remaining two premises (physical completeness and no overdetermination), whose plausibility depends on the specific theory of causation endorsed. I argued that both premises do not stand in conflict with the theory of CBNs, but that friends of non-reductive physicalism have good reasons to not accept these two conditions. In particular, both are quite strong from a metaphysical point of view. I then took a closer look at the role of these two premises within

Gebharter's reconstruction of the exclusion argument (Gebharter, 2015). It could be shown that exclusion arguments go through with much weaker premises within a CBN framework. In particular, the no overdetermination assumption is not required at all, and the completeness of the physical domain can be replaced by a weaker and more plausible premise. This premise states that if a mental property causes a physical property, then also this mental property's physical supervenience base is causally relevant for that physical property.

All in all, the results of this paper can be seen as evidence against non-reductive physicalism from the view point of causal Bayes nets. To refute non-reductive physicalism it basically suffices to either reject that mental properties are distinct from physical properties, or to accept that mental properties supervene on physical properties and that if mental properties are causes of physical properties, then also their physical supervenience bases are. The two latter assumptions seem highly plausible.

Note that the results of this paper only hold for the reconstruction of causal exclusion arguments within the CBN framework suggested by Gebharter (2015). However, a reconstruction within the theory of CBNs seems promising for several reasons. The theory seems to give us the best grasp of causation we have so far. It allows for the development of powerful discovery algorithms, for testing causal hypotheses, and even for predicting the effects of possible interventions on the basis of purely observational data (Spirtes, 2000). The theory also behaves like a modern empirical theory of the sciences. Its core axioms can be justified by an inference to the best explanation of certain statistical phenomena and several versions of the theory can be shown to have empirical content by whose means they become testable on purely empirical grounds (cf. Schurz & Gebharter 2016).

References

Esfeld, M. (2007), 'Mental causation and the metaphysics of causation', *Erkenntnis*, **67** , 207–220.

Gebharter, A. (2015), 'Causal exclusion and causal Bayes nets', *Philosophy and Phenomenological Research*, **DOI**: 10.1111/phpr.12247.

Gebharter, A., & Schurz, G. (2014), How Occam's razor provides a neat definition of direct causation, *In* Mooij, J. M., Janzing, D. , Peters,J., Claassen, T. & Hyttinen, A.(eds.), 'Proceedings of the UAI workshop Causal Inference: Learning and Prediction', Retrieved from http://ceur-ws.org/Vol-1274/uai2014ci_paper1.pdf.

Hitchcock, C. (2012), 'Theories of causation and the causal exclusion argument', *Journal of Consciousness Studies*, **19**(5–6), 40–56.

Kim, J. (2000), *Mind is a physical world*, MIT Press, Cambridge, MA.

Kim, J. (2005), *Physicalism, or something near enough*, Princeton University Press, Princeton.

Papineau, D. (1993), *Philosophical naturalism*, Blackwell, Oxford.

Pearl, J. (2000), *Causality* (1st ed.), Cambridge University Press, Cambridge.

Reichenbach, H. (1956/1991), *The direction of time*, University of California Press, Berkeley.

Schurz, G., & Gebharter, A. (2016), 'Causality as a theoretical concept: Explanatory warrant and empirical content of the theory of causal nets', *Synthese*, **193**(4), 1073–1103.

Spirtes, P., Glymour, C., & Scheines, R. (2000), *Causation, prediction, and search* (2nd ed.), MIT Press, Cambridge, MA.

Woodward, J. (2015), 'Interventionism and causal exclusion', *Philosophy and Phenomenological Research*, **91**(2), 303–347.

abstracta Linguagem, Mente & Ação ISSN 1807-9792 Volume 10, pp. 15–34, 2017

Self-Forming Actions, Snap Decisions, and Indeterminism: A Problem for Kane's Libertarianism

Neil Campbell

Philosophy Dept., Wilfrid Laurier University
75 University Avenue W.
Waterloo ONN2L3C5, Canada
necampbell@wlu.ca

Abstract: Central to Robert Kane's libertarian account of free will is the distinction between two kinds of action: undetermined self-forming actions by means of which one shapes one's character, and those actions that are subsequently determined by one's freely formed character. Daniel Dennett challenges the coherence of this distinction, but I argue that his arguments rely on highly controversial assumptions. Building on one of Dennett's less problematic examples, I argue that many actions that appear to be determined by our character might very well turn out to be indeterministic. Since such actions occur in a way the agent seldom notices, yet meet the conditions Kane sets out for self-forming actions, they represent a problem for the degree of control agents have over their self-formation.

Robert Kane's (1998) brand of libertarianism has been extremely influential. Few theories have garnered as much attention or criticism in the contemporary debate about free will. For the most part, the criticisms of Kane's view fall into one of three groups. In the first, Kane's critics challenge the role indeterminism plays in his theory. They typically argue that indeterminism hinders rather than enhances the agent's control over her choices. This is because they believe indeterminism renders the outcome of deliberation random or arbitrary and frustrates the demand for contrastive explanations for why the agent didn't choose otherwise, despite

having the ability to do so (Mele 1999, 2005, 2006; Haji 2000; Levy 2005). The second group consists of those who challenge Kane's appeal to the dual efforts of will required for undetermined choices. Here the concern tends to be either that there is something incoherent about an agent trying to perform two incompatible actions at the same time (Clarke 1999; Allen 2005; Lemos 2011), or that dual efforts of will entail that the agent is responsible for too much, since her responsibility attaches to what she *tries to do*—not just to what she *actually does* (Levy 2005). The third group is made up of those who challenge the location of the indeterminism in Kane's theory. For reasons often associated with the first two groups of criticisms, these authors claim that the indeterminism arises too late in deliberation and argue that it should be installed earlier in the process in terms of non-actional states of the agent (Franklin 2013), the ordering of preferences (Ekstrom 2000, 2003), the weighting of considerations (Greenspan 2012), or acts of creative imagination (Pitman 2012), to name a few examples.

Although these criticisms are insightful and have pushed the discussion about libertarianism forward in helpful ways, there is an important dimension of Kane's theory that has been largely neglected. Few authors have critically examined Kane's basic distinction between two categories of action: the self-forming actions that require indeterminism, and those that are determined by the agent's freely formed character. This distinction is central to Kane's view since it allows for the possibility that undetermined choices are relatively rare, and hence, make up a small but significant group of actions. It is also fundamental to Kane's attempt to transform the free will debate from one about freedom of action into one about the possibility of self-formation. To my knowledge, only Daniel Dennett (2003) has pressed Kane on this distinction in any serious way. Like Dennett, I believe there is reason to think that the difference is not as well defined as Kane assumes. Unfortunately, Dennett's arguments rely on assumptions that are even more controversial than the distinction he attacks. In the light of this, I explore and develop a related but different line of argument in order to blur Kane's distinction. Drawing on one of Dennett's examples, I argue that some relatively innocuous assumptions about what appear to be central examples of non-self-forming actions ("snap decisions"), when coupled with the naturalistic framework of Kane's theory, imply that such choices are also undetermined. While this does not show that there is no difference between self-forming actions and actions that are determined by the agent's character, it does show that many more of our actions are undetermined than Kane suggests. More significantly, since undetermined snap decisions turn out to meet the conditions Kane sets out for self-forming actions, this introduces a problem for the kind of control agents can have over the project of shaping their characters. For now, it seems as though we can—perhaps to a quite significant extent—freely shape our characters in ways of which we are largely unaware. This is related to, but distinct from, the luck objec-

tion, since my central claim is not that the pervasiveness of undetermined choices exacerbates the problem of luck for Kane. Rather, the concern is that to the extent that such choices can happen tacitly and in the background, they normally escape our notice. This is an unwelcome result because this—rather than the indeterminism—undermines the control we have over such acts of self-formation.

My discussion is divided into three parts. In part I, I provide a brief sketch of Kane's theory and highlight his distinction between self-forming actions and those actions that are determined by one's freely formed character. In part II, I explore Dennett's attack on this distinction and explain why I find his criticisms problematic. In part III, I develop a related line of objection that is more promising and argue that it has important implications for Kane's views on the nature of self-formation.

1 Kane's Theory

Since Kane's theory is quite well known my stage setting will be relatively brief. Kane's is a variety of event-causal libertarianism that purports to be wholly naturalistic. Accordingly, he advocates only event causation and eschews any reference to agent causation or, indeed, any other "mysterious" forms of causation associated with noumenal selves or Cartesian mental substances. Kane's goal is to defend the idea that human beings can have ultimate responsibility (UR), whereby agents become the authors of their own characters, goals, and purposes. For this to be possible there must be room in the world for some degree of indeterminism within the realm of human choice. Otherwise, as Strawson (1994) claims in his 'basic argument', our characters would be determined by forces beyond our control. Kane calls the indeterministic choices by means of which we shape our own characters and thereby attain UR *self-forming actions* (SFAs)(Kane 1998, 74).

During an SFA the agent has a divided will, meaning that she has moral or prudential reasons that she endorses for two (or more) incompatible actions. Kane frequently describes this in terms of the agent being torn by two different visions of the kind of person she wants to be. Drawing on the idea that the brain employs parallel processing, Kane proposes that the competing reasons and the rationales that they support are physically realized in two corresponding neural networks that interact and compete for control of behaviour. In its bid for control, each network creates resistance to the opposing network, requiring the agent to make a dual effort of will to overcome both sources of resistance at once and to resolve the conflict. This battle for control between the two networks amplifies causal indeterminacies at the synaptic level, rendering the output of the interacting networks—the agent's choosing—undetermined, meaning that the agent could have chosen otherwise. Because the agent possesses what Kane calls "plural

voluntary control" (i.e., she endorses the reasons in favour of each choice and her choice is not compelled) the resulting action is unlike Austin-style examples where indeterminism undermines the agent's control (Kane 1998, 143). Whatever the outcome, the agent will have succeeded at doing something she was trying to do for reasons she accepts. Hence, according to Kane, it would be a mistake to think that the indeterminism involved in the process implies that the resulting action is involuntary or happens by accident.

The result of such SFAs is that they allow agents to set their wills in a certain way and thereby to secure responsibility for their own goals, purposes, or characters. When the businesswoman in Kane's favourite example resolves the conflict in her will between going to an important business meeting or stopping to help an assault victim, she shapes her character in such a way that disposes her to act a certain way in the future: selfishly or selflessly—depending on the outcome in this particular case.

Kane claims that UR is, or should be, the primary focus of the free will debate:

> "Focusing on UR helps to explain why I believe the tendency in the modern era to reduce the problem of freedom of the will to just a problem of free action is a mistake and oversimplifies the problem. Free will is not just about free action. It is about self-formation, about the formation of our wills (our characters, motives and purposes), or how we got to be the kinds of persons we are, with the wills we now have. Were we ultimately responsible to some degree for having the wills we do have, or can the sources of our wills be completely traced back to something over which we had no control, such as fate or the decrees of God, or heredity or environment, upbringing or social conditioning or hidden controllers, and so on? Therein, I believe, lies the core of the traditional problem of free will." (Kane 2014a, 40)

Once an agent is ultimately responsible, subsequent actions can be determined by the agent's freely formed character and flow from it effortlessly and without reflection. These later actions inherit their freedom from the provenance of the agent's previous SFAs.

> "It would follow [...] [from UR(N. Campbell)] that agents can act of their own free wills [...] even when their acts are determined and even when they could not have done otherwise—so long as the wills from which they act were formed by them by earlier SFAs that were not determined. Thus, incompatibilists can also say that acts done "of our own free wills" need not all be undetermined. Often we act freely and responsibly (even in an *incompatibilist* sense) out of a will already formed. What is required by UR is that we ourselves freely

> formed that will, so that it is 'our own' free will." (Kane 1998, 78 *emphasis original*)

Returning to Kane's example of the businesswoman, if she decides to stop and help the assault victim, she shapes her will in such a way that her character *determines* subsequent actions. Although she could not do otherwise on these later occasions—given the kind of person she has become—these unreflective snap decisions are still free and responsible by virtue of being products of her freely formed disposition to help others in distress.

Kane's distinction between SFAs and those actions that are subsequently determined by the agent's freely formed will (henceforth, 'non-SFAs') is obviously central. In fact, this very distinction highlights the way in which he attempts to reshape the modern free will debate from one about freedom of action into one about self-formation. When we shape ourselves via SFAs we establish dispositions—ones of our choosing—that can operate more or less habitually. These habitual actions do not shape our characters, but flow from them. Hence, in Kane's view, our primary concern in the free will debate should really be on self-formation.

2 Dennett's Criticism

Although I will focus on Dennett's attack on the distinction between SFAs and non-SFAs, it is important to recognize that this is just one part of his more extensive critique of Kane's position. Dennett also argues against Kane's theory on two other fronts. He tries to show that Kane is guilty of treating the mind as a "Cartesian Theatre" by assuming that there is a place in the brain "where it all comes together"—a faculty of practical reasoning with clear boundaries between input and output where Kane can locate the indeterminism required for SFAs (Dennett 2003, 123). Such a view, according to Dennett, is fundamentally flawed because there is no such place (Ibid.). Dennett also challenges Kane's assumption that genuine indeterminism is required for SFAs by showing that deterministic pseudo-randomness would work just as well—if not better (Ibid.). I will not address these other lines of criticism here, but will focus exclusively on Dennett's attempt to collapse Kane's distinction between SFAs and non-SFAs.[1]

[1]I will, however, hazard two brief remarks about the other two prongs of Dennett's argument, though these remarks are not intended to be decisive. I don't find Dennett's claim that Kane's view assumes a Cartesian Theatre to be very compelling. Although Kane does make reference to a "self-network", and approvingly cites work by Crick and Koch (1990) about what might unify this network, it isn't entirely clear that such a picture is offensively Cartesian (Kane 1998, 137–142). After all, this network is distributed throughout the brain, and so are those networks involved in SFAs. Are these ideas any more implausible than neuroscientific theories that *localize* particular cognitive functions to specific regions of the brain? I find it difficult to see why they would be. As for Dennett's claim that pseudo-randomness will work just as well as genuine indeterminism, this might be susceptible to the concerns I identify about his verificationism later in section II, though I will not pursue this issue here.

Before launching into his argument, Dennett raises some initial doubts about this contrast. He employs the following example: His wife asks him if he can stop by the post office on his way to work to deliver a package for her. He says:

> "I reply almost instantaneously that I can't, because then I'd be late for an appointment with a student. Did I deliberate? Did I engage in a process of practical reasoning? This is not heavy-duty moral decision-making, but this is the stuff from which moral (and immoral) lives are largely composed: hundreds of thousands of minor choice points decided with a moment's consideration, usually with the background of justification kept tacit and unarticulated."
>
> (Ibid., 115)

Dennett then adds how odd it would be for him to reply to his wife as follows:

> "Well, since you are my wife and we have solemnly promised to help each other, and since I can think of no defect or problem in your request—you haven't asked me to do something physically impossible, or illegal, or self-destructive, for instance—there is undeniably a strong case for my answering 'Yes, dear.' On the other hand, I have told a student that I would meet with him at nine-thirty, and given the traffic, honouring your request would entail standing him up for at least half an hour. I could try to call him and ask his permission to reschedule, but I might not reach him, and besides, the harder question is whether my mailing the package in so timely a manner is a sufficiently important errand to warrant inconveniencing him. My making the appointment amounted to a promise to him, though not one that couldn't be forgivably broken for cause…". (Ibid.)

Dennett then makes the following interesting observation: "It is perhaps surprising to note that all these considerations (and many more!) really did contribute *somehow* to my snap answer" (Ibid.). He concludes that "[e]ven a snap judgement can be remarkably sensitive to myriad features of my world that have conspired over time to create my current dispositional state" (Ibid., 116).

The purpose of Dennett's example is to raise one of several boundary problems for Kane's theory. Here the boundary in question is the one between SFAs and non-SFAs. Dennett proposes that in the light of his alternative response to his wife it is unclear whether one should treat such snap decisions as issuing from the same faculty of deliberation as SFAs, or from a distinct faculty, leaving the former "in reserve for heavy lifting" (Ibid.). In a preliminary way, then, Dennett's example leads one to wonder just how an SFA differs from a non-SFA. Kane's answer, of course, is that only the former requires indeterminism. It is at this point in

his discussion that Dennett sets about undermining the contrast between these two types of action in earnest. His central argument is that there is no method or test by means of which we could tell the difference between an SFA and a non-SFA. Given this, the distinction is empty and ought to be abandoned.

At first blush, there appear to be two ways in which one might come to know whether or not a particular action is an SFA: either 'from the inside'—that is, from the perspective of the agent's own phenomenology of making a choice—or 'from the outside': using an empirical test or standard. According to Dennett, neither approach is helpful.

The problem with the first is fairly obvious and, given the framework of Kane's theory, quite compelling. Consider Dennett's Luther example. When Luther said, 'Here I stand. I can do no other', Kane agrees with Dennett that we need not treat Luther's inability to do otherwise as an obstacle to regarding his action as a free and responsible one. However, Kane disagrees with Dennett about the way in which this evaluation depends on Luther's earlier actions. Unlike Dennett, Kane thinks this judgement depends on the idea that Luther's action was determined by a will of his own making via earlier SFAs. We can certainly imagine Luther engaged in an inner struggle or two, becoming the kind of man he was when he finally uttered this statement. The difficulty is that from Luther's perspective it is a complete mystery whether or not his brain exhibited the required indeterminacy when he initially struggled with his allegiance to the Church of Rome to 'become his own man'. One cannot simply introspect and observe that two neural networks are interacting in such a way that they magnify indeterminacies at the synaptic level to produce an undetermined output. So the mere fact that one is *aware* of trying to choose—even in a tortured, hand-wringing way—is not a reliable indication of an SFA.[2] Hence, unless one can learn to introspect one's own brain states in the way that Paul Churchland (1985) suggests Frank Jackson's (1982) infamous Mary can, there is simply no way for an agent like Luther to know if he is in the throes of an SFA or not.

But even if one cannot tell from the inside that a particular choice is an SFA, surely one can tell from the outside, at least in principle. With the proper imaging technology, couldn't we detect an SFA since we have some idea of what we are looking for? Dennett claims we cannot: "when we get out our supermicroscopes and look at subatomic activity in the neurons, whatever we see will be equally uninformative about SFAs" (Dennett 2003, 129). Dennett justifies this bold assertion in two ways. The first hinges on the claim that genuine randomness is indistinguishable from pseudo-randomness; the second involves an example that illustrates our apparent unwillingness to allow whatever our 'supermicroscopes' tell us to override our ordinary practices of holding one another responsible.

[2]I will say more about the reliability of our introspective evidence below.

Pseudo-randomness has the appearance of genuine randomness but is fully deterministic. It is *practically* unpredictable—except for LaPlace's demon—due to the sheer number and complexity of the determining factors involved. Genuine randomness, on the other hand, is unpredictable *in principle*. Dennett's claim that what we observe in the brain couldn't possibly give us decisive evidence of an SFA is justified primarily by our inability to distinguish genuine from pseudo-randomness. Since we lack the cognitive abilities of LaPlace's demon, we simply can't tell if what we observe in the neural networks merely *appears* to be indeterministic, due to the sheer complexity of the neural connections, or *really is* indeterministic. Since we need to differentiate between these two possibilities in order to distinguish a non-SFA from a bona fide SFA, we can't do that either. Hence, even if our brain scans revealed the parallel processing Kane describes when an agent makes a difficult choice, and even if the output of these competing networks appeared to us to be undetermined, we might simply be wrong, so our 'supermicroscopes' are entirely unhelpful.

The other part of Dennett's justification for the claim that we can't identify an SFA using empirical means involves the following example. Suppose that a well-educated person with an exemplary upbringing is brought to trial for murder. Let us assume that throughout his life he has demonstrated normal moral development, understands the difference between right and wrong, and so possesses all the usual hallmarks of a responsible agent. Suppose that his team of lawyers 'demonstrates' that although he killed the victim, his brain is incapable of generating the indeterminacies Kane claims are required for SFAs. It would follow from Kane's theory that the defendant lacks UR, and so is not responsible for *any* of his actions, including the murder for which he is on trial. Dennett offers the following commentary on this example:

> "It's a tough sell. Why should the *metaphysical* feature of Ultimate Responsibility (supposing Kane has defined a coherent possibility) count more than the macroscopic features that can be defined independently of the issue of quantum indeterminism, and that are well motivated in terms of decision-making competences that agents have or lack? Indeed, why should metaphysical Ultimate Responsibility count for anything at all? If it can't be motivated as a grounds [sic] for treating people differently, why should anyone think it is a variety of free will worth wanting?" (Ibid., 131–132)

Since, according to Dennett, we would be loath to acquit the defendant on such metaphysical grounds—or better, to allow metaphysical considerations to trump our ordinary moral (and legal) standards of responsibility—even if we could detect the presence (or absence) of the indeterminism required for SFAs, our doing so would strike us as irrelevant to the issue of the defendant's responsibility. It follows that empirical evidence in the form of either the presence

or absence of indeterminism could not convince us of the difference between an SFA and any other action.

The above considerations lead Dennett to conclude that the distinction between an SFA and a non-SFA is specious, and so we should be highly sceptical of Kane's notion of ultimate responsibility. If there simply are no SFAs, they cannot serve as the source of our freedom of will or of action. In the light of this, Dennett thinks it is a non-starter to ground our responsibility in a handful of indeterministic choices. He proposes that if we want to retain something like the concept of an SFA *qua* an important character building choice-point in an agent's history, we should regard SFAs in the same way as speciation events. A speciation event, such as the occurrence of the first mammal, is identifiable only retrospectively and involves interest-sensitive properties of offspring as opposed to essential (in this case, mammalian) characteristics since there are only minute differences between each generation. *We* decide which differences are important, and our doing so necessarily involves a certain degree of arbitrariness. Similarly, Dennett proposes that an SFA is indistinguishable from other choices when they occur. Their significance becomes apparent only after the fact. The source of this significance is not that SFAs are *intrinsically* different from neighbouring choices, but is instead a matter of the way we organize such choices into the overall narrative of a person's moral life and development.

While I am sympathetic with the initial doubts Dennett raises about the contrast between SFAs and non-SFAs, I find his supporting arguments unconvincing. They are vulnerable on two fronts. The first front takes issue with Dennett's use of a verificationist principle; the second identifies some problematic assumptions at work in his example of the defendant whose brain lacks the capacity for indeterministic processes.

As we saw above, Dennett's main criticism of Kane's SFAs is that there is no way to detect them, and hence, no way reliably to differentiate between SFAs and non-SFAs. In the light of this, we should reject the distinction. Dennett is particularly fond of this style of argument. It relies on the following verificationist principle: *where there is no detectable difference, there is no difference*. Dennett has also employed this kind of argument to undermine the coherence of qualia (Dennett 1988) and of zombies (Dennett 1995). The trouble with this kind of approach, as William Seager has pointed out in the context of Dennett's discussion of qualia, is that the verificationist principle upon which it relies is at least as contentious as the concept under attack, if not more so (Seager 1999). This case seems to be no different. I think Kane could argue that we should reject Dennett's verificationist principle on the grounds that it is implausible to deny that there are distinctions we cannot reliably detect. To suggest otherwise exaggerates the relevance of our own limited abilities. By contrast, since there are plenty of theoretical reasons for thinking there are such things as SFAs—many of which are enumerated in his discussion

of parallel processing and the sciences of complexity and chaos (Kane 1998, 9–10, 117–118, 130, 186)—Kane can make a plausible case for the claim that we are *more* justified in believing in SFAs than we are in believing in Dennett's verificationist principle. Since verificationism is itself highly contentious, I see no reason to think that our inability to distinguish between genuine and pseudo-randomness should lead us to reject the distinction between SFAs and non-SFAs. Surely there can be such a difference in the absence of our ability to recognize it, and this is all that Kane's view requires. So, if we abandon Dennett's verificationist principle, it seems that much of his argument against Kane collapses.

One could reply on Dennett's behalf here that his argument does not really require the above verificationist principle, but just the claim that we are not in a position to know whether or not there are any SFAs, whereas Kane's view requires that we do know that there are such things as SFAs. While this is a tempting response, I don't think it succeeds. Kane is content to treat the existence of neural indeterminism (and hence, of the reality of SFAs) as an open question.[3] If determinism is true or if the only indeterministic processes that exist are not appropriately located in the human brain, Kane would concede that libertarianism is a lost cause. That determinism is true, or that the brain is a deterministic system are not settled issues. Until they are, it is reasonable for Kane to rely on the possibility of indeterministic neural processes without demonstrating their actuality. But then, in order for Dennett's initial criticism to hold, he must show that the concept of an SFA is incoherent, and his verificationist principle is designed to do just that.

The second problem with Dennett's argument holds even if one were to accept his verificationist principle. His example of the well-adjusted murder defendant is, in my view, unconvincing because it involves another problematic presupposition. His treatment of the example is driven by the assumption that metaphysical and neuroscientific considerations about free will and responsibility either do not or cannot inform our moral practices and judgements. Presumably, this is because he believes that such matters are entirely pragmatic, and hence, do not require any metaphysical moorings. If that sounds too strong, it is at least clear that Dennett expects most of us to agree that the defence's claim about the accused's lack of indeterministic neural processes is *irrelevant* to the assessment of his responsibility.

The trouble is that Dennett's expectation appears to be misplaced. In some intriguing recent work (Shariff, et al. 2014; Shariff & Vohs 2014) empirical research shows that when laypersons are taught about the free will debate and the neural basis of behaviour, their views on blame and responsibility change. In particular, findings suggest that such individuals prefer

[3]That this is the case comes across particularly clearly in his remarks (Kane 2014b) about Chapter 4 of Balaguer's (2010) recent book.

a less retributive form of punishment for criminals because they come to think that agents are less responsible for what they do than they originally thought. The conclusion of one such study offers the following, which has a clear bearing the way we should evaluate Dennett's example:

> "That mere exposure to modern neuroscience can be sufficient to reduce retributivist motivations may be particularly relevant to court cases. The explicit existence of free will may be rarely debated in court, but neuroscientific evidence often is. Indeed, recent research showed that judges afforded shorter sentences to hypothetical psychopathic criminals when the description of the criminals' psychopathy included a biomechanical component, compared with when it did not (Aspinwall, et al. 2012). Our findings likewise suggest that merely presenting such a perspective may move judges and jurors toward being less punitive and less retributive in general." (Shariff, et al. 2014, 1569)

In the light of this, it is perfectly reasonable to think that the jurors in Dennett's example would allow the fact that the defendant's brain is incapable of the indeterministic processes associated with SFAs to sway their decision towards a not guilty verdict. At the very least, this possibility should not strike us as incredible or absurd. Such evidence would not exist in a vacuum, after all. If we imagine (as Dennett's example requires us to) that neuroscience has advanced to the point where there could be conclusive evidence that someone's brain lacks the indeterministic processes required for SFAs, such advances would no doubt have begun to permeate our collective understanding. Given this, we can imagine the findings described by Shariff carried to the *n*th degree—of course the jury would find the defendant not guilty (Shariff et al. 2014). This suggests there is good reason to be deeply suspicious of Dennett's treatment of this example, and hence, of his support for the claim that empirical considerations about neural indeterminism are irrelevant to matters of responsibility, and so to the presence or absence of SFAs themselves.

These two considerations go a long way towards undermining Dennett's argument. If we reject Dennett's verificationism and treat the reality of SFAs as an open question, and if we disagree with Dennett's central example, his efforts to blur the distinction between SFAs and non-SFAs are not very persuasive. Since I am nevertheless sympathetic with Dennett's goal, I would like to explore another way of applying pressure to Kane's central distinction. My approach, however, will be quite different—almost the opposite, in fact. While Dennett tried to show that there are no such things as undetermined SFAs (and hence, that the distinction between SFAs and non-SFAs is specious) I will show that many of the actions that appear not to be SFAs in fact are. This is a considerably weaker conclusion than the one Dennett attempts

to motivate, but is no less important. For I will also argue that this conclusion has important implications for Kane's conception of self-formation.

3 SFAs, Snap Decisions, and Parallel Processing

We saw earlier that, according to Kane, once an agent has formed her character by means of SFAs, her character can determine subsequent choices such that they flow effortlessly from it. Following Dennett, let's call such choices 'snap decisions' since the idea is that the course of action in such cases is determined in an instant by the agent's character, and hence, bypasses the more complex indeterministic decision-making processes required by SFAs. My aim in this section is to show that even actions that appear to be snap decisions can have all the features of SFAs, and that this has troubling implications for Kane's views on self-formation. The argument, which is speculative (though no more so than Kane's own), will proceed in several steps. At each step, I will consider possible objections to the argument.

The first step is to show that, like SFAs, snap decisions actually involve deliberation and the consideration of competing reasons, all of which the agent endorses. Fortunately, we already have reason for thinking that this is the case in the light of Dennett's example of being asked by his wife to deliver a package. Upon closer examination, what initially appeared to be a non-deliberative snap decision turned out, in fact, to be much more complex. Many considerations, for and against, were actually at play, and presumably these all involved reasons Dennett endorsed. He believed that his duties as an academic and mentor provided compelling reasons to decline his wife's request, but he also believed that his duties as a partner and a husband provided a compelling reason to grant her request and to stand up his student. Hence, even in this simple case of a snap decision, Dennett displays the kind of plural rationality Kane associates with SFAs. It is also reasonably clear that this kind of thing is not an isolated phenomenon, for it seems fairly obvious that there is something to be said for and against virtually any action. Whether we notice it or not, there is reason to believe that there is a great deal of complex information processing that goes on in such cases, and hence, that before we act we are—on some level—deliberating between alternatives on the basis of conflicting reasons that we endorse.[4]

The Kanean is sure to object here and point out that the very fact one is unaware of anything resembling the kind of deliberation that occurs when one has a 'divided will' in cases like Dennett's snap decision only serves to highlight the difference between snap decisions and

[4]To be clear, the competing reasons are not simply in competition insofar as they recommend different courses of action. According to Kane, the reasons are "incommensurable" in the sense of being "noncomparable" (Kane 1998, 167). Hence, when I describe an agent as having competing reasons, I mean that these are normally incommensurable reasons.

SFAs. Dennett wasn't torn between two different visions of himself (the doting husband versus the dependable academic), nor did he have to make an effort of will in order to decide between delivering the package and meeting his student. His choice was effortless and instantaneous, which is surely very different from Kane's example of the businesswoman, who consciously struggles to resolve the conflict in her will.

There are two points here to address: the first involves the alleged differences between the speed at which SFAs and snap decisions occur; the second appeals to introspective or phenomenological differences between the two. Let's first deal with the *speed* at which Dennett makes his decision. The fact his choice was (or appeared to be) instantaneous is hardly decisive evidence that he did not deliberate about the alternatives or that his will was not divided. While it may be that some important choices are long drawn out affairs, the speed at which they occur is irrelevant to Kane's treatment of them as SFAs. Consider the businesswoman. She doesn't spend a week, a day, or even an hour deliberating about what to do—by then the crisis would be over. She has to decide then and there whether or not to intervene. Her choice, too, happens in an instant. Hence, the duration of the deliberation does not appear to be relevant to an action's status as an SFA—at least, not if we take Kane's favourite example at face value. So, the fact that Dennett's snap decision appears to happen in an instant does not show that it did not involve a divided will in just the way that SFAs do.

What, then, of the introspective differences between Dennett's choice and the businesswoman's? Surely the latter involves the subjective sense of a dilemma—even if it is just a momentary hesitation about what to do—while the former does not. Doesn't this show that Dennett's snap decision did not involve a divided will?

In response, I point to Kane's admission that introspection doesn't tell us the whole story about free will:

> "A frequently made objection is that we are not introspectively or consciously aware of making dual efforts and performing multiple cognitive tasks in self-forming choice situations. But I am not claiming that agents are introspectively aware of making dual efforts. What persons are introspectively aware of in SFA situations is that they are trying to decide about which of two options to choose and that either choice is a difficult one because there are resistant motives pulling them in different directions that will have to be overcome, whichever choice is made." (Kane 2014a, 44)

But if Kane is willing to allow that introspection doesn't tell us the full story about what is going on in an SFA, it is unclear why he insists that one must be introspectively aware of "trying to decide about which of two options to choose"(Ibid.). After all, introspection is

poorly understood and often unreliable (Nisbett & Wilson 1977), even in the context of making a choice (Nahmias, et al. 2004, Horgan & Timmons 2011). While the fallibility of introspection does not show that what it tells us about choice and deliberation can be completely discounted, the concerns about its reliability do raise serious doubts about how much introspection reveals about SFAs. After all, as we saw Dennett argue with the Luther example, even if one *is* aware of making a difficult choice, this does not guarantee the presence of the required indeterminism. Kane makes this concession himself when he discusses the similarities and differences between his theory and Balaguer's (2010) brand of libertarianism when he states the following about his own SFAs or Balaguer's torn decisions: "for all we know, decisions that feel torn might be determined" (Kane 2014b, 53). Similarly, one's lack of awareness of a deliberative conflict does not guarantee the *absence* of an indeterministic process. Hence, it does appear to be a genuine possibility that one can be in an SFA without being aware of this fact, and hence, without being aware of a deliberative conflict. Besides, if Kane claims that we are unaware of making dual efforts of will and are unaware of the cognitive processing that is going on in such situations, since the conflict *is* a matter of parallel cognitive processing it seems reasonable to expect that we can be unaware of the conflict itself. This allows for the possibility that Dennett's seemingly instantaneous snap decision really did involve a conflict of will, but one of which he was unaware—at least until he retrospectively thought about the situation more carefully.

We now have reasons to think that even Dennett's snap decision (1) involved a conflict in his will, and (2) included a process of deliberation that involved competing reasons that Dennett endorsed, and hence, plural rationality. The next stage of my argument requires the idea that two interacting neural networks physically realize these reasons. Here what I have to say will be highly speculative, but this seems fair given that Kane's own theory involves a largely hypothetical account of the supporting neurobiology. My suggestion may seem tongue-in-cheek, but is actually quite serious: *why wouldn't distinct but interacting neural networks realize Dennett's competing reasons?* Kane's reason for saying that the competing reasons in an SFA involve distinct networks is, as far as I can tell, because the reasons involve different mental contents, and hence, different information requires processing. The point of parallel processing, as Kane characterizes it, is to allow a complex system to deal with multiple kinds of information at once. Just as the colour and shape of an object are processed by different networks, so too are competing sets of reasons. Since a snap decision like Dennett's involves different reasons and contents, it seems reasonable to expect that two competing networks will also implement Dennett's conflicting reasons.

The Kanean will probably object that this part of Kane's theory involves a certain amount of conjecture, and so isn't it fair for him simply to *stipulate* that the way the competing reasons are physically implemented in an SFA differ from the way they are implemented in a snap decision? That is, can't he simply build the distinction into the underlying neurology and deny that snap decisions involve interacting parallel neural networks?

My response is, no, he can't. If we are to accept the claim that the way the competing reasons are physically realized is different for SFAs and snap decisions, this had better be because of some interesting and relevant differences between the cases and not a matter of sheer stipulation. Thus far, if my argument has been at all plausible, the differences between the businesswoman's SFA and Dennett's snap decision have not suggested *any* reasons for thinking that the way they are physically realized *must* differ in the way Kane seems to assume. We have yet to see any significant differences between the two cases that would suggest one involves parallel processing while the other doesn't. To claim otherwise is therefore entirely ad hoc.

We are now at the third and final stage of the argument. We need reasons to think that the competition between the two neural networks that realize Dennett's competing reasons magnifies quantum indeterminacies at the synaptic level, resulting in an undetermined choice. My tactic here is similar to the one in the previous step. If we agree that an SFA and Dennett's snap decision both involve parallel processing, *why wouldn't this result in an undetermined choice?* Presumably, the magnification of the synaptic indeterminacies is a product of the relations between the competing networks as they process their respective information and compete for global control. Since the information in a snap decision is implemented in the same way as that of an SFA (or so I have claimed), there is every reason to expect that it will result in an undetermined choice too. To suggest otherwise is, once again, ad hoc. Given Kane's conjecture about the way in which conflicts of will disrupt thermodynamic equilibrium in the brain, there is every reason to suppose that this happens during snap decisions as well. After all, either the resulting indeterminism is a product of the physical system, or it isn't. If it is, and we have reason to think the system is structurally the same for SFAs as it is for snap decisions, then the latter will also result in undetermined choices; if it isn't, then the indeterminism is the product of something else that looks suspiciously like magic, and Kane can no longer claim that his theory is naturalistic.

This completes the argument. By combining Dennett's example with the same naturalistic assumptions that drive Kane's theory, I have shown that there is good reason to think that even snap decisions have the same features as undetermined SFAs. If we assume that the kind of phenomenon that Dennett describes is not an isolated one, this suggests that far more of our choices are undetermined than Kane proposes. This is a more modest conclusion than

Dennett's, for it doesn't collapse Kane's distinction between SFAs and those non-SFAs that are determined by one's freely formed character. To do that one would have to show that *all* non-SFAs have the same features as the snap decision described above. I will not attempt this.

So how serious of a problem is this weaker conclusion for Kane's theory? Since he is a libertarian and embraces the idea that some of our free choices are undetermined, couldn't he welcome this conclusion? Wouldn't *more* indeterminism be *better*? I don't think so—at least, not if he is serious about reconceiving the free will debate as one about the possibility of self-formation. The problem I see for Kane is this: if snap decisions turn out to have all the hallmarks of SFAs, then they have an equal claim to a role in self-formation. This is problematic because, as we saw, snap decisions involve deliberations that become apparent to the agent only retrospectively, if at all. At the time of their occurrence, the agent is unaware of making such choices. This is a peculiar result, for it means that we make free choices and shape our characters in ways that do not involve conscious or reflective deliberation and control. Although such choices serve the same regress-stopping function as SFAs, they appear to diminish rather than to enhance the amount of control we have over self-formation.[5]

This is quite problematic for Kane's view, for it is similar to the problem he identifies for Frankfurt's approach to free will (Frankfurt 1972, 1987). Kane describes his difference of opinion with Frankfurt as follows:

> "Many years ago, I wrote a short letter to Frankfurt which mentioned several criticisms of his view. The familiar criticism was that whether or not we were wholehearted about the will we had, rather than ambivalent, could on his view be entirely a matter of social conditioning over which we had no control. The second, less familiar criticism was this: if freedom of will is being wholehearted in your commitments to what you will, without ambivalence, then no one could ever get from ambivalence to wholeheartedness 'of their own free will'. For they wouldn't *have* free will on his view until they got there." (Kane 2014a, 38)

Frankfurt answered forthrightly by saying he believed it didn't matter how you came to have free will or wholeheartedness. It might only be because of good fortune or upbringing or accidental factors or even social conditioning. He adds:

> "And no, you can't get to free will from ambivalence of your own free will in any deeper sense of that term. All that matters is that you have free will

[5]Hence, the nature of my concern isn't one about the sheer number of SFAs or about the pervasiveness of undetermined choices, though to the extent that critics like Mele are correct that indeterminism undermines control, the ubiquity of undetermined choices would be a problem. Setting aside the effectiveness of the luck objection, my worry is that if we engage in self-forming actions about which we are unaware, it will turn out that we are freely shaping our characters in ways that lack conscious control, and this is a peculiar result.

> or wholeheartedness no matter how you got it, for it is a good which makes life go well. This honest answer did nicely frame the debate between us, as compatibilist and incompatibilist, about what acting of one's own free will might mean." (Kane 2014a, 38)

Kane, then, thinks it is a *failing* if a theory about self-formation (or having the will one wants) allows for the possibility that how one comes to be that self (or have that will) is not under one's own control. Yet this is precisely what the above argument about snap decisions suggests since the agent engages in a form of self-formation that passes largely unnoticed. Although snap decisions are not the product of Frankfurt-style controllers or social conditioning—they are undetermined, after all—the fact we are largely unaware of them surely suggests they are not under our control in the way Kane thinks most SFAs are. This conclusion threatens Kane's account of self-formation, and so he needs to address the above possibility.

As I see it, Kane has two options: he can either undermine the significance of the above argument by marginalizing the effect that undetermined snap decisions have on our control over the process of self-formation, or he can show that there are principled reasons to think that snap decisions are not undetermined and lack the general features of SFAs. In the light of Kane's reaction to Frankfurt it is difficult to see how he could adopt the first approach. It is central to his view of self-formation that we have a certain amount of control over such acts. I suspect the second is more promising, but unless Kane can show that introspection is sufficiently reliable to distinguish SFAs from snap decisions, or can muster neurological evidence to show that snap decisions are (or must be) physically realized in a different way than SFAs, there is little reason for optimism about this second approach.

In conclusion, I have argued that Kane's libertarian theory faces a puzzling problem. Drawing on Dennett's attempt to undermine Kane's distinction between SFAs and non-SFAs, I have shown that there is reason to suspect that many snap decisions will turn out to have all the same features as SFAs. Both involve a conflict of will between reasons the agent endorses, both involve parallel processing of these respective reasons, and unless Kane surrenders his naturalism, it seems that the outcome of both processes is undetermined. While this might initially appear to benefit Kane's brand of libertarianism, I have suggested that it actually has the opposite result since such snap decisions undermine the control agents have over the formation of their own characters. In the light of Kane's reaction to Frankfurt's views on self-formation, this is surely a result he cannot accept.[6]

[6]I wish to express my gratitude to two anonymous referees for their helpful and insightful comments on earlier drafts of this paper.

References

Allen, R. (2005), 'Free Will and Indeterminism: Robert Kane's Libertarianism', *Journal of Philosophical Research*, **30**, 341–355.

Aspinwall, L. G., Brown, T.R. & Tabery, J., (2012), 'The Double-Edged Sword: Does Biomechanism Increase or Decrease Judges' Sentencing of Psychopaths?', *Science*, **337**, 846–849.

Balaguer, M. (2010), *Free Will as an Open Scientific Problem*, Cambridge, MA: MIT Press.

Churchland, P. (1985), 'Reduction, Qualia, and the Direct Introspection of Brain States', *Journal of Philosophy*, **82**, 8–28.

Clarke, R. (1999), 'Free Choice, Effort, and Wanting More', *Philosophical Explorations: An International Journal for the Philosophy of Mind and Action*, **2**, 20–41.

Crick, F., & Koch, C. (1990), 'Towards a Neurobiological Theory of Consciousness', *Seminars in the neurosciences*, **4**, 263–276.

Dennett, D. C. (1988), Quining Qualia, *In* Marcel, A.J., & Bisiac, E. (eds.), 'Consciousness in Contemporary Science', Oxford: Clarendon, 381–414.

———. (1995), 'The Unimagined Preposterousness of Zombies', *Journal of Consciousness Studies*, **2**, 322–325.

———. (2003), *Freedom Evolves*, New York: Viking.

Ekstrom, L. W. (2000), *Free Will: A Philosophical Study*, Boulder, Colo.: Westview Press.

———. (2003), 'Free Will, Chance, and Mystery', *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, **113**, 153–180.

Frankfurt, H. (1972), 'Freedom of the Will and the Concept of a Person', *Journal of Philosophy*, **68**, 5-20.

———. (1987), Identification and Wholeheartedness, *In* Shoeman, F. (ed.) 'Responsibility, Character, and the Emotions', Cambridge: Cambridge University Press, 27–45.

Franklin, C. E. (2013), 'How Should Libertarians Conceive of the Location and Role of Indeterminism?', *Philosophical Explorations: An International Journal for the Philosophy of Mind and Action*, **16**, 44–58.

Greenspan, P. (2012), 'Free Will and Rational Coherency', *Nous-Supplement: Philosophical Issues*, **22**, 185–200.

Haji, I. (2000), 'Indeterminism, Explanation, and Luck', *Journal of Ethics: An International Philosophical Review*, **4**, 211–235.

Horgan, T., & Timmons, M. (2011), 'Introspection and the Phenomenology of Free Will: Problems and Prospects', *Journal of Consciousness Studies*, **18**, 180–205.

Jackson, F. (1982), 'Epiphenomenal Qualia', *Philosophical Quarterly*, **32**, 127–136.

Kane, R. (1998), *The Significance of Free Will*, New York: Oxford University Press.

———. (2014a), 'Acting 'of One's Own Free Will': Modern Reflections on an Ancient Philosophical Problem', *Proceedings of the Aristotelian Society*, **114**, 35–55.

———. (2014b), 'Torn Decisions, Luck, and Libertarian Free Will: Comments on Balaguer's Free Will as an Open Scientific Problem', *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, **169**, 51–58.

Lemos, J. (2011), 'Wanting, Willing, Trying, and Kane's Theory of Free Will', *Dialectica: International Journal of Philosophy of Knowledge*, **65**, 31–48.

Levy, N. (2005), 'Contrastive Explanations: A Dilemma for Libertarians', *Dialectica: International Journal of Philosophy of Knowledge*, **59**, 51–61.

Mele, A. R. (1999), 'Kane, Luck, and the Significance of Free Will', *Philosophical Explorations: An International Journal for the Philosophy of Mind and Action*, **2**, 96–104.

———. (2005), 'Libertarianism, Luck, and Control', *Pacific Philosophical Quarterly*, **86**, 381–407.

———. (2006), *Free Will and Luck*, Oxford: Oxford University Press.

Nahmias, E., Morris, S., & Nadelhoffer, T. (2004), 'The Phenomenology of Free Will', *Journal of Consciousness Studies*, **11**, 162–179.

Nisbett, R. E., & Wilson, T. D. (1977), 'Telling More Than We Can Know: Verbal Reports on Mental Processes', *Psychological Review*, **84**, 231–259.

Pitman, M. M. (2012), 'Freedom, Indeterminism and Imagination', *South African Journal of Philosophy*, **31**, 369–383.

Seager, W. (1999), *Theories of Consciousness: An Introduction and Assessment*, London: Routledge.

Shariff, A. F., Greene, J. D., Karremans, J. C., Luguri, J. B.,Clark, C. J., Schooler, J. W., Baumeister, R. F., & Vohs, K. D. (2014), 'Free Will and Punishment', *Psychological Science*, **25**, 1563–1570.

Shariff, A. F., & Vohs, K.D, (2014), 'The World without Free Will', *Scientific American*, **310**, 76–79.

Strawson, G. (1994), 'The Impossibility of Moral Responsibility', *Philosophical Studies*, **75**, 5–24.

abstracta Linguagem, Mente & Ação ISSN 1807-9792 Volume 10, pp. 35–44, 2017

Cultural artefacts and neglect of the materials from which they are made

Terence Rajivan Edward

School of Social Sciences,The University of Manchester
Arthur Lewis Building, Manchester
United Kingdom, M13 9PL
t.r.edward@manchester.ac.uk

Abstract: This paper discusses an explanation, offered by Tim Ingold, for why social and cultural anthropologists have so far paid little attention to the materials from which artefacts are composed. The explanation is that these anthropologists accept a certain argument. According to the argument, what an anthropologist should focus on when examining an artefact are those qualities that make it part of a culture, and this is not the materials from which the artefact is composed. I show that Ingold has not made a compelling case against this argument, but also that it is not sound.

1 Introduction

Since at least the 1980s, there have been explanations offered for why social and cultural anthropologists pay little attention to artefacts in their works (Strathern 1990: 37–38; Miller 1987: 3–4; Henare et al. 2007: 9). In his essay 'On weaving a basket', Tim Ingold provides an explanation that addresses something more specific: why such anthropologists have paid little attention to the materials from which artefacts are composed (Ingold 2000: 340). Some artefacts are made of stone, some of wood, some of metal, and so on, yet this material constitution has been neglected he says and explains why. When discussing this explanation below, I use 'anthropology' as short for social and cultural anthropology, and 'artefact' as short for material artefact.

According to Ingold, the reason why anthropologists do not focus on the material constitution of artefacts is because the vast majority of them implicitly accept a certain argument (Ingold 2000: 340). I shall present the argument as three premises and a conclusion inferred from these premises:

(1) Anthropology is the study of different cultures.

(2) If anthropology is the study of different cultures, then what an anthropologist should focus on when examining an artefact is the quality that makes it part of a culture.

(3) The materials from which an artefact is composed do not make it part of a culture.

Therefore:

(4) When examining an artefact, the materials from which it is composed are not what an anthropologist should focus on.

If most anthropologists do indeed accept this argument, then we would have an explanation for why they pay little or no attention to the materials from which artefacts are composed.[1]

At this stage, it is worth elaborating a little on premise (3). According to Ingold, previous anthropologists hold that one should not infer that an object is part of a culture purely on the basis of its material constituents. For if an object with identical material constituents came into existence without being intentionally designed, through some natural process, then it would not be part of a culture at all (Ingold 2000: 340). The relevant intention is required. An artefact counts as cultural because: the maker, or makers, had an idea of that object beforehand; and then intentionally produced an object to correspond with their idea; and this idea participates in a shared tradition.

'Previous anthropologists' here means anthropologists who wrote prior to Ingold on artefacts, or at least the vast majority of them. What previous anthropologists apparently thought is that the artefact maker combines or shapes pre-existing materials – in other words, gives them a form – to produce an object that corresponds to a design within their mind. (In this paper, I will include qualifications like 'apparently' because I have doubts about how accurate Ingold is about past history. See the appendix.) Here is Ingold attempting to capture how previous anthropologists thought:

> "For nothing about their material composition *per se* qualifies artefacts for inclusion within culture… It is the form of the artefact, not its substance, that is attributed to culture. This is why, in the extensive archaeological and anthropological literature on material culture, so little attention is paid to actual

[1] I have written of 'the quality' in premise (2), but some anthropologists may think that what makes an artefact part of a culture is a plurality of qualities. The argument can easily be reformulated to cope with this point. Premise (3) is meant to exclude the thought that the quality of being composed from such and such materials makes an artefact part of a culture.

> materials and their properties. The emphasis is almost entirely on issues of meaning and form – that is, on culture *as opposed* to materiality. Understood as a realm of discourse, meaning and value inhabiting the collective consciousness, culture is conceived to hover over the material world but not to permeate it." (Ingold 2000: 340)

I take this passage to be Ingold attributing the (1) to (4) argument above to previous anthropologists.

This argument is of interest not just as part of an explanation, but also as a justification. That is its primary interest for philosophy of social science. If the argument is sound, anthropologists would have a justification for not paying attention to the material composition of artefacts. Ingold does not think that this argument is sound. I agree, but the reason I shall present is different to Ingold's reason. I will dispute premise (2). After disputing this premise, I will consider Ingold's reason.

2 Premise (2)

Recall the content of premise (2): if anthropology is the study of different cultures, then what an anthropologist should focus on when examining an artefact is the quality that makes it part of a culture. Ingold does not dispute this premise, but I will. I will start by pointing out three inconsistencies, before making a more fundamental objection to the premise. The inconsistencies concern how premise (2) fits with the rest of Ingold's portrait of previous anthropologists, rather than whether it is simply false.

(i) Ingold attributes this premise to previous anthropologists, but he also attributes to them a certain account of what a culture is: that a culture is something within minds (or else within a group mind, if there is such a thing). He writes: "Understood as a realm of discourse, meaning and value inhabiting the collective consciousness, culture is conceived to hover over the material world but not to permeate it." (Ingold 2000: 340) The main way of thinking about culture that concerns Ingold here is as follows: a culture consists of shared ideas[2]; ideas are concepts and beliefs within minds; and minds are non-material entities or else brains (see Sperber 1996: 1; Ingold 2000: 2–3).

However, it does not seem that previous anthropologists who think in this way can consistently count artefacts as part of a culture at all. For if a culture is within minds, then every part of a culture is within minds, but artefacts are not within minds. 'Artefact' in this paper

[2]A standard point within anthropology is that cultures cannot be neatly divided off from each other, culture A consisting of one set of ideas, culture B consisting of another set, a non-identical set. But much the same points can be made while thinking that human culture consists of ideas but denying that we can neatly divide human culture into culture A, culture B, culture C, etc.

means material artefact. Even if an artefact is the result of someone trying to make an object that corresponds to an idea of theirs, the artefact itself is not within their mind. So Ingold's previous anthropologists cannot consistently endorse premise (2), because this premise assumes that an artefact can be part of a culture.

Previous anthropologists could consistently endorse this premise if they were idealists. 'Idealists' here means people who think that the world consists only of ideas and minds. For idealists, what seem to be material objects are in fact ideas. But Ingold does not portray previous anthropologists as idealists. Rather they are either dualists, who think that there is matter but minds are non-material entities, or else they are materialists, specifically ones who think that minds are brains. Either way, they cannot consistently include artefacts as part of a culture. For these anthropologists, artefacts are material entities external to minds, even if they were intentionally produced by creatures with minds and even if they entirely realize design intentions.

(ii) Ingold's previous anthropologists deny that the material constituents of an artefact are part of a culture, but they do think of the form as cultural. A further question is whether they can consistently think this. The form of an object is the way in which its material constituents are combined together, or the way they are shaped. Whether this is the result of design or not, the form of an object does not itself seem to be something within a mind. An idea you have for an object's form is something within your mind, but not the form itself that you end up giving an object, regardless of how much it realizes the idea. In light of how Ingold presents previous anthropologists and their theory of culture as within the mind, along with their background metaphysics, I suspect what he should have said is: "Previous anthropologists studied the form of an artefact to learn about the idea of the form that was in the maker's mind, because it is this idea, or some component of it, that they thought of as part of a culture." (Ingold could contest that he should have said this by giving textual evidence that previous anthropologists thought of the form of an object as itself an idea, and not merely as the realization of an idea. But he does not provide such evidence.)

(iii) There is yet another inconsistency if previous anthropologists, as Ingold portrays them, say that the form of an artefact is part of a culture. Their argument against counting the material constituents as part of a culture is that the material constituents do not in themselves make an artefact part of a culture. If one knows the material constituents of an object but nothing else, one would not have grounds for thinking that it is a cultural object. Some natural process may have caused that object to be. But an analogous argument can be made about the form. If one knows the form of a material object but nothing else, e.g. one is told by a reliable informant that it has a spiral form, one would not have grounds for thinking that the material

object is a cultural object. Snail shells, galaxies and more have form, but they are not parts of any culture, for Ingold's previous anthropologists (Ingold 2000: 340). What is missing in these cases is the relevant intention: these are not objects that have been produced in order to realize a design, a design that participates in a cultural tradition. Given their reason for discounting the material constituents of an artefact as part of a culture, Ingold's previous anthropologists should also discount forms as part of a culture.

Let us now leave aside the three inconsistencies above and turn to a more fundamental objection to premise (2). It can be stated quite briefly. Suppose that there is some quality of an artefact that does not make it part of a culture but is very useful for gaining information about something that is part of a culture. Surely an anthropologist should pay attention to this quality as well. So premise (2) is false, because it denies this. According to premise (2), an anthropologist should focus on only the quality of an artefact that makes it part of a culture. This is far too strict a doctrine.

We can apply this general point to realize that previous anthropologists, as Ingold characterizes them, have an internal reason to pay attention to the material qualities of artefacts. By an internal reason, I mean a reason that emerges from within their worldview, once purged of the premise we have found to be false.

Ingold's previous anthropologists are interested in cultures as sets of shared ideas. Now the idea of an artefact in a maker's mind can include more than just an idea of the form. It can also include an idea of the purpose and the meaning, as Ingold himself implies (Ingold 2000: 340). Here the crucial point to note is this: the idea may also include a choice of materials. For example, an artist may conceive of a stone artwork, not a wood artwork or a metal artwork. For another example, a religious group may believe that they must use a silk dress in a marriage ceremony, leading a group member to make a dress from silk. The material constitution of an artefact may therefore provide a clue about what was included in the maker's idea. Much as an anthropologist has reason to examine the form of an artefact in the hope of gaining information about the idea in the maker's mind, so they have reason to examine the material constitution in the hope of gaining such information. Of course, in both cases the anthropologist may have to ask questions to the maker, or makers, but the basic point remains. Since the maker's idea may include a specification of the materials to be used, as well as a specification of how the materials are to be combined or shaped in the final object, there is no reason for neglecting the materials from which the artefact is made.

I anticipate Ingold protesting that what has been said so far still does not allow the material constituents of artefacts to be of interest in themselves to anthropologists. They are merely of interest as a means of acquiring information about ideas, because the ultimate object of study,

a culture, is conceived of as shared ideas within minds. This protest is correct, but it is also a modification of the stance in his essay. To repeat the point I made in the preceding paragraph: we can grant much of the worldview that Ingold attributes to previous anthropologists and yet they have reason to pay attention to material constituents, contrary to the impression he gives in his essay. We do not need to radically challenge this worldview to reveal such a reason. Also note that this worldview, purged of premise (2), does not licence paying more attention to the form of an artefact, as opposed to material constituents, contrary to what Ingold says. For people who accept this worldview, the reason to pay attention to form is as a means of acquiring information about shared ideas, and this is equally a reason to pay attention to material constituents.

3 Premise (3)

It is clear that Ingold rejects the argument we have been considering, but why does he reject it? In his essay, 'On weaving a basket', he does not challenge either premise (1) or (2). The first premise is disputed elsewhere in his writings (Ingold 2008: 69), but not in the essay we are focusing on.

Recall the content of premise (3): the materials from which an artefact is composed do not make it part of a culture. At this point, it is useful to imagine previous anthropologists saying the following, even though by now it will sound very familiar: "An artefact qualifies as part of a culture not because of the materials it is composed from, rather because its form is the result of intentionally producing an object with that form, to correspond with a prior idea of the object in the maker's mind. Furthermore, this idea must be part of a shared tradition." A worry about this statement is that some artefacts count as cultural because they are made from artificial materials, such as plastic. Ingold dismisses this worry (Ingold 2000: 340). But he thinks that the statement involves a series of assumptions and he is against a number of these assumptions (Ingold 2000: 339–340, 345–347). In 'On weaving a basket', his response to the argument we have been considering is therefore to reject it because of the dubious assumptions involved in the justification of premise (3). But I think some of these assumptions can be moderated or abandoned while still endorsing this premise.

Below are some of the assumptions Ingold is against:

(a) What distinguishes an artefact from certain superficially artefact-like productions, such as a beehive, is that the artefact-maker has an idea of what they want to make beforehand and the artefact is a realization of that idea.

(b) There is a mutually-exclusive distinction between artefacts and growths. Something cannot be both.

(c) For any artefact, the maker of the artefact has an idea of the artefact's entire form in their mind beforehand and the artefact is the realization of this idea.

(d) For any artefact, there is a distinction between its form and its substance, i.e. its material constituents.

(e) A culture is something within minds (or else within a group mind, if there is such a thing). And minds here are either non-material entities or brains.

Readers are bound to wonder why Ingold thinks that the person who justifies premise (3) in the way described depends on these assumptions. He mostly leaves us to guess the answer. My guess is that Ingold thinks of this justification as rooted in a philosophy which generally comes as a whole (2000: 339). One cannot coherently accept one element of such a whole without accepting the rest, or else the deviations one can make are insignificant (2000: 344), hence he attributes all the assumptions above to his opponent. I disagree with this holism. In the essay of his that I have been focusing on, Ingold's most developed challenge is to assumption (c). In the rest of this paper, I want to consider this assumption.

Ingold challenges the assumption by offering a counterexample. His example is the coiled basket:

> "This is not the case with basketry, however, which involves the bending and interweaving of fibres that may exert a considerable resistance of their own… One could say that the form unfolds within a kind of force field, in which the weaver is caught up in a reciprocal and quite muscular dialogue with the material." (Ingold 2000: 342)

This example is very useful. But it is natural to respond to Ingold by saying that, even if the exact form was not in the basket-maker's mind beforehand, could there not have been some more vague idea of the form in their mind? After all, the basket-maker set out to make a coiled basket, and not something else, such as a necklace.

Ingold concedes that there may be a vague idea of the form beforehand (Ingold 2000: 342), but he claims that the standard view of artefacts is that the entire form of the artefact was in the maker's mind before trying to make it:

> "Now it is very often assumed, in the study of both organisms and artefacts, that to ask about the form of things is, in itself, to pose a question about design, as if the design contained a complete specification that has only to be 'written out' in the material… the artefact is supposed to pre-exist, fully represented as a 'virtual object' in the mind, even before a finger has been lifted in its construction." (Ingold 2000: 343)

I do not think that Ingold provides sufficient textual evidence for this reading of previous anthropologists. I see hardly any evidence, and I suspect that it portrays previous anthropologists as more extreme than they actually were. The claim that for any artefact, the entire form of the artefact was conceived beforehand is a very extreme claim, and I think it would be remarkable if many previous anthropologists committed themselves to it. It goes against the common experience of having to alter one's plans as one makes something. I suspect previous anthropologists were mostly committed to the moderate claim that there was some idea of the form beforehand. (See the appendix below.)

Furthermore, previous anthropologists do not need to make an assumption as strong as assumption (c) when arguing that the material constituents of an artefact do not make it part of a culture. Even if they do somehow make this assumption, they do not need to. In the case of a certain coiled basket, they can say that it counts as cultural because the intention was to make a coiled basket, which is a kind of product within this culture. It does not matter if many of the more specific details are not the result of previously conceiving these details and then shaping the material to conform to this conception. By drawing attention to and challenging a number of assumptions, Ingold makes a very valuable contribution. But at present I cannot see how to extract from his essay a compelling case against premise (3) or against the justification he associates with it, because this justification need not involve an assumption as strong as the one which he attacks.

4 Appendix: historical note

When attributing the argument to previous anthropologists, Ingold writes as if they do not explicitly make it. Rather it is affecting their minds at a level that is not fully conscious:

> "This is precisely the kind of view that lies at the back of the minds of anthropologists and archaeologists when they speak of artefacts as items of so-called 'material culture'. The last thing they mean to suggest, in resorting to this phrase, is that in the manufactured object the domains of culture and materiality somehow overlap or intermingle." (Ingold 2000: 340)

Explaining the thinking at the back of the minds of anthropologists is a risky venture, and Ingold provides almost no evidence for his reading. I am not a historian of anthropology and so I have largely proceeded without disputing his portrait of how most previous anthropologist thought, but below I wish to point out three challenges to his reading of history.

(i) The first challenge comes from the divide between British social anthropology and American cultural anthropology. Traditionally, British social anthropologists did not think of cultures as their object of study. So even if they did not count the material constituents of an

artefact as part of a culture, this would not give them a reason to ignore these constituents. Perhaps Ingold believes that they were committed to a parallel argument, where the object of study was thought of as different societies but materials were not thought of as part of any society.

(ii) When searching for anthropologists who accept the argument, it is natural to look at the writings of those cultural anthropologists who think of a culture as a set of ideas within people's minds, or who have a closely related understanding of cultures (e.g. Geertz 1973: 5; Schneider 1980: 1; Sperber 1996: 1). The second challenge comes from other understandings of what a culture is, ones which give more space for treating the material constituents of an artefact as part of a culture. For example, Malinowski divides cultures into customs and artefacts and Melville Herskovits understands culture as the man-made part of the environment (see Prinz 2016).

(iii) At least some previous anthropologists knew that artefacts are not simply realizations of a design which is fully-specified in the maker's mind beforehand. Ingold acknowledges this point. He says:

> "Effectively, the form of the basket emerges through a pattern of *skilled movement*, and it is the rhythmic repetition of that movement that gives rise to the regularity of form. This point was made long ago by Franz Boas, in his classic work *Primitive Art*." (Ingold 2000: 342)

So it does not seem that Franz Boas held that an artefact is the realization of a fully-specified design idea. But Boas was also influential and he expressed an alternative view in a classic work. There is a worry then that there would have been, or simply are, a significant number of anthropologists who were influenced by Boas and did not accept what Ingold regards as the dominant conception of an artefact.

I anticipate someone proposing that this third challenge fails to understand which period in the history of anthropology Ingold is focusing on. The proposal is that he is attributing the dominant conception of artefacts to a period which lasts at least until when he was writing but begins after the heyday of Boas's influence. However, Ingold himself provides no such specification. He actually quotes Marx and then writes as if this is how most anthropologists think (Ingold 2000: 340), which makes it seem as if he is uncovering an assumption that has been there throughout anthropology's professional history, occupying a place that lies beneath the levels where most of the changes happen. The Boas reference challenges this view.

References

Geertz, C. (1973), Thick Description: Toward an Interpretive Theory of Culture, *In* Geertz, C. (ed.) 'The Interpretation of Cultures: Selected Essays', New York: Basic Books, pp. 3–30.

Henare, A., Holbraad, M. and Wastell, S. (2007), Introduction, *In* Henare, A., Holbraad, M. & Wastell, S. (eds.), 'Thinking Through Things', London: Routledge, pp. 1–31.

Ingold, T. (2000), *The Perception of the Environment: Essays on livelihood, dwelling and skill*, London: Routledge.

Ingold, T. (2008), 'Anthropology is Not Ethnography', *Proceedings of the British Academy* **154**, 69–92.

Miller, D. (1987), *Material Culture and Mass Consumption*, Oxford: Blackwell.

Prinz, J. (2016), Culture and Cognitive Science, *In* Zalta, E. N. (ed.) *The Stanford Encyclopedia of Philosophy* Fall 2016 Edition. Accessed on 17th Feb. 2016 from: https://plato.stanford.edu/archives/fall2016/entries/culture-cogsci/

Schneider, D. (1980), *American Kinship: A Cultural Account*, 2nd Edition, Chicago: University of Chicago Press.

Sperber, D. (1996), *Rethinking Culture*, Oxford: Blackwell.

Strathern, M. (1990), Artifacts of history: events and the interpretation of images, *In* Sikkala, J. (ed.), 'Culture and History in the Pacific', Helsinki: Finnish Anthropological Society, pp. 25–44.

abstracta Linguagem, Mente & Ação ISSN 1807-9792 Volume 10, pp. 45–64, 2017

Nonconceptualism or De Re Sense? A New Reading of Kantian Intuition

Roberto Horácio de sá Pereira

Philosophy Dept.
University of Rio de Janeiro
Av. Pedro Calmon, 550
Cidade Universitária
Rio de Janeiro - RJ, 21941-901
Brasilia
robertohsp@yahoo.com.br

Abstract: The aim of this paper is to offer a critically review the recent nonconceptualist reading of the Kantian notion of sensible intuition. I raise two main objections. First, nonconceptualist readers fail to distinguish connected but different anti-intellectualist claims in the contemporary philosophy of mind and language. Second, I will argue that nonconceptual readings fail because Kantian intuitions do not possess a representational content of their own that can be veridical or falsidical in a similar way to how the content of propositional attitudes are true of false. In this paper, I will support my own reading that sensible intuition is better seen as what Evans and McDowell have called a *de re* sense, whose main characteristic is object-dependence. In this sense, Kantian sensible intuitions can be seen as a sensible *mode of donation* of objects. In my reading, the Kantian opposition between intuitions and concepts is best seen as the opposition between the objectual *de re* perception of something and the propositional *de dicto* apperception *that* something is the case rather than the opposition between nonconceptual and conceptual contents. However, if Kantian sensible intuition is not a mental state with a nonconceptual content, it is certainly in the general anti-intellectualist neighborhood.

1 Introduction

At the beginning of the recent contemporary debate over the nonconceptual content of sense perception, Kant was often regarded as advocating the side of the conceptualists. While nonconceptualists see the alleged "Kantian model of experience" as the greatest challenge to anyone claiming that sense perception possesses nonconceptual content (Gunther, 2003: 23), conceptualists such as McDowell attack nonconceptualism, referring to the alleged "Kantian insight" that conceptual capacities are supposedly required "to make it intelligible that experience is not blind" (1994: 60). Those on both sides of the controversy seem to agree that Kant was the *founding father* of conceptualism (Hanna, 2011) in the contemporary philosophy of perception. At the beginning of the controversy, the pivotal passage was Kant's famous adage that, without thoughts or concepts, sensible intuitions are blind (A51/B75). As Gunther emblematically puts it:

> "In its slogan: 'thoughts without intuitions are empty, intuitions without concepts are blind,' Kant sums up the doctrine of conceptualism. [...] According to conceptualism, no intentional content, however portentous or mundane, is a content unless it is structured by concepts that the bearer possesses." (Gunther, 2003: 1)

However, since the emergence of a series of insightful papers and books by Hanna (2005; 2006; 2008) and Allais (2009; 2010; 2012), a new trend in Kantian scholarship has begun. Hanna (2005; 2006; 2008), Allais (2009; 2010; 2012), McLear (2011), and Tolley (2012) have gathered overwhelming textual evidence and succeeded in building a strong case if not in favor of nonconceptualism, certainly in favor of an anti-intellectualist buttoned-up reading of the First *Critique*. Nevertheless, the mainstream of Kantian scholarship (Allison, 2004; 2015; Longuenesse, 1998) has never taken this new trend seriously. A few noteworthy exceptions are Gomes (2014) and Griffith (2012).

What I intend to do here is open a new battlefront in this recent debate. Even though I am on the side of the anti-intellectualist readers of Kant, I am far from being convinced that the Kantian distinction between sensible intuition and concepts corresponds to "the contemporary distinction between non-conceptual cognitions and their content, and conceptual cognitions and their content, is essentially the same as Kant's distinction between intuitions and concepts" (Hanna, 2006: 85). I have at least two reasons for this. First, nonconceptualist readers such as Hanna and Allais are running together connected but different anti-intellectualist claims. Second, and most important, Kant has never taken sensible intuition to be a mental representation with a representational content of its own that can be veridical or falsidical (independently of judgments) in a similar way to how the content of propositional attitudes are true or false.

In this paper, I will support my own anti-intellectualist reading that sensible intuition. In the reading I am proposing, Kantian intuition is better seen as what Evans and McDowell have called a *de re* sense, whose main characteristic is object-dependence. In this sense, Kantian intuitions can be seen as a sensible *mode of donation* of objects. The crucial opposition between intuitions and concepts is better seen as the opposition between the objectual *de re* perception of what appears in space and time and the propositional *de dicto* aperception *that* something is the case rather than the opposition between nonconceptual and conceptual contents. However, if Kantian sensible intuition is not a mental state with a nonconceptual content, it is certainly in the general anti-intellectualist neighborhood.

This paper is conceived in the following sections. The first is a brief section devoted to describing the present status of the debate. I argue there briefly that conceptualist readers of the Deduction confuse intentionality with objectivity. What Kant has achieved at the end of his B-Deduction is showing that categories are conditions for representing what appear *as objects*, that is, as mind-independent entities (objectivity thesis), rather than conditions for representing what appears (intentionally thesis).

The second section is devoted to distinguishing the several different anti-intellectualist claims that are bluntly brought together under the wide umbrella of "nonconceptualism." In this section it is necessary to clarify the different senses of anti-intellectualism that are run together in the secondary literature on Kant. This clarification also paves the way for my own reading in the last section of the paper.

The third section is devoted to defending my reading of Kantian sensible intuition as an unmediated relation (relational view). Rather than being representationalist (content view) in the contemporary sense of having a content that is veridical or falsidical in a similar way to how the content of propositional attitudes are true or false, Kantian representation (Vorstellung) is a relation that puts us in direct contact with objects.

The fourth and last section is devoted to defending my own reading of Kantian sensible intuitions as a sensible *mode of donation* of objects, what I call a "pre-conceptual reference without nonconceptual content." Based on Evans and McDowell's conception of a *de re* reference, and on Kant's claim of object-dependency, I propose to read Kantian anti-intellectualism as a pre-conceptual mode of the donation of objects. If I am right, Kant is less preoccupied in classifying mental states (nonconceptualism, according to Crane, 1992). Rather, his main concern is the determination of the reference of mental states: sensible intuition with consciousness is a *de re* perception of what appears in space and time, while concepts are a *de dicto* apperception that what appears in space and time is such-and-such.

2 The Current State Of The Debate

The bone of contention in the current debate over the nonconceptual content of sense perception is no longer the Kantian adage of A51/B75, but rather the core of the Transcendental Deduction where Kant describes why such a Deduction is unavoidable:

> "Objects can indeed appear to us *without necessarily having to be related to functions of the understanding.*" (A89/B122. Emphasis added)

> "Appearances could after all be so constituted that the understanding would not find them in accord with the conditions of its unity.... [and] in the succession of appearances nothing would offer itself that would furnish a rule of synthesis and thus correspond to the concept of cause and effect, so that this concept would be entirely empty, nugatory, and without significance. Appearances would nonetheless offer objects to our intuition, *for intuition by no means requires the function of thinking.*" (A90–1/B122–3. Emphasis added)

According to the conceptualist reading, Kant is suggesting a mere epistemic possibility to be eliminated later as an unreal metaphysical possibility (Gomes, 2014: 6; Griffith, 2012: 195; Grüne, 2011). In the same vein, following Henrich (1982), Allison (2004, 160) suggests that Kant is here evoking a "specter" to exorcise later, at the end of the B-Deduction. He reiterates the same reading in his recently published book (2015):

> "I refer to this possibility as a specter because its realization would result in a cognitive chaos, and I argue that the Transcendental Deduction can be regarded as Kant's attempt to exorcise it. Although this specter may call to mind the famous Cartesian specter it is significantly different from it. While the latter is at the bottom of the worry about the lack of correspondence between our experience and a mind-independent reality, the Kantian specter concerns the fit between two species of representation in the Kantian specter the problem is that nothing would be recognizable and our experience would be nothing but what William James famously referred to as 'one great booming, boozing confusion"'(Allison, 2015: 54).

In contrast, anti-intellectualists (such as myself) have taken A90–1/B122–3 as one of the best pieces of textual evidence for Kantian anti-intellectualism. We assume that Kant was alluding to a real metaphysical fact or, as I prefer to say, to an empirical fact of human and animal cognition rather than a mere epistemic possibility to be ruled out at the end of the B-Deduction. Hanna, for example, reads the passage (correctly, according to my judgment) as the Kantian statement of what Hanna calls *Priority-to-Thought*:

> "*Priority-to-thought.* Kant says that "the representation that can be given prior to all thinking is called *intuition*" (CPR B132), and all thoughts essentially involve concepts, so intuitions can be given prior to all concepts. Furthermore it is clear that this priority of intuition to thought is both cognitive and semantic. Thus an act of intuition can occur without any corresponding act of conceptualization, and also an intuition can be objectively valid independently of any concept." (Hanna, 2006: 102)

Commenting on the same passage, Allais adds:

> "*Prima facie* textual evidence against the McDowellian claim that intuition does not make an even notionally separable contribution to cognition is provided by the passages A89/B122) in which Kant simply asserts that intuition makes an independent representational contribution." (Allais, 2009: 387)

So how to settle the dispute? Taking a closer look at what Kant says in the controversial footnote Kant of § 26:

> "Space, represented **as object** (as is really required in geometry), contains more than the mere form of intuition, namely the comprehension of the manifold given in accordance with the form of sensibility in an intuitive representation, so that the form of intuition merely gives the manifold, but the formal intuition gives unity of the representation. In the Aesthetic I ascribed this unity merely to sensibility, only in order to note that it precedes all concepts, though to be sure it presupposes a synthesis, which does not belong to the senses but through which all concepts of space and time first become possible. For since through it (as the understanding determines the sensibility) space or time are first given as intuitions, the unity of this *a priori* intuition belongs to space and time, and not to the concept of the understanding (§ 24)." (B160n. Original emphasis)

As I have argued in another paper (Pereira, 2017), Kant is not saying here that categories are not conditions *for what appears* or conditions *for representing what appears* (let us call this the intentionality thesis). Instead, what he is saying is that categories are conditions for representing what appears *as objects* (in the case in point to represent the very space *as an object*). In other words, categories are conditions for representing what appears in space and space itself as mind-independent things (let us call this the objectivity thesis). The moral to be drawn is as simple as that: Conceptualist readers are confusing conditions for *intentionality* with conditions for *objectivity*.

Yet, the best evidence demonstrating that Kant meant his statements at A89/B122 and A90–1/B122–3 as suggesting a real metaphysical possibility is his Transcendental Aesthetic.

How could Kant claim therein that space and time are not discursive concepts but pure intuitions, if he did not *truly* believe that objects can appear without necessarily having to be related to functions of the understanding?

Conceptualist readers usually appeal to the footnote of B161, where Kant seemingly argues that unity of space (what he there calls formal intuition depends on categories). But what is the price to be paid if we do not distinguish the pure intuition of the Aesthetic from the formal intuition of the B-Deduction? Longuenesse (1998) is the only conceptualist/intellectualist reader coherent in this respect. She clearly sees that if we do admit pure intuitions without categories, we must reread the Transcendental Aesthetic (1998: 216). The question is why Kant did not do this rereading himself in the B-Edition.

3 The Nonconceptual Readings Of Kant

According to its standard definition, conceptualism is the claim that mental states only possess a representational content when the subject of them possesses the required concepts to specify canonically the putative content that the mental state is representing (Bermúdez, 1998). In contrast, according to its standard definition, nonconceptualism is the opposite claim that a creature's mental state may have content even when she lacks the required concepts to specify whatever she is representing.

One important distinction in the contemporary debate that is relevant to my discussion of Kant is the distinction between "state" and "content" nonconceptualism (Heck, 2000). According to content nonconceptualism, the content of conceptual contents is composed of concepts, while the nonconceptual content is fundamentally different in the negative sense of not being conceptually structured. Therefore, one cannot represent the same content conceptually and nonconceptually. In contrast, according to the so-called state nonconceptualism, one can represent the very same content conceptually and non-conceptually because what matters is not the content but rather how the content is represented by both states. A state is nonconceptual when the subject in that state does not need to possess the required concepts to specify whatever the state represents.

This opposition between state and content nonconceptualism can be traced back to the different major views of the representational content of experience. Content nonconceptualists are neo-Fregean: Bermúdez, Peacocke, Burge, etc. State nonconceptualists, in contrast, are neo-Russellians: Tye, Dretske, etc. Yet, as Heck himself recognizes, the main motivation to introduce the very notion of nonconceptual content was to differentiate perceptual states from cognitive states (Heck, 2000: 2). In the same vein, Crane complains: "the purpose of introducing the notion of nonconceptual content is to identify such a form of representation,

which is in some way more primitive, more basic than belief" (Crane, 2009: 466). To assume that the representational content of sensible intuition is nonconceptual is just to assume that the subject *is in a mental state* (sensible intuition), and does not provide her with the concepts required to specify canonically what that intuition represents.

In an extensive series of papers and books in 2005, 2006, and 2008, Hanna, the ingenious pioneer of the nonconceptualist reading of Kant, gathers a manifold of textual evidence if not in support of Kant's nonconceptualist reading then certainly in support of a buttoned-up anti-intellectualist reading. Allais (2009; 2010; 2012) makes a very convincing case in favor of the anti-conceptual reading of the Kantian view on space (2009). Inspired by Burge (2010), McLear (2011) has also provided a strong case in favor of Kant's claim about the independence of an animal's perception of concepts. There is abundant textual evidence that supports McLear's reading. Tolley (2012) provides an interesting reading of Kantian sensible intuition as a Fregean Sinn or mode of presentation of objects (*Art des Gegebenseins*). In my own modest contribution, I have tried to show, among other things, that Kantian anti-intellectualism dates back to his pre-critical writings.

Yet, the question is, do those huge amounts of textual evidence really support the nonconceptualist reading of Kantian sensible intuition?

To begin with, I see with reservation the common appeal to the independency of sensible intuition from judgments and thoughts as support for the nonconceptualist reading, what Hanna calls *Priority-to-Thought* claim (2006: 102). To be sure, Kant has claimed that sensible intuitions are independent from judgments (A90–1/B122–3). Moreover, he defines concepts as predicates of possible judgments. Still, conceptualism is not what I call here "Predicativism," for lack of a better name. Predicativism is a claim that dates back to Reid (2002), and according to it, to see or perceive *a* as *F* is the same as to judge or to think that *a is F*. To be sure, anti-predicativism is a form of anti-intellectualism. Still, conceptualism does not entail Predicativism. MacDowell (1994) is the best counter-example I know. He is certainly the most notorious proponent of conceptualism today, and he rejects Predicativism. According to McDowell, the conceptual perception of *a* being *F* is not the judgment that *a* is *F*. Thus, to show that Kantian intuition is anti-predicative is not yet to prove that Kantian intuition has a nonconceptual content.

Given this, it seems easier to accommodate Hanna's Priority-to-Thought as the Kantian claim that sensible intuitions are anti-predicative rather than the claim that sensible intuitions possess a nonconceptual content. Therefore, by claiming that "the representation that can be given prior to all thinking is called *intuition*" (B132), Kant is saying just that in order to see *a* being *F*, I do not need to judge or think that *a* is *F*. But at least if McDowell and Sellars' reading of Kantian intuition still stands, I cannot see *a* being *F* without concepts and beliefs.

The Priority-to-Thought claim motivates a further misunderstanding: the frequent identification of nonconceptualism with non-propositionalism. Indeed, for neo-Fregeans (Burge, Peacocke, Bermúdez), those doctrines are one and the same: nonconceptual contents are contents that are not structured or composed of concepts. That is the so-called content nonconceptualism. Still, as we have seen above, nonconceptualism is a notion introduced to differentiate mental states (epistemic from non-epistemic) rather than a distinction about contents themselves. Given that for Kant concepts are nothing but predicates of possible judgments, it seems easier to accommodate the Priority-to-Thought claim as the opposition between *objectual* and *propositional knowledge* rather than the opposition between nonconceptual and conceptual contents.

Moreover, the Priority-to-Thought claim motivates a further misunderstanding of nonconceptualism: the assimilation of nonconceptualism to the Russellian knowledge by acquaintance. This assimilation is untenable for various reasons. First, Russell is the *founding father* of the relational view of experience in the early 20th century. For him, knowledge by acquaintance is devoid of content that could be true or false. Only knowledge by description has content. This is the reason why he restricted knowledge by acquaintance to sense data and emphatically denied the possibility of acquaintance with bodies.

However, even if we leave aside Russell's peculiar view, Russell's opposition between objectual knowledge and propositional knowledge is not the same as the contemporary opposition between nonconceptual and conceptual contents. What it is at stake for Russell is not a classification of representations, but rather the classification of *kinds of cognitions*. Interestingly, when we take into account that Kant's crucial opposition is also described as the opposition between *Kenntnis* and *Erkenntnis*, it seems easier to accommodate the Priority-to-Thought claim as the opposition between objectual and propositional knowledge rather than the opposition between nonconceptual and conceptual contents.

That best textual evidence for the idea that what Kant had in mind was the opposition between objectual and propositional knowledge (rather than the opposition between nonconceptual and conceptual contents) comes from an opuscule of the pre-critical period of Kant's career (*FSS*):

> "I would go still further and say: it is one thing to differentiate (unterscheiden) things from each other, and quite another thing to recognize the difference between them (den Unterschied der Dinge zu erkennen). The latter is only possible by means of judgments and cannot occur in the case of animals, who are not endowed with reason. The following division may be of great use. Differentiating logically means recognizing that (*erkennen dass*) a thing A is not B; it is

> always a negative judgment. Physically differentiating (physisch unterscheiden) means being driven to different actions by different representations. The dog differentiates the roast from the loaf, and it does so because the way in which it is affected by the roast is different from the way in which it is affected by the loaf (for different things cause different sensations)." (*FSS.*, § 6, Ak., 2: 60; p. 104)

Kant's contrast can be couched in the terms of the contrast that Dretske draws between non-epistemic and epistemic seeing. The dog sees (*kennt*) things, the roast and the loaf, insofar as he is able to physically discriminate between them (non-epistemic seeing), but he is driven to different actions by the different sensations that they cause in him. However, he does not see (*erkennen*) *that* the roast is not a loaf or *that* the loaf is not a roast (epistemic seeing) in so-called categorical propositions. The capacity to know things by acquaintance (*kennen, noscere*) does not entail the capacity to know (*erkennen, dass*) that something is the case (the truth of propositions), and vice versa. I may know Paris by acquaintance without knowing, for example, that the French Revolution took place. Furthermore, I may know truths about Paris without knowing it by acquaintance. Kant and Russell's idea is that without knowledge by acquaintance (*kennen*), no genuine knowledge of the external world would be possible. Therefore, blindness of intuitions without concepts might well be understood as a lack of propositional knowledge: we know a thing without knowing any truth about it.

I also have reservations about the usual appeal to the Kantian descriptivist definition of concepts as *representatio per notas comunis*, in opposition to sensible intuition as *immediate* and *singular* representations. This also motivates a further misunderstanding of nonconceptualism: the assimilation of nonconceptualism to what I call a *mental referentialism*. To be sure, mental referentialism (direct mental reference) is a further form of anti-intellectualism. However, mental descriptivism is not the same as conceptualism in the same way that mental referentialism is not the same as nonconceptualism. First, not all conceptualists are descriptivists, and McDowell (1994) is, once more, a clear counterexample. His demonstrative-like concepts refer *immediately* in the relevant sense that reference determinated independently of the satisfaction of the features (*Merkmale*) contained in any description-like representation. Second, not all concepts are general. We also possess singular concepts.

Now if we take the Kantian definition of concepts as description-like representations, it seems easier to accommodate the Priority-to-Concept claim as the opposition between *de re* perception and *de dicto* apperception rather than the opposition between nonconceptual and conceptual contents. To be sure, *de re* attitudes is a further form of anti-intellectualism. Still, while nonconceptualism is a claim about a key distinction between mental states or representa-

tions, *de re* attitudes is a claim about *how a reference is determined*. According to Burge (1977), in a *de re* mental state, the reference is direct and not determined by concepts. Still, a *de re* mental state may be a complex thought composed of concepts. In the same vein, Bach claims that *de re* representations are those whose reference is determined relationally rather than *satisfactionally* (1987: 12).

I could summarize my disagreement with the nonconceptual readers of Kant in the following terms. For the new anti-intellectualist trend, nonconceptualism is an umbrella term that covers all kinds of anti-intellectualist claims that have emerged in different grades at different times in the history of western philosophy. It first emerged with the British Empiricists, as what Hanna calls a "super-weak version of nonconceptualism: a pure sensationalist nonconceptualism" (Hanna, 2006: 87). It then emerged with Kant, and then with Russell, Evans, Dretske, etc. My view is different. As I see it, nonconceptualism is a highly specific contemporary thesis that had several anti-intellectualist precursors; the most important in the remote past was certainly that suggested by Kant with his view on sensible intuitions.

4 Kantian Relationalism

Yet, my main concern is with the putative Kantian representationalism (content view). Hanna clearly describes Kant as a representationalist (holding a content view):

> "The central fact about the mind is its capacity to represent *vorstellen*, which is to say that the mind "puts something before itself," and this something is what Kant calls "content" *Inhalt* (A6/B9), namely *Bedeutung* of the representational state (A239–40/B298–9). (. . .) More precisely, however, for Kant the form of a conscious representation is what for lack of a better name I will call its *representational character*." (Hanna, 2006: 95)

While Hanna seems to endorse a content nonconceptualist reading of Kant, Allais clearly endorses a state nonconceptualist reading:

> "I am concerned here to argue only for the attribution to Kant of what Speaks calls 'relative', as opposed to 'absolute', non-conceptual content. The idea is that only the latter asserts that perception and belief have an intrinsically different structure; the former merely claims that a subject can have a perceptual representation with a certain content without herself possessing relevant concepts to describe that content." (Hanna, 2009: 386)

Be that as it may, regardless of whether one endorses state or content nonconceptualism (Heck, 2000), one thing is for sure: it only makes sense to talk about nonconceptualism for those who endorse the opposite so-called *content view* of experience (or representationalism).

The central tenet of representationalism (also known as the *content view*) is the claim that experiences have a content *that can be veridical or falsidical in a similar way to how propositional attitudes have a propositional content that can be true or false.* In Dretske's famous words, according to representationalism, all mental facts are representational facts (not only the so-called propositional attitudes). The mind is the representational interface of the brain. How this content should be understood is an open question.

In contrast, according to the relationalist, perception is *just a matter of putting us in direct contact with the world*. Perception does not possess any content of its own. The idea here is to take perception etymologically as a factive verb: there is no perception (seeing, hearing, touching, intuiting, etc.) when there is no object being seen or being touched. This leads relationalism forcefully to embrace disjunctivism. Even though hallucinations and experiences may be phenomenologically identical, hallucinations are not experiences. Versions of this view were popular among early 20th-century Oxford Realists like Russell (1912), but the recent work of Campbell (2002), Travis (2004), Johnston (2004; 2006), Brewer (2006), Fish (2009), and Martin (2002; 2004) has brought the relational view back into discussion. Martin (2002; 2004) calls his position "naïve realism", while Brewer (2006) calls his own the "object view". I prefer Campbell's label (2002): the "relational view".

As a committed representationalist, I have nothing to say against the "representational character of conscious representation," as Hanna puts it (2006: 95). I also do agree totally with him and Allais, that if Kant is a representationalist and if we endorse an epistemic reading of Kant's Transcendental Idealism, we also must be "direct perceptual realists": there are no intermediary entities between the mind and the world, neither Cartesian ideas, nor Humean sense-impressions, nor Russell's sense-data, etc.

However, as a Kantian reader, I have my reservations. And these are the main claim of this short paper. If we leave aside Kant's Transcendental Idealism of his phenomenalist version, Kantian writings offer overwhelming evidence of Kant's position being closer to relationalism than to representationalism. The first textual evidence is the following. As Hanna recognizes (2006: 102), Kantian sensible intuitions are *object-dependent* in the relevant sense that there is no "Vorstellung" when there is no object. Allais (2009: 389) also appeals to the same characterization. In Kant's own words, "our mode of intuition is dependent on the existence of the object" (B72).

Hanna usually calls this "veridical perception" in opposition to "non-veridical illusions:"

As mentioned in n. 7, I am using the term "veridical perception" in a precisified way to mean *sense perception that requires the actual existence of the object perceived but not necessarily an accurate representation of it.* For example, I inaccurately and thus incorrectly, but still veridically,

see that actual rose as a tulip. By contrast, I am using the term "correct perception" in a similarly precisified way to mean *sense perception that requires both the actual existence of its object and also an accurate representation of it.* For example, I accurately and thus correctly see that actual rose as a rose. Correct perception entails veridical perception, but not the converse (2006: 45ff.).

Non-veridical illusions are phenomenal representations without any existing objects, and can vary radically in content from context to context and from perceiver to perceiver. By sharp contrast, veridical illusions—e.g., the straight stick in water that appears to be bent—imply the actual existence of the object perceived, and how we represent them remains essentially the same across contexts and perceivers. Also Kant holds that perceivers can stand in non-epistemic and non-conceptual dynamical community with the objects of veridical illusion (2006: 77ff.).

However, if all sensible intuitions must be veridical in the sense that the object must exist (Object-Dependency, *CRP* B72), otherwise there is no representation of a mental state, non-veridical mental states cannot be sensible intuitions or "phenomenal representations". Moreover, it makes little sense to talk about "non-veridical illusion" because in non-veridical cases, there is no falsidical content in the first place. All conspire to the conclusion that the Kantian intuition has no content.

Moreover, for relationalists, Kant in B72 seems to take those verbs expressing experiences as *factive*: there cannot be a seeing or intuiting, unless the seen object exits; there cannot be a perceiving, unless the perceived object exits (likewise with all verbs of perception and with remember too). Indeed, if we take Kant's "Vor-stellung" etymologically, as Hanna has done in his book (2006: 113), the relationalist suspicion increases because there cannot be a *Vorstellung* when there is nothing before the mind. Thus, etymologically, sensory states that do not put us before anything are not actually *Vorstellungen* in the proper sense, but only hallucinations or imaginations. Now, in these terms, Kant is not really a representationalist but rather a relationalist and a disjunctivist.

Nevertheless, one might try to circumvent the relationalist suspicion, alleging that Kant was never rigorous with his technical terms. If sensible intuition really requires the existence of the affecting object (B72), he could never speak of "intuitions in me" (BXXXIX, footnote). Likewise, if *Vorstellung* should be understood etymologically, as putting something before the mind, Kant could never speak of "mere representations (blosse Vorstellungen)" as simple mental states devoid of any knowing objective reference (B275), or define representations as the "ground of determination of my existence that can be found in me" (BXXXIX, footnote).

The second textual evidence in favor of the relationist reading of Kant is unavoidable. Kant, both in the First *Critique* and in the *Anthropology*, emphatically asserts that sensibility *per se* never errs. In the First Critique, Kant puts this as follows:

> "Truth and illusion are not in the object, insofar as it is intuited, but in the judgment about it insofar as it is thought. Thus it is correctly said that the senses do not err; yet not because they always judge correctly, but because they do not judge at all. Hence truth, as much as error, and thus also illusion as leading to the latter, are to be found only in judgments, i.e., only in the relations of the object to our understanding." (A294/B50)

Exactly the same line of reasoning is found in *Anthropology*:

> "The senses do not deceive. This proposition is the rejection of the most important but also, on careful consideration, the emptiest reproach made against the senses; not because they always judge correctly, but rather because they do not judge at all. Error is thus a burden only to the understanding. Still, sensory appearances (species, apparentia) serve to excuse, if not exactly to justify, understanding. Thus the human being often mistakes what is subjective in his way of representation for objective (the distant tower, on which he sees no corners, seems to be round; the sea, whose distant part strikes his eyes through higher light rays, seems to be higher than the shore (*altum mare*); the full moon, which he sees ascending near the horizon through a hazy air, seems to be further away, and also larger, than when it is high in the heavens, although he catches sight of it from the same visual angle). And so one takes appearance for experience; thereby falling into error, but it is an error of the understanding, not of the senses." (*Anthr*., § 11, Ak., 7: 146; 258)

In other words, the error only occurs when the understanding, under the unnoticed influence of the faculty of sensible intuition, mistakes what subjectively appears to our senses as the real way that things are. Thus, there is no place for illusions in Kant's view of intuition. This is why Kant calls the object of intuitions "*Erscheinungen*" as opposed to appearances (*Schein*). Therefore, it is not our senses that deceive us (*betrügen*), but rather our ability to judge (*Urteilskraft*), by taking what appears to the senses to be real when this is not the case.

Now, if this is right, then representationalism never crossed Kant's mind: *sensible intuitions do not possess a representational content of their own that could be veridical or falsidical, independent of the content of judgment.* To my mind, representationalism is actually a very recent doctrine. When we read the classics, we see that they always discuss ideas, intuitions, perceptions etc., but never seem to attribute the possibility of mistake to the sensibility. For them, too, perceptual experience is *a relation*. The only difference to contemporary relationalists is that they conceive that relation *as indirect* or *mediated*,while the relationalist conceives it *as direct*. In this sense, contemporary relationalists are closer to tradition than representationalists

are. Indeed, even Dretske in his seminal work of 1969, when distinguishing non-epistemic from epistemic seeing, was not yet the representationalist he became in 1981:

> "[v]isual differentiation, as I am employing this phrase, is a pre-intellectual, pre-discursive sort of capacity which a wide variety of beings possess [and it] is an endowment which is largely immune to the caprices of our intellectual life." (Dretske, 1969: 29)

For one thing, at that time, he saw that non-epistemic seeing was a nonconceptual *relation* to an object rather than a nonconceptual *representation*. I suspect that representationalism was born with the seminal paper of Harman (1990). To my knowledge, he was the first to claim clearly that perceptual experience has a content of its own in comparison to the propositional content of propositional attitudes.

5 Kantian De Re Awareness

Let us assume for the sake of argument that Kantian sensible intuitions are really independent of any concepts. I myself have no doubt about such independency. Thus, in this regard, I am totally on the side of the nonconceptualist readers of Kant. Still, the point is that, even so, there cannot be mental states with nonconceptual content *because they do not possess a representational content of their own in the first place.*

To be sure, Kant is not a nonconceptualist as we understand that label today: neither sensible intuition nor perception possesses a representational content of its own. There is no such thing as hallucinatory perception for Kant. My point is the following: nothing changes about Kantian anti-intellectualism. Nothing substantively changes if Kant is a relationalist rather than a representationalist. *If Kantian sensible intuition is not a mental state with a nonconceptual content,is certainly in the general anti-intellectualist neighborhood.* For one thing, we can still maintain that our fundamental cognitive relation to the world, the sensible intuition (*Kenntnis*), is direct and totally independent from any kinds of concepts.

That said, McDowell and Sellars are still wrong when they claim that Kantian sensible intuitions are demonstrative-like concepts. One does not need the concept of a house (not even the demonstrative concept THIS) to see (as factive verb) a house. One needs concepts to *understand* and know (cognition) what your *sensible intuition puts before your mind*. Moreover, the mainstream of Kantian scholarship (Longuenesse, Allison, etc.) is still wrong when claiming that without categories intuitions lack objects: sensible intuition puts objects before our mind regardless of whether we understand what they stand for and regardless of whether we know that those things before our mind are mind-independent. Categories are conditions for

representing what appears as mind-independent (objectivity thesis), rather than conditions for representing what appears (intentional thesis).

Let us recall the results of the other sections. First, the Kantian Priority-to-Thought claim is better understood as the Kantian claim that sensible intuitions are anti-predicative rather than the claim that sensible intuitions possess a nonconceptual content. Second, the same Priority-to-Thought is also better understood as the opposition between objectual and propositional knowledge rather than the opposition between nonconceptual and conceptual contents. Finally, the Priority-to-Concept claim is better understood as the opposition between *de re* perception and *de dicto* aperception rather than the opposition between nonconceptual and conceptual contents. The remaining question is, how should we understand such anti-predicative, objectual, and *de re* awareness?

Let me begin by reviewing some well-known Kantian claims. "Representation" (*Vorstellung; repraesentatio*) is the foremost Kantian word for mental states whose function is to put us in relation to something. When representation is considered only as a mental state (*Modifikation des Gemüts*) resulting from the affection of the mind (*Afektion*), it is called sensation. However, when representation is considered in its referential relation to an object, it is called cognition (A320/B376). There are two kinds of cognition: intuition and conceptual. Conceptual (cognition/Erkenntnis) is the representation of objects that takes the form of propositional knowledge (*cognoscere*). Sensible cognition is the representation of objects that takes the form of knowledge by acquaintance (*noscere*).

What Kant describes as singular representation is the state of mind that puts us in direct relation to objects by means of affection (*Afektion*) or by means of acquaintance. In contrast, the mental state that refers to objects by means of functions ("the unity of action of ordering different representations under a common one" A68/B93) is what Kant describes as a general representation. General representations refer to objects indirectly, in the sense that the reference is mediated by reference to other representations (either mediated by reference to other concepts or ultimately mediated by reference to sensible intuitions). Importantly, this means that general representations refer to an object only insofar as the subject recognizes that (*erkennen dass*) the object in question falls under the extension of the concept by fulfilling one of the features (*Merkmale*) contained in the intention of the concept. This is how Kant characterizes general representations both as representations by means of notes (*Merkmale; repraesentatio per notes communes*) and as thoughts, or discursive representations (*JL*., first section, §I, Ak., 9: 91, p. 589).

However, if general or common representations refer to an object only to the extent that the subject thinks that the object falls under the extension of a concept, we may question what

it means to represent or refer to an object intuitively. An initial approximation suggests that singular representations refer to objects immediately in the negative sense that their reference is independent of any conceptual reference to them. This is what Kant has in mind when he says "that intuition is called the representation which can be given prior to all thinking" (B132).

But what does it mean to say that singular representations refer to objects non-immediately? We have seen, first, that sensible intuition is the relation to objects in a way that takes the form of knowledge by acquaintance: that is, a knowledge that is based on some direct contact with what appears. We also have seen that singular representation refers to an object insofar as it results from the affection (*Afektion*) of the mind by the object.

Here, we can appeal to Russell's opposition between knowledge by acquaintance and knowledge by description to clarify how singular representations refer to their objects. Knowledge by description is the propositional knowledge of truths or facts: that is, knowledge that something is the case. A rational being knows (by description) that a roast is not a loaf, and that a loaf is not a roast (categorical propositions). In contrast, knowledge by acquaintance is objectual knowledge gained in virtue of the fact that the subject is put by her sensible intuitions in direct in contact with an object or, in Kantian terms, due to the fact that the object affects the sensible mind. Thus, the following picture emerges. Singular immediate representations are mental states that refer *de re* to an object in the crucial sense that the reference is nothing but a direct relation of cognitive contact (Kenntnis; kennen). For example, when I see a house, my sensible intuition puts me in direct epistemic contact with what appears in my visual field and affects my sensibility.

Now, there are different ways of understanding the *de re* reference in contemporary literature. According to Bach, for example, *de re* modes of presentation are mental types whose tokens determine a different referent with respect to a context (Bach, 1987: 12). Following this view, it would be possible to argue that singular representations possess context-independent *de re* modes of presentation. They are type-individuated by the sensations and forms that are normally connected to the type of objects whose presence they evidence. Thus if the linguistic mode of presentation of a demonstrative "that house" is equal to the salient object referred to by this demonstrative (that house in the distance), the *de re* manner of presentation of the objects of singular representation can be connected to the object that normally causes this sensation with this form.

However, the *de re* manner of presentations à la Bach does not fit well for characterizations of singular representations that are understood as sensible intuitions. For one thing, for Kant, sensible intuitions are object-dependent. This means that if for sensible representation, in general, singular representations are type-individuated by sensations and forms, then for intuitions,

specifically, they are also token-individuated by the very objects that they present. It is in this sense that Kant says that "our mode of intuition is dependent on the existence of the object" (B72).

Therefore, for intuitions specifically, singular representations do have *de re* senses in the way suggested by McDowell (1991) after Evans, rather than *de re* modes of presentation à la Bach. In opposition to the *de re* modes of presentation of reference à la Bach, the distinctive feature of the *de re* sense à la McDowell is its strong object-dependence: it would not exist if the object it represented did not exist (*CRP*, B72). Likewise, for Kant, if the putative object of a sensible intuition does not exist, then there is no authentic sensible intuition (*Prol.*, §9, Ak., 4: 282; 34). In this sense, I describe Kantian sensible intuitions as *modes of donation* of objects.

My proposal is as follows. Even though sensible intuitions cannot be seen as demonstrative concepts à la McDowell (1991), they cannot be seen as mental states with nonconceptual content à la Hanna and Allais either, because they do not possess a representation content of their own in the first place. Kantian sensible intuitions are better understood as the *mode of donation* of the objects and their attributes. Thus, to say that sensible intuitions are blind without concepts is to say, like Russell does, that without concepts sensible intuition puts us in direct relational contact with objects and properties, a form of blind knowledge by acquaintance. They are *de re* perceptions of what appears as opposed to the *de dicto* apperception that something is the case.

6 Works Of Kant

References to Kant's works are given in the German Academy edition: *Gesammelte Schriften, herausgegeben von der Königlich Preussischen Akademie der Wissenschaften*, 29 vols. (Berlin: 1902–1983; 2nd ed., Berlin: De Gruyter, 1968, for vols. I–IX). They are indicated as follows: abbreviation of the title of the work, followed by Ak., volume, and page. For the *Critique of Pure Reason*, the references are shortened, in keeping with current practice, to the pagination of the original edition indicated by A for the 1781 edition, and B for the 1787 edition.

Anthr.: *Anthropologie in pragmatischer Hinsicht*, Ak. 7 (1798). *Anthropology from a Pragmatic Point of View*, trans. V. L. Dowell, rev. and ed. H. H. Rudnick (Carbondale: Southern Illinois University Press, 1978).

FSS.: *Die falsche Spitzfindigkeit der vier syllogistischen Figuren* Ak. 2 (1762). *The False subtlety of the four syllogistic figure*, trans. David Walford in collaboration with Ralf Meerbote. Ed. Paul Guyer and Allen W. Wood (Cambridge: Cambridge University Press, 1992).

"Logic," ed. J. B. [Jäsche Logic], in *Lectures on Logic*, ed. and trans. J. Michael Young (Cambridge: Cambridge University Press, 1992), 521–640.

Prol.: *Prolegomena zu einer jeden künftigen Metaphysik, die als Wissenschaft wird auftreten konnen*, Ak. 4 (1783). *Prolegomena to Any Future Metaphysics That Will Be Able to Come Forward As Science, in Philosophy of Material Nature*, trans. J. W. Ellington (Indianapolis: Hackett, 1985).

References

Allais, L. (2009), 'Kant, non-conceptual content and the representation of space', *Journal of the History of Philosophy* **47**(3), 383–413.

———. (2010), 'Kant's argument for transcendental idealism in the transcendental aesthetic', *Proceedings of the Aristotelian Society* **110**(1), 47–75.

———. (2012), Transcendental Idealism and the Transcendental Deduction, *In* Schulting, D., & Verburgt, J. (eds.) 'Kant's Idealism: New Interpretations of a Controversial Doctrine', Dordrecht: Springer, pp. 91–107.

Allison, H. E. (2004), *Kant's transcendental idealism. An interpretation and defense*, New Haven: Yale University Press.

———. (2015), *Kant's Transcendental Deduction: An Analytical-Historical Commentary*, Oxford: Oxford University Press.

Bach, K. (1987), *Thought and Reference*, Oxford: Oxford University Press.

Bermudez, J. L. (1998), *The Paradox of Self-Consciousness*, Cambrdige, MA: MIT Press.

Brewer, B. (2006), 'Perception and Content', *European Journal of Philosophy* **14**, 165–181.

Burge, T. (1977), 'Belief de re', *Journal of Philosophy* **74**(6), 338–362.

Burge, T. (2010), *Origins of Objectivity*, Oxford: Oxford University Press.

Campbell, J. (2002), *Reference and Consciousness*, Oxford: Oxford University Press.

Crane, T. (1992), The nonconceptual content of experience, *In* Crane, T. (ed.) 'The Contents of Experience', Cambridge: Cambridge University Press, pp. 136–157.

———. (2009), 'Is perception a propositional attitude?', *The Philosophical Quarterly* **59**(236), 452–469.

Dretske, F. (1969), *Seeing and Knowing*, Chicago: University of Chicago Press.

Fish, W. (2009), *Perception, Hallucination, and Illusion*, Oxford: Oxford University Press.

Gomes, A. (2014), 'Kant on Perception: Naive Realism, Non-Conceptualism, and the B-Deduction', *Philosophical Quarterly* **64**(254), 1–19.

Griffith, A. M. (2012), 'Perception and the Categories: A Conceptualist Reading of Kant's Critique of Pure Reason', *European Journal of Philosophy* **20**(2), 193–222.

Grüne, S. (2011), 'Is there a Gap in Kant's B Deduction?', *International Journal of Philosophical Studies* **19**(3), 465–490.

Gunther, Y.H. (2003), General Introduction, *In* Gunther, Y.H.(ed.) 'Essays on nonconceptual content', Cambridge: MIT Press, pp. 1–20.

Hanna, R. (2005), 'Kant and nonconceptual content', *European Journal of Philosophy* **13**(2), 247–290.

———. (2006), *Kant, Science, and Human Nature*. Oxford: Oxford University Press.

———. (2008), 'Kantian non-conceptualism', *Philosophical Studies* **137**(1), 41–64.

———. (2011), 'Kant's Non-Conceptualism, Rogue Objects, and The Gap in the B Deduction', *International Journal of Philosophical Studies* **19**(3), 399–415.

Harman, G. (1990), 'The intrinsic quality of experience', *Philosophical Perspectives* **4**, 31–52.

Heck, R. (2000), 'Nonconceptual content and the 'space of reasons'', *Philosophical Review* **109**(4), 483–523.

Henrich, D. (1982), The proof-structure of Kant's transcendental deduction, *In* Walker, R.C.S. (ed.) 'Review of Metaphysics', Oxford: Oxford University Press, pp. 640–659.

Johnston, M. (2004), 'The Obscure Object of Hallucination', *Philosophical Studies* **120**, 113–183.

———. (2006), Better than mere knowledge? The function of sensory awareness, *In* Gendler, T.S., & Hawthorne, J. (eds.) 'Perceptual Experience', Oxford: Oxford University Press, pp. 260–290.

Longuenesse, B. (1998), *Kant and the capacity to judge*, Princeton University Press.

Martin, M. G. F. (2002), 'Particular thoughts and singular thought', *Royal Institute of Philosophy Supplement* **51**, 173–214.

———. (2004), 'The limits of self-awareness', *Philosophical Studies* **120**, 37–89.

McDowell, J. (1991), Intentionality "De Re", *In* Lepore, E., & Van Gulick, R. (eds.) 'John Searle and His Critics', Oxford: Basil Blackwell, pp. 215–226.

———. (1994), *Mind and world*, Cambridge, MA: Harvard University Press.

McLear, C. (2011), 'Kant on Animal Consciousness', *Philosophers' Imprint* **11**(15), 1–16.

Pereira, R. (2017), 'A nonconceptualist reading of the B-Deduction', *Philosophical Studies* **174**, 425–442.

Reid, T. (2002), *Essays on the Intellectual Powers of Man*, Pennsylvania State University Press.

Russell, B. (1912), *The Problems of Philosophy*, Oxford: Oxford University Press.

Tolley, C. (2012), 'Kant on the Content of Cognition', *European Journal of Philosophy* **20**(4), 200–228.

Travis, C. (2004), 'The Silence of the Senses', *Mind* **113**, 57–94.