

## Dretske on Introspection and Knowledge

I. In *Naturalizing the Mind*, Fred Dretske articulates and defends a naturalistic theory of the mind which he calls «the Representation Thesis»<sup>1</sup>. In brief, this thesis states that «(1) All mental facts are representational facts, and (2) All representational facts are facts about information functions»<sup>2</sup>. From this it follows that introspective knowledge, the mind's direct knowledge of its own states, is a case of «displaced perception» – that is, knowledge of mental (i.e., representational) facts through an awareness of external (i.e., physical) objects<sup>3</sup>. In an earlier work, Dretske presents a general account of knowledge which is intended to circumvent Gettier-type counterexamples<sup>4</sup>. According to this view, *S* knows that *P* if and only if: (1) *S* believes, without doubt, reservation, or question, that *P* is the case; (2) *P* is the case; (3) *S* has a reason, *R*, for believing that *P* such that, if *P* were not the case, then *S* would not have *R* (i.e., *S* has a conclusive reason, *R*, for believing that *P*). In this paper, my aim is to explore the relation between these two theories. After summarizing both in brief detail, I argue that Dretske's account of introspective knowledge cannot be reconciled with his account of knowledge broadly construed.

<sup>1</sup> F. Dretske, *Naturalizing the Mind*, Cambridge (MA), MIT Press, 1999, p. XIII.

<sup>2</sup> *Ibidem*.

<sup>3</sup> F. Dretske, *Naturalizing the Mind*, cit., p. 41.

<sup>4</sup> F. Dretske, *Conclusive Reasons*, in *Knowledge: Readings in Contemporary Epistemology*, ed. by S. Bernecker and F. Dretske, New York, Oxford University Press, 2000, pp. 43-62.

II. The theory of mind which Dretske presents in *Naturalizing the Mind* is intended to demystify «the baffling problems of phenomenal experience»<sup>5</sup>. On this view, which he calls «Representational Naturalism», qualitative aspects of mental life, such as sense perception, are ultimately explained with reference to the representational purpose or function of consciousness<sup>6</sup>. By «representation», Dretske means a function within a system which indicates (i.e., provides information about) a property over a given domain of objects<sup>7</sup>. For example, a speedometer is a system whose function is to represent – i.e., provide information about – the speed of a given automobile. When the speedometer is functioning properly, it occupies a series of different states (pointer positions «24», «37» etc.) which correctly correspond to different automobile speeds (24 mph, 37 mph etc.). As Dretske puts it, «The fact that the speedometer has a speed indicating function, and the fact that pointing at “37” means 37 mph are representational facts about the instrument and this state of the instrument»<sup>8</sup>. Facts about the physical design of the instrument, in contrast, are *not* representational facts, since they say nothing about what sort of information the instrument is supposed to indicate<sup>9</sup>.

According to Dretske, the difference between representational facts and facts about representational systems is analogous to the difference between the mind and the brain. Detailed information about the physical operation of the brain, like detailed information about the mechanics of speedometers, fails to provide knowledge of representational facts. This is because neurophysiological facts (e.g., that human brains are divided into two hemispheres) do not by themselves supply information about what the mind represents, or what it has the function of representing<sup>10</sup>. In order to ascertain representational facts about the mind, one must determine «what

<sup>5</sup> F. Dretske, *Naturalizing the Mind*, cit., p. XIII.

<sup>6</sup> F. Dretske, *Naturalizing the Mind*, cit., p. XIV; Dretske provides a similar analysis of propositional attitudes – particularly beliefs and desires – in *Explaining Behavior*, Cambridge (MA), MIT Press, 1988.

<sup>7</sup> F. Dretske, *Naturalizing the Mind*, cit., p. 2.

<sup>8</sup> *Ibidem*.

<sup>9</sup> F. Dretske, *Naturalizing the Mind*, cit., p. 3.

<sup>10</sup> F. Dretske, *Naturalizing the Mind*, cit., p. 3.

kind of representation a mental representation is supposed to be»<sup>11</sup>.

It is unnecessary for present purposes to provide an exhaustive overview of Dretske's Representational Thesis. Still, a few points need to be addressed before looking at his theory of introspection. First, Dretske makes an important distinction between what he calls *conventional representations* and *natural representations*<sup>12</sup>. The former are informational functions which come about as a result of a system's being designed in a certain way, as when a speedometer is designed by an automobile manufacturer to provide information about automobile speeds. The latter are representations that are not conventional in this way. Dretske thinks that some informational functions – e.g., those of the senses – are acquired naturally through the process of species evolution. Thus he thinks that there are some natural representations<sup>13</sup>. The natural representations produced by the senses have a certain content that is not determined by intentional design.

Second, although all natural mental representations are representations of fact, not all such representations are *conceptual* representations. For example:

Thoughts and beliefs are classified with experiences [...]. Both are forms of natural representation [...]. One can see or hear a piano being played without believing a piano is being played, and one can believe a piano is being played without seeing or hearing it played. Seeing a piano being played is constituted, in part, by a visual experience, hearing by an auditory experience. Until these experiences occur one has not seen or heard the piano. Experiences of piano playing do not require the concept of a piano [...]. Believing is something else. It requires the concept of a piano, some understanding of what a piano is<sup>14</sup>.

The idea here is that conceptual representations of facts, such as beliefs and judgments, presuppose an ability to *say* what one is aware of *vis-à-vis* the possession of relevant concepts<sup>15</sup>. The same is not true of sensory representations. A cat, for example, can have sensory awareness of burning

<sup>11</sup> F. Dretske, *Naturalizing the Mind*, cit., p. 6.

<sup>12</sup> F. Dretske, *Naturalizing the Mind*, cit., pp. 6-7.

<sup>13</sup> F. Dretske, *Naturalizing the Mind*, cit., p. 7.

<sup>14</sup> F. Dretske, *Naturalizing the Mind*, cit., p. 9.

<sup>15</sup> F. Dretske, *Naturalizing the Mind*, cit., p. 10.

toast, since a cat can smell. It cannot, however, smell the burning toast *as* burning toast, since a cat does not know what «toast» is (i.e., it lacks the concept of «toast»).

Third, and lastly, all sensory experiences, beliefs, and thoughts are representations, and all representations are particular (token) states or events<sup>16</sup>. A token state whose indicator function is derived from the system of which it is a state is called a *systemic representation*, and its indicator function is called a *systemic indicator function*. A token state whose indicator function is assigned or acquired independently of its systemic function(s) is called an *acquired representation*, and its indicator function is called an *acquired indicator function*. Dretske supplies the following helpful example to illustrate this distinction<sup>17</sup>.

Suppose there is a simple speedometer mechanism which represents the speed of a car by measuring the rotation of the car's axle. Suppose, too, that this speedometer can be used in cars with different tire sizes, and that the owner of the car must calibrate the speedometer according to the size of the car's tires. In car A, which has larger tires, the owner puts the number «60» at position *x* of the pointer; in car B, which has normal tires, the owner puts the number «50» at position *x* of the pointer. Thus, when the axles of car A and car B are rotating at a rate corresponding to pointer position *x*, car A is going 60 mph and car B is going 50 mph. In both cars, pointer position *x* has the same *systemic indicator function* – viz., indicating an axle rotation rate of *N* rpm. However, pointer position *x* has different *acquired indicator functions* in car A and car B, since what it represents about speed varies from car to car.

Dretske thinks that the representational properties of thoughts, beliefs, and other conceptual states are *acquired* and that the representational properties of sensory experiences are *systemic*<sup>18</sup>. With regard to the former, Dretske claims that the representational content of beliefs and thought depends upon a conceptual system which is brought about, and can be changed, by learning<sup>19</sup>. With regard to the latter he says:

<sup>16</sup> F. Dretske, *Naturalizing the Mind*, cit., p. 12.

<sup>17</sup> F. Dretske, *Naturalizing the Mind*, cit., pp. 13-14.

<sup>18</sup> F. Dretske, *Naturalizing the Mind*, cit., p. 15.

<sup>19</sup> *Ibidem*.

Experiences have their representational content fixed by the biological functions of the sensory systems of which they are states [...]. The quality of a sensory state – how things look, sound, and feel at the most basic (phenomenal) level – is thus determined phylogenetically<sup>20</sup>.

This distinction explains the difference between a sensation of redness and the belief that a given object is red. Both sensations and beliefs are natural representations, but only sensations (as well as experiences and feelings) have systemic indicator functions.

Much more could be said about the Representational Thesis, but the foregoing has provided a sufficient overview of its most important features. That said, I turn now to Dretske's account of introspective knowledge. On his view, introspective knowledge is «the mind's direct knowledge of itself», and introspection is «the process by means of which we come by such knowledge»<sup>21</sup>. What one comes to know through introspection are facts about one's «mental life» – facts about internal representations<sup>22</sup>. However, the objects and events by means of which one learns these facts are not mental: «One becomes aware of representational facts by an awareness of physical objects. One learns that *A* looks longer than *B*, not by an awareness of the experience that represents *A* as longer than *B*, but by an awareness of *A* and *B*, the objects the experience is an experience of»<sup>23</sup>. For this reason, Dretske thinks, introspection is an instance of *displaced perception* – knowledge of internal (i.e., mental, representational) facts via an awareness of external (i.e., physical) objects.

In general, displaced perception involves seeing that some object *k* is *F* by seeing, not *k* itself, but some other object *h*. Dretske explains this definition with the following examples:

I see how much I weigh by looking at the bathroom scale on which I stand. The object I see is the bathroom scale. The fact I learn is a fact about me – that I weigh 170 pounds. This pattern – perceptual object in one place, perceptual fact in another – is familiar. One looks at a gauge on the dash-

<sup>20</sup> *Ibidem*.

<sup>21</sup> F. Dretske, *Naturalizing the Mind*, cit., p. 39.

<sup>22</sup> *Ibidem*.

<sup>23</sup> *Ibidem*.

board in order to see how much gas remains in the gas tank – a fact about an object, a tank beneath the car, that one does not see<sup>24</sup>.

In cases of displaced perception, there is a conceptual, but no corresponding sensory, representation of  $k$ <sup>25</sup>. When I look at a gas gauge, for instance, I have a conceptual representation of the gas tank bearing a certain property (e.g., containing 5 gallons of gasoline), but I acquire this representation by means of a sensory representation of the gauge. The properties sensuously represented to me are properties of the gauge, not the gas tank. However, the property conceptually represented to me – i.e., containing 5 gallons of gasoline – is a property of the gas tank, not the gauge.

On Dretske's view, again, introspective knowledge is knowledge of mental (hence, representational) facts. Furthermore, as the foregoing makes clear, introspective knowledge involves a conceptual representation of a representation – i.e., a conceptual representation of the fact that something is a representation or has a certain representational content. It is thus what Dretske and others call «metarepresentational»<sup>26</sup>. A metarepresentation isn't merely a representation of a representation, as when I represent a photographic image of  $x$  as a piece of paper weighing two grams. Rather, it is a representation of a representation *as* a representation, as when I represent a photographic image of  $x$  as being an image *of*  $x$  – that is, as being a representation bearing certain representational content<sup>27</sup>.

In the same way, introspective knowledge is a representation of an experience or belief – which are themselves representations – as experiences *of this* or beliefs *about that*. As Dretske puts it, «If E is an experience (sensory representation) of blue, then introspective knowledge of this experience is a conceptual representation of it as an experience of blue

<sup>24</sup> F. Dretske, *Naturalizing the Mind*, cit., p. 41.

<sup>25</sup> *Ibidem*.

<sup>26</sup> F. Dretske, *Naturalizing the Mind*, cit., p. 43; cf. J. Perner, *Understanding the Representational Mind*, Cambridge (MA), MIT Press, 1991, p. 35; Z.W. Pylyshyn, *When is Attribution of Beliefs Justified?*, «The Behavioral and Brain Sciences», I, 1978, pp. 592-93.

<sup>27</sup> F. Dretske, *Naturalizing the Mind*, cit., p. 43.

(or of color)»<sup>28</sup>. In introspection, then, a given experience is conceptually represented to me, via a sensory representation, as being an experience with a certain kind of representation content. This sensory representation, moreover, is a representation not of the experience itself, but rather of some displaced object. Thus, for example, one comes to introspectively know that she is experiencing blue by experiencing, not the experience of blue, but rather some blue object.

Dretske likens human minds to instruments, such as pressure gauges, the function of which is to represent facts about other objects in the world<sup>29</sup>. Suppose there is an instrument,  $S$ , whose function is to indicate the  $F$  of some object  $k$  (that is, to represent  $k$  as being  $F$ , where  $F$  is some quantifiable attribute of  $k$  such as velocity, weight, etc.). Suppose, further, that  $S$  indicates the  $F$  of  $k$  by means a pointer device, and the pointer occupies position  $P$ . Now, the fact that  $S$  represents  $k$  is a representational fact which further entails that (1)  $k$  stands in some relation,  $C$ , to  $S$ , and (2) for some quantifiable property  $F$ ,  $S$  represents the  $F$  of whatever  $k$  stands in relation  $C$  to  $S$ <sup>30</sup>. The fact that there is an object whose  $F$  is represented by  $S$  is not itself a representational (hence, mental) fact about  $S$ . The same is true of (1) – i.e., the fact that there is an object  $k$  which stands in some relation  $C$  to  $S$ . Therefore, Dretske says, «To know the “mind” of an instrument is to know (2) but not (1). If instruments could introspect, if they could know their own representational states, we would expect them to know how they represent what (if anything) they represent, what determinate value of  $F$  they represent it as having, but not what object – or that there is an object – that they represent in this way»<sup>31</sup>.

The question is: how might we go about determining what pointer position  $P$  on instrument  $S$  indicates about object  $k$ ? In other words, «How [...] would we find out how an instrument represents the objects (if any) it represents? How would we find out the representational facts, the facts that define what was going on in the representational “mind”

<sup>28</sup> F. Dretske, *Naturalizing the Mind*, cit., p. 44.

<sup>29</sup> F. Dretske, *Naturalizing the Mind*, cit., pp. 45-63.

<sup>30</sup> F. Dretske, *Naturalizing the Mind*, cit., p. 45.

<sup>31</sup> F. Dretske, *Naturalizing the Mind*, cit., p. 46.

of an instrument?»<sup>32</sup>. Suppose that  $S$  is a functioning pressure gauge. If the pointer mechanism on the gauge is clearly marked off in pounds per square inch (psi), and if the pointer occupies position «14», we know that the gauge is representing whatever it is connected to as having a pressure of 14 psi<sup>33</sup>. But suppose that the pointer positions are illegible or otherwise uninterpretable. In such a case, how would we go about determining what the gauge is representing when its pointer occupies position  $P$ ?

In order to answer this question, Dretske says, we must find out «what information  $P$  is supposed to carry about pressure, what information  $P$  has the [systemic function] of providing»<sup>34</sup>. If we know that  $S$  is a pressure gauge connected to  $k$ , and if we know that  $S$  is working properly, then we would merely have to determine, by independent means, what the pressure in  $k$  is when the gauge's pointer occupies position  $P$ <sup>35</sup>. If it turns out that the pressure in  $k$  is 14 psi, then we know that position  $P$  on the gauge has the function of indicating that  $k$  has a pressure of 14 psi. Dretske refers to this as a «calibrational process» which specifies «what states [ $P$ ] mean by comparing what the system [ $S$ ] says, when it speaks truly, with the facts [the  $F$  of  $k$ ] about which it speaks»<sup>36</sup>. In the case of human minds, we determine whether a mental «system» is functioning properly by analyzing the kind of information it was «designed», by means of natural selection, to provide<sup>37</sup>.

To summarize: in order to determine how a system  $S$ , whether natural or artificial, is representing a given object  $k$ , one must know what  $S$ 's reaction to  $k$  means<sup>38</sup>. This, in turn, is determined by analyzing what value of  $F$  brings about this reaction (i.e.,  $S$ 's occupying position  $P$ ) when  $S$  is functioning properly. As Dretske says, «If [ $S$ 's] reaction to  $k$  is  $P$ , and,

<sup>32</sup> *Ibidem*.

<sup>33</sup> F. Dretske, *Naturalizing the Mind*, cit., p. 46.

<sup>34</sup> F. Dretske, *Naturalizing the Mind*, cit., p. 48.

<sup>35</sup> *Ibidem*.

<sup>36</sup> F. Dretske, *Naturalizing the Mind*, cit., p. 49.

<sup>37</sup> F. Dretske, *Naturalizing the Mind*, cit., pp. 49-50. Dretske admits that we might be mistaken about the natural indicator functions of human mental states, but he does not regard this as problematic.

<sup>38</sup> F. Dretske, *Naturalizing the Mind*, cit., p. 50.



when it is functioning properly,  $P$  is [ $S$ 's] way of reacting to a pressure of 10 psi, then [ $S$ ] is representing  $k$

as having a pressure of 10 psi»<sup>39</sup>. The question becomes: how might an instrument, such as a pressure gauge, come to know what pointer position  $P$  represents the pressure in  $k$  as? Obviously a gauge cannot determine whether it is working properly, nor can it independently determine the actual pressure of  $k$  for the purposes of calibrating itself<sup>40</sup>. If human minds are comparable to pressure gauges in this way, how could introspective knowledge be possible?

The answer, as Dretske explains, has to do with the difference between representational systems and external observers: «When I, an external observer, try to determine what state  $P$  means in system  $S$ , I do not, whereas  $S$  does, occupy the state whose representational content is under investigation.  $S$ , therefore, *has* information – whatever information is carried by state  $P$  – that I do not»<sup>41</sup>. While external observers can only acquire this information by means of the calibrational process outlined above,  $S$  has instant and reliable access to it merely by representing  $k$  as  $F$ , on the assumption that the value of position  $P$  correctly represents  $k$  as  $F$  if  $S$  is working properly<sup>42</sup>. The system, unlike the external observer, has knowledge about its own representational states, and this knowledge specifies what the world would be like if it were functioning properly. Whereas the external observer must examine both  $S$  and  $k$  to determine what  $S$  is representing  $k$  as,  $S$  itself «need only look to the world, at whatever it is already ‘looking’ at (i.e.,  $k$ ) to get this information»<sup>43</sup>. This, on Dretske’s view, is the source of the first-person authority of introspective knowledge.

What distinguishes human minds from other representational systems is the capacity for metarepresentation – that is, the mind’s capacity to represent itself or its states as representations of the world. In introspection, I know that my experience is an experience of blue (rather than red or yellow)

<sup>39</sup> *Ibidem*.

<sup>40</sup> F. Dretske, *Naturalizing the Mind*, cit., p. 51.

<sup>41</sup> *Ibidem*.

<sup>42</sup> *Ibidem*.

<sup>43</sup> F. Dretske, *Naturalizing the Mind*, cit., p. 53.

by perceiving, not the experience itself (or any other internal state), but rather a blue object in the world. My sensory experience need not be veridical in order for me to ascertain facts about how I am representing the world to be<sup>44</sup>. This information is obtained immediately and non-reflectively by the sensory experience itself. As Dretske says,

To know I am experiencing blue, to know that that is the kind of color experience I am having, I need only the experience of blue [...]. Given that I understand the concept of experience and its qualities, I have, in my experience of blue, all I need to know what kind of experience I am having<sup>45</sup>.

We could say more about Dretske's theory of introspection, but I think the foregoing has provided a sufficient explanation of what he takes introspective knowledge to be. We can summarize this in the following manner:

- (1) All mental facts are representational facts.
- (2) Introspective knowledge is knowledge of facts about one's own mental states.
- (3) I introspectively know my mental state  $b$  is  $F$ , not by perceiving that  $b$  is  $F$ , but by perceiving that  $k$  is  $F$  (where  $k$  is a physical object).
- (4) Therefore, introspective knowledge is knowledge of (conceptual) representational facts ascertained by way of (sensory) representational facts about the world.

Although we may assume that introspective knowledge is a species of knowledge broadly construed, Dretske fails to provide a definition of the latter in *Naturalizing the Mind*. For this we must look elsewhere.

III. According to a long-standing tradition in the history of philosophy, there are three necessary and sufficient conditions for someone's knowing a given proposition<sup>46</sup>. These may be summarized as follows:

<sup>44</sup> F. Dretske, *Naturalizing the Mind*, cit., p. 61.

<sup>45</sup> F. Dretske, *Naturalizing the Mind*, cit., p. 63.

<sup>46</sup> This view of knowledge seems to have originated with Plato (*Theaetetus* 201; *Meno* 98); cf. R. Chisholm, *Perceiving: A Philosophical Study*, Ithaca (NY), Cornell University Press, 1957, p. 16; A.J. Ayer, *The Problem of Knowledge*, London, Macmillan, 1956, p. 34.

S knows that  $p$  if and only if:

- (1)  $p$  is true.
- (2) S believes that  $p$ , and
- (3) S is justified in believing that  $p$ .

In an influential article published in 1963, Edmund Gettier attempts to demonstrate that this conception of knowledge is false<sup>47</sup>. He begins by noting that (1) «for any proposition  $P$ , if  $S$  is justified in believing  $P$ , and  $P$  entails  $Q$ , and  $S$  deduces  $Q$  from  $P$  and accepts  $Q$  as a result of this deduction, then  $S$  is justified in believing  $Q$ ; and (2) in that sense of «justified» in which  $S$ 's being justified in believing  $P$  is a necessary condition of  $S$ 's knowing that  $P$ , it is possible for a person to be justified in believing a proposition that is in fact false»<sup>48</sup>. On the basis of these points, Gettier presents two cases which are intended to show that the above definition of knowledge fails to provide sufficient conditions for someone's knowing a given proposition. In the interest of brevity, I will only summarize the first of these<sup>49</sup>.

Suppose that Smith and Jones have both applied for a certain job. Suppose, further, that Smith has strong evidence for believing that  $P$ : «Jones is the man who will get the job, and Jones has ten coins in his pocket»<sup>50</sup>.  $P$  therefore entails  $Q$ : «The man who will get the job has ten coins in his pocket»<sup>51</sup>. Now, on the assumption that Smith recognizes that  $P$  entails  $Q$  and believes that  $Q$  on the basis of his justified belief that  $P$ , it follows that Smith is justified in believing that  $Q$ . But suppose that Smith, not Jones, is actually the man who will get the job, and that he, too, has ten coins in his pocket. From this it follows that  $Q$  is true, even though  $P$ , from which  $Q$  was inferred, is false. Moreover, it follows that: (1)  $Q$  is true; (2) Smith believes that  $Q$  is true; and (3) Smith is justified in believing that  $Q$  is true. But, as Gettier points out, «it is equally clear that Smith does *know* that [ $Q$ ]

<sup>47</sup> E.L. Gettier, *Is Justified True Belief Knowledge?*, «Analysis», XXIII, 1963, pp. 121-23; reprint in *Knowledge: Readings in Contemporary Epistemology*, ed. by S. Bernecker and F. Dretske, Oxford, Oxford University Press, 2000, pp. 13-15.

<sup>48</sup> E.L. Gettier, *Is Justified True Belief Knowledge?*, cit., p. 13.

<sup>49</sup> E.L. Gettier, *Is Justified True Belief Knowledge?*, cit., p. 14.

<sup>50</sup> *Ibidem*.

<sup>51</sup> *Ibidem*.

is true, for [Q] is true in virtue of the number of coins in Smith's pocket, while Smith does not know how many coins are in Smith's pocket, and bases his belief in [Q] on a count of the coins in Jones's pocket, whom he falsely believes to be the man who will get the job»<sup>52</sup>. Therefore, knowledge cannot be *merely* justified true belief.

We need not be concerned with whether Gettier's argument works<sup>53</sup>. For our purposes, it is enough to note that Dretske's «conclusive reasons» account of knowledge is specifically intended to rule out Gettier-type counterexamples. As he puts it, «In having conclusive reasons to believe that *P* is the case one's epistemic credentials are such as to eliminate the possibility of mistake»<sup>54</sup>. By «conclusive reasons», Dretske means grounds for believing that *P* which guarantee that a subject *S* «could not be wrong about *P* or, given these reasons, it is false that he might be mistaken about *P*»<sup>55</sup>. In other words, suppose that:

- (1) *S* knows that *P* and he knows this on the basis (simply) of *R*.
- (2) *R* would not be the case unless *P* were the case.

Dretske thinks that if (2) is true, then *R* is a conclusive reason for *P*<sup>56</sup>. This follows from the fact that, if (2) is true, we are forced to deny both that not-*P* is the case and that not-*P* *might* be the case. In order to rule out Gettier-type counterexamples, Dretske says, any knowledge that *P* that must rest on evidence, grounds, or reasons that are conclusive in this way<sup>57</sup>. We will return to this point presently.

Dretske examines the relation between (1) and (2) above in great detail<sup>58</sup>. We need not concern ourselves here with

<sup>52</sup> *Ibidem*.

<sup>53</sup> For further reading on this subject, see: R. Almeder, *The Invalidity of Gettier-Type Counterexamples*, «Philosophia: Philosophical Quarterly of Israel», XIII, 1983, pp. 67-74; R. Slaughter, *Is Justified True Belief Knowledge? A Selective Critical Survey of Recent Work*, «Philosophy Research Archives», III, 1977, pp. 1-135; S. Sturgeon, *The Gettier Problem*, «Analysis», LIII, 1993, pp. 156-64.

<sup>54</sup> F. Dretske, *Conclusive Reasons*, cit., p. 42.

<sup>55</sup> *Ibidem*.

<sup>56</sup> *Ibidem*.

<sup>57</sup> *Ibidem*.

<sup>58</sup> F. Dretske, *Conclusive Reasons*, cit., pp. 43-50.

what he says; it is enough to note that  $R$  is a conclusive reason for  $P$  if and only if  $R$  would not be the case unless  $P$  were the case. Notice that  $R$  can be either *logically* or *empirically* conclusive in this sense.  $R$  is logically conclusive just in case the truth of (2) can be demonstrated on «purely logical and definitional grounds»<sup>59</sup>.  $R$  is empirically conclusive, in contrast, just in case the conditional in (2) is true, but not logically true. Now, since  $R$  can be a conclusive reason for believing that  $P$  even if no one believes  $P$  (or believes  $P$  on the basis of  $R$ ), further conditions must be specified to show when  $S$  has a conclusive reason for believing that  $P$ . These can be summarized as follows.  $S$  has a conclusive reason,  $R$ , for believing that  $P$  if and only if:

- (A)  $R$  would not be the case unless  $P$  were the case (i.e.,  $R$  is a conclusive reason for  $P$ );
- (B)  $S$  believes, without doubt, reservation, or question, that  $P$  is the case and he believes this on the basis of  $R$ .
- (C)  $S$  knows that  $R$  is the case<sup>60</sup>.

On Dretske's view,  $S$ 's having conclusive reasons for believing that  $P$  is both a necessary and a sufficient condition for his knowing that  $P$ . As Dretske points out, «If  $S$  has conclusive reasons for believing  $P$ , then it is *false* to say that, given these grounds for belief, and the circumstances in which these grounds served as the basis for his belief,  $S$  might be mistaken about  $P$ »<sup>61</sup>. This rules out the aforementioned counterexample from Gettier, since  $Q$  does not constitute a conclusive reason for Smith's believing that  $P$ . The proposition «The man who will get the job has ten coins in his pocket» could clearly be true even if the proposition «Jones is the man who will get the job, and Jones has ten coins in his pocket» is false.

On Dretske's view, then,  $S$  knows that  $P$  if and only if:

<sup>59</sup> F. Dretske, *Conclusive Reasons*, cit., p. 53.

<sup>60</sup> *Ibidem*. To condition (C) Dretske adds «or  $R$  is some experiential state of  $S$  (about which it may not make sense to suppose that  $S$  knows that  $R$  is the case; at least it no longer makes much sense to ask *how* he knows)».

<sup>61</sup> *Ibidem*.

(1)  $R$  would not be the case unless  $P$  were the case (i.e.,  $R$  is a conclusive reason for  $P$ )

(2)  $S$  believes, without doubt, reservation, or question, that  $P$  is the case and he believes this on the basis of  $R$ .

(3)  $S$  knows that  $R$  is the case.

(4)  $P$  is the case.

Two qualifications are worth noting here. First, in presenting the conclusive reasons condition, Dretske does not mean to suggest that, in having  $R$ ,  $S$  would necessarily be in a position to *give*  $R$  as his reason for believing that  $P$ . As he puts it, « $R$  may simply be a certain experience which  $S$  has undergone and, having undergone this experience, come to the belief that  $P$  was the case on the basis of (as a result of) this experience. He may find it difficult, or impossible, to give verbal expression to  $R$ . He may have forgotten it. Or it may consist in something's looking a particular way him which he finds difficult to describe»<sup>62</sup>. Second, Dretske does not mean to suggest that, in having  $R$ ,  $S$  has necessarily deduced  $P$  from  $R$ , or otherwise reasoned his way to  $P$  from premises involving  $R$ <sup>63</sup>.

Dretske discusses all of these points in great detail and goes to great lengths to defend his account from possible objections. For our purposes, it is unnecessary to address all of this, since we are willing to stipulate that Dretske's account is correct. Our main topic of interest is this: if introspective knowledge is a species of knowledge, and if knowledge requires conclusive reasons, then doesn't introspective knowledge also require conclusive reasons? If so, what *sorts* of conclusive reasons are required for introspective knowledge? Can such reasons be given? I propose to answer these questions in the following, and final, section of this paper.

IV. We must begin by recalling that «displaced perception», on Dretske's view, means «seeing that  $k$  is  $F$  by seeing, not  $k$ , but some other object,  $h$ ». Introspective knowledge about an experience – say, of the color blue – «is an instance in which an experience [...] is conceptually represented as an experience of blue via a sensory representation not of the

<sup>62</sup> F. Dretske, *Conclusive Reasons*, cit., p. 55.

<sup>63</sup> F. Dretske, *Conclusive Reasons*, cit., p. 56.

experience, but of some other [displaced] object»<sup>64</sup>. The displaced object in question is usually some object in the external world which is, in turn, the object of the experience of blue. Thus, one comes to possess introspective knowledge that one is experiencing blue by experiencing something other than the experience of blue – namely, a blue object.

Let us say, then, that *S* has introspective knowledge of *P* just in case:

(1) *P* is a proposition about one or more of *S*'s mental states (say, that *b* is *F*).

(2) *S* knows believes, without doubt, reservation, or question, that *P* is the case, and she believes it on the basis of *R*, where *R* is a proposition about some physical object *k* (say, that *k* is *F*).

(3) *P* is the case.

Recall that, for Dretske, (3) is guaranteed to be true as long as *S* is representing *b* as *F*: «I do not have to *truly* represent the color of *k* in order to get information about myself from my sensory representation of *k*»<sup>65</sup>. From this it follows that *R* need not be the case in order for *P* to be the case. In fact, it seems that *S* doesn't even need to believe that *R* is the case in order to believe that *P* on the basis of *R*. For example, *S* could say «I believe that *P* because it *seems* that *k* is *F*».

However, if introspective knowledge is truly a species of *knowledge* as Dretske defines this concept in «Conclusive Reasons», then it would seem that *S* has introspective knowledge that *P* if and only if:

(1) *P* is a proposition about one or more of *S*'s mental states (say, that *b* is *F*)

(2) *S* believes, without doubt, reservation, or question, that *P* is the case, and she believes this on the basis of *R*.

(3) *S* has a reason, *R*, for believing that *P* is the case such that, if *R* were not the case, then *P* would not be the case (i.e., *S* has a conclusive reason *R* for believing that *P*).

(4) *R* is a proposition about one or more physical objects (say, that *k* is *F*).

(5) *P* is the case.

<sup>64</sup> F. Dretske, *Naturalizing the Mind*, cit., p. 44.

<sup>65</sup> F. Dretske, *Naturalizing the Mind*, cit., p. 61.

Since introspective knowledge is a kind of displaced perception (i.e., seeing that  $k$  is  $F$  by seeing, not  $k$ , but some other object,  $b$ ), it follows that  $R$  will be a proposition about an external (physical) object. Now, suppose  $R$  is the proposition «the chair is blue», and  $P$  is the proposition «I am experiencing the color blue». In order for  $S$  to have introspective knowledge that  $P$ , one must have a conclusive reason,  $R$ , for believing that  $P$ . But  $R$  is a conclusive reason for  $P$  just in case, among other things,  $R$  would not be the case unless  $P$  were the case. Thus, in order for  $R$  to be a conclusive reason for  $S$  to believe that  $P$ , the proposition «the chair is blue» would have to be such that it wouldn't be true unless the proposition «my experience of the chair is blue» were also true. But clearly this is not the case, since the chair might be blue even though I misrepresent it as red. The same is true, I submit, of any possible proposition of this sort – and since Dretske admits the possibility of perceptual mistake, he clearly agrees. *Therefore, one cannot have introspective knowledge about one's own mental states.*

Now, it may be argued that there is no reason to assume that the above proposition about physical objects needs to be identical to the *conclusive reason* which justifies  $S$ 's belief that  $P$ . For example,  $S$  could believe that  $P$  («my present experience is an experience of blue») on the basis of  $R$  («the object I am presently experiencing is a blue object»), *and* on the basis of  $Q$ , where  $Q$  is a separate proposition that constitutes a conclusive reason for believing that  $P$ . But what sort of proposition would  $Q$  be? It's not clear how  $Q$  could be a proposition about *another* physical object. In fact, it seems that  $Q$  *must* be a proposition about one or more of  $S$ 's mental states. But what sort of proposition about one or more of  $S$ 's mental states would not be the case unless  $P$  were also the case? Furthermore, wouldn't this proposition, by virtue of stating a representational fact, necessarily be held on the basis of a *further* proposition,  $T$ , which is itself a proposition about a physical object? It seems obvious that this sort of reasoning will very quickly lead to an infinite explanatory regress.

It may also be argued that conclusive reasons are unnecessary in the case of introspective knowledge since, as we have mentioned,  $P$  is guaranteed to be true. The fact that I can-



not be mistaken about  $P$  follows, not from the conclusiveness of  $R$ , but rather from the indefeasibility of (3) – i.e., « $P$  is true». After all, as long as I am in the state described by  $P$ , then  $P$  necessarily cannot fail to be true. If this is the case, however, it follows that there is at least one type of true belief that is properly termed «knowledge» even though it is unsubstantiated by conclusive reasons – viz.,  $S$ 's belief that  $P$ , where  $P$  is a proposition about one of  $S$ 's own mental states and  $P$  is true. And if this is the case, Dretske's «conclusive reasons» definition fails to account for at least one special type of knowledge, and possibly others as well. This is a serious shortcoming, since Dretske's own account of introspective knowledge appears to provide a counter-example to his theory of knowledge broadly construed.

One final question merits consideration. Suppose that  $S$ 's introspection knowledge that  $P$  merely requires that (1)  $P$  is a proposition about one or more of  $S$ 's mental states; (2)  $S$  believes, without doubt, reservation, or question that  $P$  is the case, and she believes this on the basis of  $R$  (where  $R$  is a proposition about a physical object); and (3)  $P$  is true. Suppose, further, that  $S$  mistakenly refers to all her blue color experiences as green. If this is the case, then:

(1)  $P$  is a proposition about one or more of  $S$ 's mental states (say, that her experience is an experience of the color which she calls «green»).

(2)  $S$  believes, without doubt, reservation, or question that  $P$  is the case, and she believes this on the basis of  $R$  (where  $R$  is a proposition about a physical object, say, that the chair  $S$  is experiencing is the color which she calls «green»)

(3)  $P$  is the case.

Now, we want to say that (3) is false, since  $S$ 's experience is really of blue, not green. But  $S$ 's experience is, among other things, a conceptual representation, and as such, it is limited by her extant conceptual scheme. If she lacks the concept «blue», then it makes no sense to say that (3) is false, since  $S$ 's experience is clearly an experience of *something*, and it can't be an experience of anything other than green, since «green» is what she represents the color of the chair *as*. If it were otherwise, it wouldn't be  $S$ 's experience at all. In order to avoid this kind of problem, it seems that we would have

to add an additional condition which rules out the possibility of misapplied concepts. But I shall not investigate this point further.

V. I began this paper by discussing Dretske's views on introspective knowledge as outlined in his book *Naturalizing the Mind*. I then examined his account of knowledge broadly construed as presented in the article «Conclusive Reasons». Finally, I argued that the former account cannot be reconciled with the latter, since, on Dretske's view, *S* does not require conclusive reasons to have introspective knowledge about her own mental states. If this argument is successful, it shows either that (1) introspective knowledge is not really knowledge, at least as Dretske understands this concept; or (2) at least some forms of genuine knowledge fall outside the scope of Dretske's definition of knowledge, in which case we have reason to regard this definition with suspicion. I ended by broaching a potential problem with Dretske's account of introspection that involves the possibility of conceptual misapplication. Together, I think these criticisms constitute a strong objection to Dretske's view overall.

*Summary. Dretske on Introspection and Knowledge*

In *Naturalizing the Mind*, Fred Dretske articulates and defends a naturalistic theory of the mind which he calls «the Representation Thesis». In brief, this thesis states that «(1) All mental facts are representational facts, and (2) All representational facts are facts about information functions». From this it follows that introspective knowledge, the mind's direct knowledge of its own states, is a case of «displaced perception» – that is, knowledge of mental (i.e., representational) facts through an awareness of external (i.e., physical) objects. In an earlier work, Dretske presents a general account of knowledge which is intended to circumvent Gettier-type counterexamples. According to this view, *S* knows that *P* if and only if: (1) *S* believes, without doubt, reservation, or question, that *P* is the case; (2) *P* is the case; (3) *S* has a reason, *R*, for believing that *P* such that, if *P* were not the case, then *S* would not have *R* (i.e., *S* has a conclusive reason, *R*, for believing that *P*). In this paper, my aim is to explore the relation between these two theories. After

summarizing both in brief detail, I argue that Dretske's account of introspective knowledge cannot be reconciled with his account of knowledge broadly construed.

*Riassunto. Dretske su introspezione e conoscenza*

In *Naturalizing the Mind* Fred Dretske articola e difende una teoria naturalistica della mente che chiama la «tesi della rappresentazione». In breve questa tesi afferma che «(1) Tutti i fatti mentali sono fatti rappresentazionali e che (2) tutti i fatti rappresentazionali sono fatti concernenti funzioni di informazione». Da ciò segue che la conoscenza introspettiva, ossia la conoscenza diretta che la mente ha dei propri stati, è un caso di «percezione dislocata», vale a dire una conoscenza di fatti mentali (rappresentazionali) attraverso una consapevolezza degli oggetti esterni (fisici). In un suo lavoro precedente Dretske presentava una concezione generale della conoscenza che voleva sottrarsi ai controesempi del tipo di quelli proposti da Gettier. Secondo questa concezione, *S* conosce che *P* se e solo se: (1) *S* crede senza nessun dubbio, nessuna riserva o obiezione che *P* sussiste; (2) *P* sussiste; (3) *S* ha una ragione *R* per credere che *P* tale che, se *P* non sussistesse, allora *S* non avrebbe *R* (cioè, *S* ha una ragione conclusiva, *R*, per credere che *P*). Il presente articolo intende analizzare la relazione tra queste due teorie. Dopo averle ricapitolate brevemente entrambe, avanzo la tesi che la concezione della conoscenza introspettiva di Dretske non si concilia con la sua concezione generale della conoscenza.

*Keywords:* Gettier, Knowledge, Introspection, Representation.

