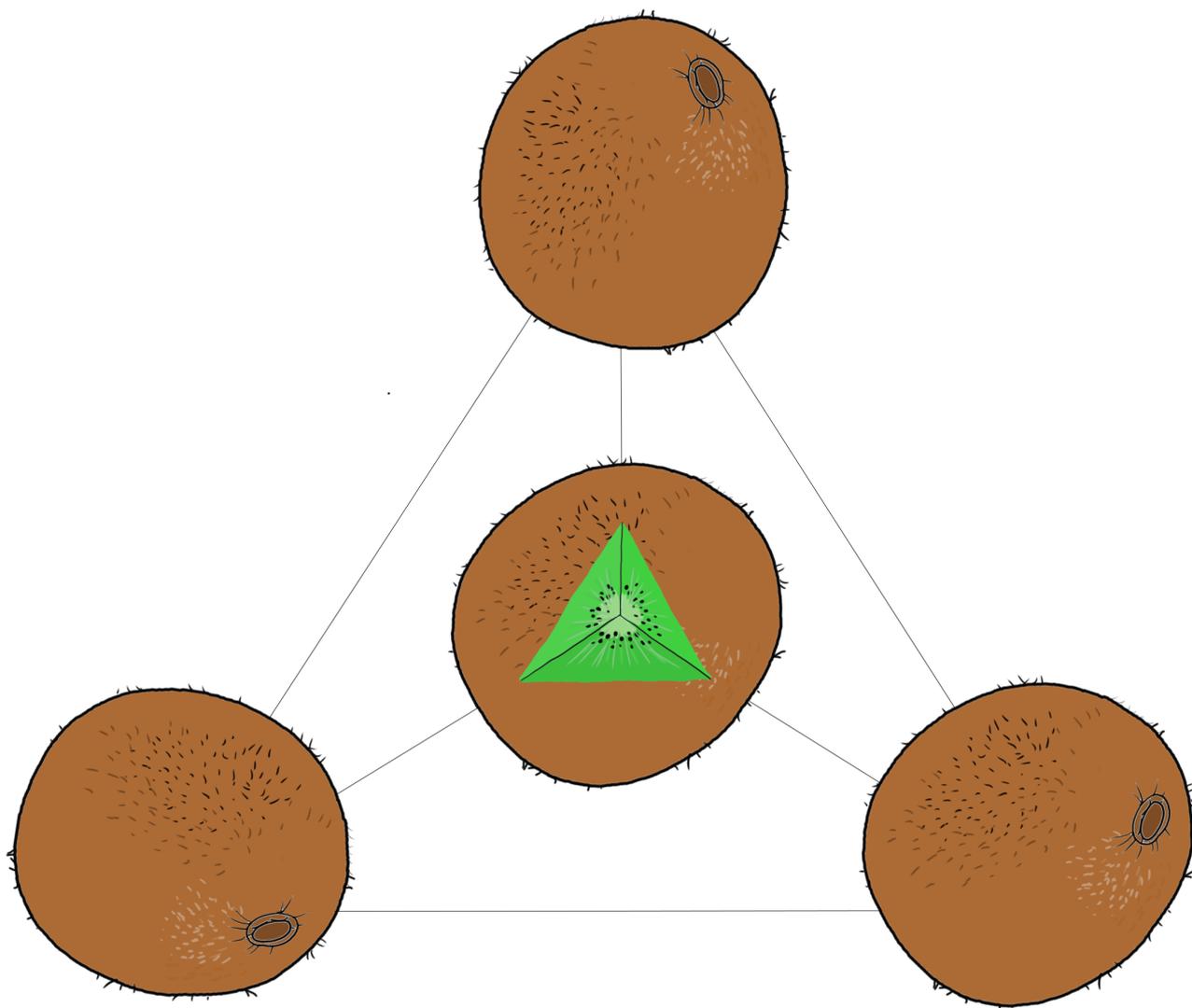


# R I F A J

RIVISTA ITALIANA DI FILOSOFIA ANALITICA JUNIOR

- PEER REVIEWED JOURNAL -



PATROCINATA  
DALLA SOCIETÀ  
ITALIANA DI  
FILOSOFIA  
ANALITICA  
DAL 2011

A CURA DI:

PIETRO CASATI  
ANDREA TOGNI

[WWW.RIFANALITICA.IT](http://WWW.RIFANALITICA.IT)  
ISSN: 2037-4445

REDAZIONE:

STEFANO CANALI  
MARCO GROSSI  
GIULIA LORENZI  
MATTIA SORGON

VOLUME 9  
NUMERO 1

2018



## Rivista Italiana di Filosofia Analitica Junior Vol. 9, n. 1

### Indice

#### Editoriale

Andrea Togni

*Editoriale* ..... 1

#### Snapshots in the Philosophy of Perception

Giulia Lorenzi and Andrea Togni

*What about Sounds in a Vacuum?* ..... 3

#### Report

Giulia Lorenzi

*Nasce a Torino la Società Italiana per le Donne in Filosofia  
(SWIP Italia)* ..... 10

#### Articoli

Francesco Beccuti

*La finitezza del mondo rispetto al tempo nella Critica della Ragion Pura* . 14

Alberto Bergamini

*Sul rapporto tra il sistema di riferimento cosale e gli altri sistemi  
di riferimento in Empiricism, Semantics and Ontology* ..... 25

Matteo Casarosa

*Alcune considerazioni sulla possibilità metafisica dei supercompiti* ..... 40

Marco Facchin

*Il progetto NOC (Natural Origins of Content): un bilancio critico* ..... 53



---

Giorgio Mazzullo	
<i>Le allucinazioni sono immagini mentali?</i> .....	78
Rodrigo Neira	
<i>Evidentialism and Transparency: a Critical Approach</i> .....	98
Andrea Raimondi	
<i>Wilfrid Sellars e la normatività del significato</i> .....	110
Marco Giovanni Stucchi	
<i>Alternative Possibilities in Frankfurt-style Cases</i> .....	136



## Editoriale

*Andrea Togni*

Care lettrici, cari lettori,

È con piacere che presentiamo il numero 9.1 di RIFAJ, il primo del 2018. La pubblicazione prevede due tratti salienti. Primo, il numero è introdotto da due rubriche curate da membri della redazione: si tratta di un breve *paper* tecnico e di un *report* di un evento filosofico. Secondo, il lettore potrà confrontarsi con diversi articoli scritti da giovani filosofi interessati alle più svariate discipline e tematiche: alcuni riguardano autori classici del calibro di Kant, Carnap e Sellars, altri si occupano di metafisica, si ritrovano incursioni nell'epistemologia, nella filosofia della mente, fino a giungere all'etica. Siamo fiduciosi che i lettori non avranno di che annoiarsi.

Il numero si apre con la nuova rubrica *Snapshot in the Philosophy of Perception*, il cui scopo consiste nel presentare in modo agile alcuni problemi curiosi riguardanti il percepire: in questo caso, i membri del comitato editoriale Giulia Lorenzi e Andrea Togni discutono alcune soluzioni filosofiche concernenti l'esistenza dei suoni nel vuoto. Giulia Lorenzi propone anche un *report* sulla neo-costituita Società Italiana per le Donne in Filosofia (SWIP), tra le cui finalità rientrano la promozione del lavoro delle filosofe in ambito accademico e non, e l'istituzione di rapporti di collaborazione con le altre società filosofiche italiane e internazionali.

La sezione *Articoli* si compone di otto lavori.

Francesco Beccuti, nel *paper* "La finitezza del mondo rispetto al tempo", indaga la coerenza delle argomentazioni kantiane sul cominciamento e la fine del mondo. A tale scopo, si serve anche del lavoro di Schopenhauer, Kemp Smith, Russell e Fried.

L'articolo di Alberto Bergamini "Sul rapporto tra il sistema di riferimento co-sale e gli altri sistemi di riferimento in *Empiricism, Semantics and Ontology*" propone un'analisi critica della posizione di Carnap in merito alla possibilità della ricerca metafisica e al suo rapporto con i sistemi di riferimento linguistici.

Nel lavoro "Alcune considerazioni sulla possibilità metafisica dei supercompiti", Matteo Casarosa mette al centro del palcoscenico i supercompiti, ovvero

i «processi fisici aventi un numero infinito di passaggi successivi e compientisi in un tempo finito».

Marco Facchin, ne “Il progetto NOC (Natural Origins of Content): un bilancio critico”, analizza i vantaggi e gli svantaggi del programma di ricerca *Natural Origins of Contents* proposto da Hutto e Satne. Il programma NOC si inserisce all’interno del paradigma enattivista in filosofia della mente, e cerca di indagare, tra le altre cose, quale ruolo svolgono i contenuti rappresentazionali nella cognizione.

L’articolo di Giorgio Mazzullo “Le allucinazioni sono immagini mentali?” si confronta con la recente proposta di Nanay, Allen e Pagondiotis di trattare le allucinazioni come appartenenti al genere delle immagini mentali.

Il *paper* “Evidentialism and Transparency: a Critical Approach” di Rodrigo Neira affronta il dibattito epistemologico sulla nozione di trasparenza. Tra i punti di riferimento dello studio rientra la posizione di Steglich-Petersen.

Andrea Raimondi, nel *paper* “Wilfrid Sellars e la normatività del significato”, la posizione di Sellars secondo cui i significati delle espressioni linguistiche naturali posseggono una componente normativa, la quale fissa «uno standard di correttezza per l’applicazione o l’uso (da parte dei parlanti) delle espressioni stesse».

Marco Giovanni Stucchi, nell’articolo “Alternative possibilities in Frankfurt-style cases”, si concentra sul tema della responsabilità etica personale e della sua dipendenza o indipendenza dalla possibilità di agire altrimenti.

La Redazione tiene molto a ringraziare gli autori per aver scelto RIFAJ come sede di pubblicazione del proprio lavoro. Ai lettori auguriamo di trovare negli articoli uno stimolo in grado di stuzzicare la loro curiosità filosofica.

*Spoiler alert!* L’attuale numero presenta numerosi e avvincenti articoli, ma il prossimo non sarà da meno. Si tratterà di una pubblicazione dedicata alla logica, in cui i lettori potranno non solo apprezzare nuovi lavori, ma anche ritrovare articoli pubblicati in passato da RIFAJ sul tema. RIFAJ è una rivista attenta al futuro della ricerca filosofica, ma non per questo dimentica della propria storia.



## Snapshots in the Philosophy of Perception What about Sounds in a Vacuum?

*Giulia Lorenzi and Andrea Togni*

**Abstract.** In the first *Snapshot in the philosophy of perception*, we tackle the question of whether there are sounds in a vacuum. The answer depends heavily on the metaphysics one is willing to accept. On the one hand, Berkeley points out that there are no empirical, perceptual grounds to settle the issue. On the other hand, realists try to provide reasons to address the question. Mechanistic philosophers believe that there are no sounds in a vacuum, because no sound waves can be found there. Casati and Dokic claim that there are sounds in a vacuum, because they are to be identified with the vibrations of the sound source. O'Callaghan denies the existence of sounds in a vacuum: he submits that the presence of a medium is a necessary condition for the existence of sounds even though he doesn't identify sounds with sound waves. In what follows, we briefly examine these approaches.

**Keywords.** Sounds, Vacuum, Auditory Perception.

In the *Three Dialogues between Hylas and Philonous in opposition to Sceptics and Atheists*, Hylas, a fictional realist philosopher who debates Berkeley's alter-ego Philonous, presents a peculiar argument about sounds: "They don't inhere in the sounding bodies. We know this, because when a bell is struck in a vacuum, it sends out no sound. So the subject of sound must be the air"<sup>1</sup>. Some contemporary philosophers add some details to the thought experiment<sup>2</sup>. For example, Casati and Dokic ask us to imagine "a vacuum jar which has the property of immediately creating a vacuum upon closing the lid and of immediately recalling air upon opening the lid. Take now a sounding object like a tuning fork at 440Hz and have it vibrate, supposing that vibration fades and become inaudible after 10 seconds. What you hear is an A that becomes feebler and feebler until it disappears. Now, place the tuning fork inside the jar, have it vibrate as before, and repeatedly open-and-close the lid of the jar, say once in a bit less than a second"<sup>3</sup>. The main difference between the two thought experiments is that Hylas focuses on the waves traveling in the air, while Casati and Dokic focus on the vibrating tuning fork. This difference depends mainly on historical reasons. In the eighteenth century, the standard mechanistic view was that sounds are vibrations of the medium (air or water) between emitting bodies and listeners, while contemporary distal theories give a prominent role to the resonating source.

According to Hylas and the mechanistic view, there are no sounds in a vacuum, because no vibrations of the medium can be found there. Berkeley rejects Hylas' statement. First, Philonous points out that it is curious to maintain that sounds are vibrations, because vibrations are normally perceived thanks to the senses of touch and sight, not through hearing. This is not a very strong argument. Although it is true that some mechanistic approaches describe vibrations as the proper objects of both hearing and touch, they are compatible with the idea that the two sensory modalities can be distinguished by the different sense organs that elaborate those vibrations, namely, the ears and the skin. To be sure, Berkeley's real target is the thesis according to which sounds may exist independently of any perceiver<sup>4</sup>. Indeed, his metaphysics is based on the principle that we are allowed to claim that something exists only in the presence of empirical, perceptual evidence. To understand this point, we have to approach the second argument offered by Philonous against Hylas. Berkeley rejects the distinction endorsed by Hylas between real sounds, which may exist even if nobody

<sup>1</sup>Berkeley (1713, p. 8).

<sup>2</sup>They are *thought* experiments because no human being has created or encountered a perfect vacuum so far.

<sup>3</sup>Casati and Dokic (2010).

<sup>4</sup>For a more detailed exposition of Berkeley's arguments against materialism and realism, see Togni (2017).

perceives them, and sounds as we perceive them. Of course, Berkeley cannot *prove* that non-perceived sounds don't exist, but he correctly points out that the burden of the proof is on his opponent to show that there are such "real", possibly non-perceived sounds and that those "real" sounds are lacking in a vacuum. Berkeley's argument holds because it is the realist who is affirming the positive thesis that sounds can exist independently of any perceiver. As shown by the question about the existence of sounds in a vacuum, realists try to develop a theoretical view of the metaphysical existence of sounds (and of other perceptual entities or events), but pay less attention to what is perceptually experienced. While Hylas claims that there are no sounds in the absence of a vibrating medium, Philonous and Berkeley submit that *it is not possible to know* whether there are sounds in a vacuum, because we lack empirical evidence to answer the question<sup>5</sup>. Berkeley holds the straightforward view that we cannot say anything about the metaphysical existence of sounds that are not heard. Of course, it is possible to study what happens to the bell, what is the relation between its vibrations and the vacuum, and so on, but these considerations don't say too much about the existence or non-existence of *sounds* in a vacuum: this last point can be settled only at the empirical level, which, in the thought experiment submitted by Hylas, is lacking. Thus, Berkeley's view is not that Hylas' mechanistic position on the non-existence of sounds in a vacuum is wrong, but that it cannot be empirically verified or falsified.

While Berkeley doesn't discuss theories that locate sounds near the vibrating body, it is fair to assume that he would move against them the same criticisms that he moves against Hylas. To claim or deny that a vibrating body emits a sound in a vacuum is to try to address the metaphysical question about the existence of sounds theoretically rather than empirically. From a Berkeleyan perspective, the fact that we *know* that a body is vibrating in a vacuum doesn't tell anything about the existence or non-existence of a *sound* that is not *heard*.

Some contemporary philosophers who defend a distal theory of sounds as events located at their source<sup>6</sup> reject Berkeley's stance and try to give reasons to accept or deny the idea according to which there are sounds in a vacuum. This is clearly shown by the debate between Casati and Dokic on the one hand, and O'Callaghan on the other.

Casati and Dokic consider the presence of the air, or of another medium, as a non-necessary condition for the existence of a sound. Even if the air is required to bring information from the vibrating object to the ears and to reveal

<sup>5</sup>First, a perfect vacuum has never been created. Second, if a perfect vacuum were created, one would hear sounds or she wouldn't: whatever the outcome, philosophers wouldn't be allowed to draw theoretical conclusions regarding "real" sounds that may exist in the absence of perceivers. According to Berkeley, giving the senses the last word is the best strategy when it comes to metaphysics. Thus, if sounds were heard in a vacuum, then they would exist as perceived sounds.

<sup>6</sup>In this paper, we don't address proximal and a-spatial theories of sounds.

sounds to perceivers, it is not essential for the existence of sounds<sup>7</sup>. Instead, O'Callaghan supports a relational event theory: sounds are events where both the sources and the surrounding medium are necessary components. He claims that "a sound is the event of an object or interacting bodies disturbing a surrounding medium in a wavelike manner"<sup>8</sup>.

To stress the difference between their position and O'Callaghan's, Casati and Dokic submit the above-mentioned variation of Berkeley's thought experiment. Their version is more complex and involves two extra elements that are perceptually relevant: the repetition of the open-and-close motion of the lid of the jar and the temporal length of the action. Casati and Dokic suggest that in this case perceivers would have the impression of a revealed sound that persists unheard when the lid is closed. Thus, O'Callaghan's metaphysical position<sup>9</sup> appears counter-intuitive because it implies that the sound of the tuning fork pops into and out of existence every time the lid is opened and closed.

To support their argument, Casati and Dokic articulate three visual analogies<sup>10</sup>. First, they consider a coloured object that is put into a dark room and then brought to the light: normally, it looks like colours are revealed to perceivers. Second, they compare the tuning fork in the jar to other two cases of visual tunnel effect: a circle moving behind a screen and the switching on-and-off of the light in a refrigerator when we open and close its door. Their point is that, if there is nothing wrong in assuming the existence of unperceived visual objects, the same can be said regarding sounds in a vacuum that are not actually heard. Furthermore, they consider the vacuum in the jar as like other ordinary unfavourable conditions of hearing: those adverse circumstances don't affect the nature and the existence of the perceived entities. Hence, the lack of information, deriving from the absence of the medium, doesn't involve the non-existence of a sound.

On the contrary, O'Callaghan's goal is to show that the presence of a medium is not only a condition for veridical sound perception, but also for the existence of sounds themselves. His reasoning<sup>11</sup> can be reconstructed as follow: (i) Casati and Dokic's visual analogies do not work; (ii) A better visual analogy about indistinguishable objects leads to conflicting intuitions; (iii) Those intuitions are not useful at the metaphysical level, but only at the perceptual level; (iv) The functioning of perception is not the issue here; (v) The absence of audible qualities in a vacuum provides philosophical reasons to maintain that the presence of a medium is a necessary condition for the existence of sounds.

<sup>7</sup>See Casati and Dokic (1994, pp. 44-46, 61-64). See also Casati and Dokic (2009, ch. 5).

<sup>8</sup>See O'Callaghan (2007, p. 110).

<sup>9</sup>This is true also for the medial waves theories.

<sup>10</sup>See Casati and Dokic (1994, pp. 42-44), Casati and Dokic (1998, pp. 36-37), Casati and Dokic (2010).

<sup>11</sup>See O'Callaghan (2007, pp. 49-55).

O'Callaghan supports (i) by pointing out that sounds, unlike colours, do not appear as features or properties of objects that persist unchanged in time. In addition, he stresses that silence ordinarily signals the conclusion of a sound and allows the beginning of another. Instead, darkness seems to be a barrier behind which objects and their colours can continue to exist. Obscurity is more akin to a sufficiently intense sound that does not allow us to hear anything else. Also, O'Callaghan suggests to compare the case of the tuning fork in the jar with a situation where an object disappears completely from a perceiver's visual field and, then, an indiscernible copy of it reappears in front of her. Given that this situation does not lead to univocal reactions, O'Callaghan can find support for (ii). It is also possible to argue that the moving circle and the refrigerator light pop into and out of existence, regardless of the fact that they are *perceived* as continuously present behind the white screen and the closed door. Even if we consider these two cases good analogues of the jar situation, O'Callaghan's position that sounds do not exist in a vacuum is not disproved despite Casati and Dokic's opposite perceptual intuitions.

From what we have stated above, it is possible to conclude that studying the reactions to various visual and auditory situations can say something about the functioning of perception, but not so much about the ontological nature of the entities that are perceived (iii). O'Callaghan, like Berkeley, recognises that it is not possible to obtain empirical refutations or confirmations of the presence of sounds in a vacuum, but, unlike Berkeley, he tries to find reasons to deny their existence by focusing on the metaphysical level (iv). For this reason, he focuses on the demonstration of the necessity of a medium, regardless of Casati and Dokic's thought experiment.

According to the standard physical theory and to some common understanding of sounds, the qualities (pitch, timbre and loudness) of sounds depend on the surrounding medium: this implies that a vibrating event without a surrounding material does not present any audible qualities (v). Unless a special category of sounds that has no standard qualities exists, the tuning fork in a vacuum emits no sounds. While the light doesn't change objects' visual qualities, the medium seems to play a more substantial role in defining sounds' properties. This is shown by the fact that it is possible to specify standard visual conditions (the full spectrum of light) that reveal the genuine colours of objects, while a similar set of standard conditions cannot be found for the auditory world<sup>12</sup>. It is important to point out that the necessity of the medium for the existence of sounds does not imply that they exist entirely in the surrounding materials. For this reason, O'Callaghan's theory of sounds does not collapse into a mechanistic medial theory.

Even if the case of sounds in a vacuum is considered a classic argument against

<sup>12</sup>O'Callaghan would have done well to provide more details about these last two points.

mechanistic theories of sounds and it creates problems for a hybrid theory that take into account both the source and the surrounding medium, it cannot be decisive from a metaphysical point of view. Nevertheless, Berkeley's thought experiment and its latter-day versions remain an interesting field where theories and intuitions about hearing can be tested.



## References

- Berkeley, George (1713). *Three Dialogues between Hylas and Philonous in opposition to Sceptics and Atheists*. URL: <http://www.earlymoderntexts.com/assets/pdfs/berkeley1713.pdf>.
- Casati, Roberto and Jérôme Dokic (1994). *La philosophie du son*. Nîmes: Editions Jacqueline Chambon.
- (1998). “Lo spazio del suono”. In: *Rivista di Estetica* 9.
- (2009). “Some varieties of spatial hearing”. In: *Sounds & Perception. New Philosophical Essays*. Ed. by Matthew Nudds and Casey O’Callaghan. Oxford: Oxford University Press, Chapter 5.
- (2010). *Sounds*. Ed. by Edward N. Zalta. The Stanford Encyclopedia of Philosophy. URL: <https://plato.stanford.edu/entries/sounds/>.
- O’Callaghan, Casey (2007). *Sounds. A Philosophical Theory*. Oxford: Oxford University Press.
- Togni, Andrea (2017). *Sulla via di San Tommaso Apostolo: Berkeley su realismo e scetticismo*. Ed. by E. Baccarini. Dialeghesthai. URL: <https://mondodomani.org/dialeghesthai/atog01.htm>.



## Nasce a Torino la Società Italiana per le Donne in Filosofia (SWIP Italia)

[Torino, 14 giugno 2018]

*Giulia Lorenzi*

Lo scorso 14 giugno a Torino alle ore 17, al termine del simposio dedicato all'ultimo libro di Kate Manne *Down Girl: The Logic of Misogyny*<sup>1</sup>, ha avuto luogo l'Assemblea Costituente della SWIP Italia: Società Italiana per le Donne in Filosofia.

Vera Tripodi, Laura Caponetto e Federica Berdini, ideatrici e promotrici del progetto, hanno aperto, presieduto e coordinato i lavori della Costituente, alla cui realizzazione hanno contribuito, collaborando in varie fasi dell'iniziativa, anche Marina Sbisà, Claudia Bianchi, Carla Bagnoli e Clotilde Calabi. L'aiuto della linguista Manuela Manera ha inoltre permesso una formulazione corretta in un'ottica di genere dei documenti da sottoporre all'attenzione dell'assemblea. Fra le attività principali della Costituente vi sono state: la discussione e la ratifica dello Statuto della Società, l'elezione del primo Consiglio Direttivo e la distribuzione delle cariche all'interno di quest'ultimo. Segnaliamo inoltre che è stato possibile partecipare all'evento tramite teleconferenza, modalità che ha permesso a diversi interessati di prendere agevolmente parte all'iniziativa.

La Società Italiana per le Donne in Filosofia nasce con uno Statuto, votato in sede assembleare, che prevede due categorie di appartenenza: quella delle 'socio' e quella degli 'amici'. Secondo quanto sancito dal documento statutario, la prima categoria accoglierà 'chiunque si identifichi come donna' (clausola restrittiva la cui validità è prevista fino alla raggiunta parità di genere in ambito accademico), mentre la seconda sarà destinata a coloro che vorranno sostenere la Società pur non riconoscendosi nella definizione precedente.

La SWIP Italia si prefigge come obiettivi quelli di sostenere, promuovere e diffondere il lavoro delle donne in filosofia, nonché di favorirne l'interazione. Come chiaramente espresso nell'articolo 2 comma 1 dello Statuto, intenti della Società sono altresì la sensibilizzazione della comunità accademica riguardo la

<sup>1</sup>Manne (2017)

contenuta presenza di donne in filosofia e il contrasto delle discriminazioni di genere all'interno di questo ambito di ricerca.

Allo scopo di perseguire tali finalità, la Società Italiana per le Donne in Filosofia intende attivare, nel prossimo futuro, una serie di iniziative e progetti. Fra i propositi vi è quello dell'avvio di un programma volto al sostegno delle giovani studiose tramite il loro affiancamento ad un mentore. Tale intenzione è stata ispirata dall'esistenza, nel Regno Unito, di un simile sistema di tutorato ideato dalla SWIP UK e volto, fra gli altri scopi, ad impedire l'isolamento delle donne in un ambiente a prevalenza maschile. Un ulteriore progetto è destinato alla creazione di collaborazioni con altre società filosofiche italiane, per esempio, tramite la proposta di ospitare, all'interno dei loro convegni e conferenze, sessioni SWIP Italia dedicate. È altresì desiderio della Società Italiana per le Donne in Filosofia il proporre alle SWIP "sorelle" l'organizzazione di eventi congiunti. Infine, vi è l'intenzione di raccogliere dati riguardanti la presenza delle donne in filosofia in Italia, allo scopo di ottenere una visione completa e fedele delle discriminazioni di genere in quest'area di ricerca. Attualmente in corso è invece l'organizzazione congiunta di un Workshop dal tema "Le donne filosofe dall'antichità ai giorni nostri" che si terrà presso l'Università degli Studi di Bergamo i prossimi 13 e 14 dicembre.

Il primo Consiglio Direttivo della Società, in carica per i prossimi tre anni, sarà guidato dalla presidenza di Marina Sbisà, professoressa di prima fascia presso l'Università degli Studi di Trieste, coadiuvata nel suo ruolo dal vicepresidente Vera Tripodi, assegnista di ricerca presso l'Università di Torino. Le cariche di segretaria e tesoriera sono state invece affidate a Laura Caponetto, dottoranda presso l'Università Vita e Salute San Raffaele, e Federica Berdini, attualmente *part-time lecturer* presso la Ryerson University e successivamente impegnata presso la Ludwig-Maximilians-Universität München. Lo spirito con il quale si è proceduto nella votazione e nella ripartizione dei ruoli all'interno del direttivo è stato quello di coinvolgere rappresentanti della comunità filosofica appartenenti a fasce accademiche diverse e di affiliazione geografica ed istituzionale diversificata. Il direttivo attualmente in carica coinvolge così studiose impegnate in fasi diverse della carriera accademica e di provenienza geografica dissimile, restituendo un quadro eterogeneo che permette la rappresentanza del più ampio numero possibile di realtà vissute dalle donne impegnate in filosofia in Italia.

Tutt'altro che un unicum, la società italiana nasce inserendosi in una rete già consistente di associazioni affini, considerabili ramificazioni della *Society for Women in Philosophy* fondata nel 1972. Numerose sono infatti le società per le donne in filosofia costituite in varie regioni del pianeta, ognuna delle quali presenta proprie specificità inerenti agli ambiti di ricerca promossi e alle realtà territoriali in cui è sorta e si è sviluppata. Tra le società "sorelle" possiamo sicuramente menzionare la SWIP UK, il cui Statuto ha ispirato quello della neo-nata

società italiana. Altra realtà attiva in questo contesto è la SWIP Ireland, la cui sesta conferenza annuale tenutasi lo scorso maggio ha visto la partecipazione di più di una cinquantina di donne impegnate in ambito filosofico. La SWIP Analytic, invece, fondata nel 2013 nell'area newyorkese, ha come principale interesse la promozione dell'attività di ricerca delle donne in ogni settore della filosofia: dalla metafisica alla filosofia del linguaggio, dalla logica alla filosofia della mente. Fra le SWIP di più recente fondazione possiamo ricordare la SWIP Switzerland, impegnata in intensi programmi di seminari su tematiche riguardanti femminismo e differenze di genere, e la SWIP Turkey, la cui prima conferenza annuale si è svolta a Bilkent lo scorso ottobre.

In conclusione, a chiunque fosse interessato all'operato della SWIP Italia, segnaliamo il sito internet ufficiale ([www.swip-italia.org](http://www.swip-italia.org)) contenente tutte le informazioni concernenti l'identità e le attività della Società Italiana per le Donne in Filosofia. Indicazioni sulle modalità di associazione, nonché un archivio che raccoglie i verbali delle Assemblee delle Socie, lo Statuto della Società e l'Atto Costitutivo firmato dalle Socie fondatrici, sono altrettanto visualizzabili nel medesimo spazio web.

A chi volesse ricevere ulteriori informazioni segnaliamo anche l'indirizzo email: [swip.italia@gmail.com](mailto:swip.italia@gmail.com).

## **Riferimenti bibliografici**

Manne, Kate (2017). *Down Girl: The Logic of Misogyny*. Oxford: Oxford University Press.





# La finitezza del mondo rispetto al tempo nella Critica della Ragion Pura<sup>1</sup>

*Francesco Beccuti*

**Abstract.** Esaminerò gli argomenti kantiani esposti nella dialettica trascendentale per la necessità di un cominciamento temporale del mondo (tesi della prima antinomia) e per la necessità dell'assenza di un cominciamento del mondo (antitesi della prima antinomia). So-sterrò, servendomi di argomenti già di Arthur Schopenhauer, Kemp Smith, Bertrand Russell e Milton Fried, che l'argomento kantiano per la tesi è interamente da rigettare. Viceversa lo stesso non può dirsi dell'argomento per l'antitesi, che risulta logicamente valido: ipotizzare un inizio del mondo è contraddittorio. Concluderò infine con alcune riflessioni a proposito del simmetrico problema della fine del mondo, la quale, se intesa come un passaggio dall'essere al nulla, è anch'essa contraddittoria. Alternativamente, essa potrebbe configurarsi, seguendo Schopenhauer (e una interpretazione olistica del secondo principio della termodinamica), come l'ingresso del mondo in una fase definitiva di stasi.

**Keywords.** Filosofia del tempo, Cosmologia razionale, Kant.

<sup>1</sup> Ringrazio Brunello Lotti e un revisore anonimo che hanno contribuito con i loro commenti a migliorare significativamente questo articolo.

È sempre più facile mostrare i difetti e gli errori nell'opera di un grande spirito che non dare una chiara e completa spiegazione del suo valore.

(Schopenhauer 2006, p. 1586)

Mi occuperò in questo breve intervento della prima parte della prima antinomia presentata da Immanuel Kant nella sua *Critica della Ragion Pura*, ossia del problema dell'infinità del mondo rispetto al tempo<sup>2</sup>. Tratterò dunque nelle prime due sezioni, a partire da Kant e da alcuni suoi critici, gli argomenti formulati a proposito della seguente perenne domanda filosofica:

Il mondo ha avuto inizio oppure è sempre esistito?

Nella conclusione mi occuperò invece brevemente del simmetrico problema dell'esistenza di una fine del mondo, da Kant completamente ignorato. La riflessione sulle antinomie gioca in effetti un ruolo fondamentale nell'opera del filosofo di Königsberg, il quale, attraverso una doppia *reductio ad absurdum*, tentò di mostrare l'impossibilità di asserire sia che il mondo abbia avuto un inizio sia che il mondo non l'abbia avuto. A partire da ambedue le premesse, infatti, si giungerebbe, secondo Kant, in entrambi i casi ad un assurdo ed il riconoscimento di questo conflitto aprirebbe la via indiretta per la dimostrazione dell'idealità trascendentale di tutti i fenomeni (in aggiunta alla via diretta presentata nell'estetica trascendentale)<sup>3</sup>. "Sarebbe dunque possibile provare due proposizioni mutualmente esclusive, ossia che il mondo abbia avuto inizio nel tempo e che non l'abbia avuto". E allora "Kant conclude che l'unica soluzione possibile del problema è la sua soluzione critica". Tuttavia, come vedremo, tale contraddizione deriva in realtà da ambiguità semantiche e da "una procedura logica che prima semina contraddizioni e poi simula sorpresa" al loro fiorire. Dunque "le dimostrazioni delle antinomie non sono altro che un tour de force filosofico che esibisce al più una serie di manovre linguistiche" (Fried 1940, pp. 206-207)<sup>4</sup>.

## 1 La tesi dell'antinomia

La correttezza del severo giudizio di Fried è soprattutto evidente se si considera il primo corno dell'antinomia. Ripercorriamo l'argomento supponendo con Kant che il mondo non abbia avuto un inizio nel tempo. Da ciò segue necessariamente che, sino ad un qualsiasi istante dato, è passata un'infinità di tempo, e

<sup>2</sup>Tralascierò dunque quella parte dell'antinomia che tratta dell'infinità del mondo rispetto allo spazio.

<sup>3</sup>In breve, se il mondo (la serie complessiva delle apparenze) esistesse come cosa in sé, allora esso dovrebbe essere finito o infinito. Dato che si può dimostrare che deriva una contraddizione sia assumendone la finitezza sia assumendone l'infinitezza, allora il mondo non esiste come cosa in sé, ma solo come ideale trascendentale (Gram 1967, p. 500).

<sup>4</sup>La traduzione del passo di Fried (e dei successivi estratti in lingua inglese) è mia.

dunque che è trascorsa (fino all'istante presente, o fino a qualsiasi altro istante nel passato o nel futuro) una serie infinita di stati di cose. Tuttavia, osserva Kant, vi sarebbe una contraddizione logica fra il definire una serie infinita e definirla già trascorsa:

Si supponga infatti, che il mondo non abbia alcun inizio nel tempo: in tal caso, sino ad un qualsiasi istante è passata un'eternità, e quindi è trascorsa nel mondo una serie infinita di stati susseguentisi delle cose. Peraltro l'infinità di una serie consiste proprio nel fatto che quest'ultima non può mai essere completata mediante una sintesi successiva. Dunque, è impossibile un'infinita serie del mondo già trascorsa, e quindi, un inizio del mondo è una condizione necessaria della sua esistenza (Kant 1976, B294).

Come molti commentatori hanno notato, da Arthur Schopenhauer fino a Kemp Smith, passando per Bertrand Russell, la dimostrazione della tesi è errata ed è annoverabile storicamente fra quelle fallacie logiche commesse dai filosofi (anche i più grandi) quando l'infinito ha fatto capolino fra i loro ragionamenti. In poche parole, Kant confuse la mancanza di principio di una serie con la sua mancanza di fine. Nonostante alcuni tentativi dei commentatori di recuperare l'argomento kantiano, mi sembra del tutto evidente che esso non possa essere punto salvato e che dunque non possa essere salvata la succitata 'seconda via' indiretta per la dimostrazione dell'idealità dei fenomeni. Non vi è infatti alcuna contraddizione logica intrinseca nel pensare una serie senza cominciamento. Come già notò Schopenhauer<sup>5</sup>:

[...] invece della mancanza di principio della serie degli stati, di cui si era prima parlato, viene introdotta improvvisamente la loro mancanza di fine (infinità), e viene poi dimostrato, cosa che nessuno mette in dubbio, che a questa contraddice logicamente l'esser compiuta e che tuttavia ogni presente è la fine del passato. Ma si può sempre pensare la fine di una serie senza principio, senza recare pregiudizio alla sua mancanza di principio; come anche si può pensare, viceversa, il principio di una serie senza fine (Schopenhauer 2006, pp. I586-I587).

E come rimarcò ancor più chiaramente Russell:

Che cosa è accaduto nella sua [di Kant] immaginazione fu una cosa di questo genere: partendo dal presente e andando indietro nel tempo, abbiamo, se il mondo non ha avuto inizio, una serie infinita di eventi. Come è chiaro dalla parola 'sintesi', egli immaginò una

<sup>5</sup>Il quale, circa le quattro antinomie, sostenne che le dimostrazioni delle tesi non sono altro che sofismi, mentre le affermazioni delle antitesi sono tutte vere. Cfr. la bella esposizione della questione in D'Iorio (1995, pp. 76-83).

mente nel tentativo di afferrarli successivamente, *nel senso contrario* rispetto al senso in cui essi sono avvenuti, ossia dal presente al passato. Questa serie non ha chiaramente fine. Ma la serie di eventi fino all'istante presente ha una fine, dato che finisce nel presente. A causa dell'invecchiato soggettivismo proprio della sua abitudine mentale, egli non notò che aveva invertito il verso delle serie, sostituendo la sintesi all'indietro con l'accadere in avanti, e dunque egli suppose che fosse necessario identificare la serie mentale, che non ha fine, con la serie fisica, dotata di fine ma non di inizio. Fu questo errore che, credo, operando inconsciamente, lo portò ad attribuire validità ad una così singolarmente fragile porzione di ragionamento fallace (Russell 1915, p. 157).

In effetti, per i matematici contemporanei non v'è alcuna contraddizione nel pensare un'infinità, per così dire, aperta a sinistra e chiusa a destra. La successione ordinata dei numeri interi non positivi

$$\dots - 3, -2, -1, 0$$

è forse il più semplice esempio di successione senza elemento minimo. Un altro esempio elementare è la successione dei reciproci degli interi positivi:

$$\dots, \frac{1}{3}, \frac{1}{2}, 1.$$

Tali successioni vengono usate dai matematici nei loro ragionamenti senza cadere punto in contraddizione<sup>6</sup>.

Tuttavia, anche supponendo che Kant non abbia confuso, per così dire, l'indietro con l'avanti, rimangono comunque due fondamentali critiche al ragionamento impiegato per dimostrare la tesi:

1. L'introduzione surrettizia della soggettività trascendentale in un ragionamento che, invece, dovrebbe essere dogmatico in senso metafisico, in quanto servirebbe a Kant proprio per negare la validità dei ragionamenti dogmatici esemplificati dalle antinomie. In altre parole, Kant trasforma una impossibilità di esaurire un percorso di una serie di successivi stati di apprendimento soggettivo (piano epistemico) in una impossibilità di esistenza oggettiva della serie stessa (piano ontologico).
2. Pur ammettendo la compatibilità del ricorrere alla soggettività trascendentale con l'originaria intenzione kantiana, non è comunque evidente

<sup>6</sup>Un esempio più complesso è il seguente: all'interno di un modello non-standard dell'aritmetica (Kaye 1991, pp. 10-13), alcuni elementi (detti appunto non-standard) hanno la peculiarità di possedere infiniti predecessori. Un altro esempio è invece quello presupposta da un modello della teoria degli insiemi con la negazione dell'assioma di fondazione (Drake 1974, pp. 8-9), contenente insiemi che appartengono a sé stessi o che hanno sé stessi come sottoinsiemi:  $\dots A \in A \in A \dots$  oppure  $\dots A \in B \in C \in A \in B \in C \in A \dots$

che la sintesi mentale dei successivi elementi di una serie non possa effettivamente essere completata anche solo sul piano epistemico.

In effetti, riguardo al secondo punto, “Se si potesse leggere sempre più veloce, ossia se si potesse dire [o pensare] ‘zero’ in un secondo, ‘meno uno’ in mezzo secondo, ‘meno due’ in un quarto di secondo, ‘meno tre’ in un ottavo di secondo, e così via, allora tutti i nomi [dei componenti della serie] potrebbero essere letti in due secondi soltanto”<sup>7</sup>. Dunque l’argomento di Kant presuppone un’ulteriore ipotesi limitativa (rispetto alla mente), ossia deve presupporre che “lo stesso intervallo di tempo deve essere impiegato per leggere [o pensare] ogni numero” (Fried 1940, pp. 211-212). Un ulteriore assunto, dunque, si nasconde nel ragionamento kantiano: che la serie successiva degli stati del mondo debba essere pensata (e non solo accaduta) ad intervalli di tempo uguali. Il ragionamento per la dimostrazione della tesi è quindi errato.

## 2 L’antitesi dell’antinomia

Supponiamo in secondo luogo con Kant che il mondo abbia avuto un inizio in un determinato istante di tempo. Ne segue che fino a quell’istante, non v’era nulla. Ma un qualcosa (il mondo) non può sorgere dal nulla (assenza del mondo), da cui la contraddizione:

Poniamo infatti che esso [il mondo] abbia un inizio. In quanto l’inizio è un’esistenza, preceduta da un tempo in cui la cosa non è, deve esserci già stato in precedenza, un tempo in cui il mondo non era, cioè un tempo vuoto. In un tempo vuoto, peraltro, non è assolutamente possibile che sorga una qualsiasi cosa [...] Di conseguenza, nel mondo possono bensì prendere inizio parecchie cose, ma il mondo stesso non può avere alcun inizio, ed è perciò infinito rispetto al tempo passato (Kant 1976, B295).

Tale dimostrazione si basa sull’antico precetto logico *ex nihilo nihil*: se ad un certo punto non v’era nulla allora non sussistono le condizioni causali necessarie perché si generi qualcosa. Inoltre la dimostrazione presuppone una qual certa primarietà ontologica della dimensione temporale rispetto alla dimensione spaziale. Tale primarietà del tempo rispetto al mondo è del resto nascosta nella domanda iniziale. Nel chiedersi se il mondo ha avuto inizio nel tempo oppure no, si suppone già implicitamente che il tempo abbia in qualche modo una

<sup>7</sup>Si noti *en passant* che questa è la soluzione dei paradossi eleatici fornita dall’analisi matematica moderna: per chi li ricorda diremo semplicemente che la freccia raggiunge il bersaglio perché attraversa ogni successiva metà del percorso a velocità sempre raddoppiantesi.

preminenza ontologica sul mondo: si suppone cioè che il mondo esista prima di tutto *nel tempo*<sup>8</sup>.

Ad ogni modo, il ragionamento per dimostrare l'antitesi, per quanto stringato, è, a parer mio, del tutto corretto ed inattaccabile dal punto di vista strettamente logico. Tuttavia, come ha sostenuto Kemp Smith, la validità dell'argomento dipende dal fatto che per 'mondo' intendiamo 'tutto ciò che esiste' e non 'il mondo materiale', giacché, sotto quest'ultima ipotesi, altre esistenze non materiali potrebbero fornire le condizioni causali sufficienti per la creazione del mondo<sup>9</sup>.

Nel caso in cui, invece, assumessimo che il 'mondo' nel testo kantiano sia posto a significare 'tutto ciò che esiste', Smith lamenta comunque la fallacia dell'argomento: si tratterebbe di assumere un'impossibilità (l'inesistenza di tutto ciò che esiste) e derivarne retoricamente un assurdo già implicito nella prima assunzione impossibile:

Se, d'altro canto, il mondo per Kant vuol dire 'tutto ciò che esiste', allora l'assunzione della sua non-esistenza significa l'assunzione della non-esistenza di tutte le sue possibili cause. ... Da questo punto di vista, l'argomento consiste nel fare un'assunzione impossibile, e, successivamente, far notare la conseguenza impossibile che ne deriva necessariamente (Smith 1918, pp. 487-488)

Anche se le cose stessero così, ossia se il ragionamento kantiano fosse solo retorico e l'impossibilità derivata dal ragionamento fosse già implicita nell'assunzione iniziale, lo stesso riconoscimento dell'assurdità dell'assunzione mostrerebbe comunque la difficoltà di pensare il mondo (inteso come l'insieme di tutto l'esistente) come iniziato. Al contrario, lo stesso non sembra valere per la prima *reductio* formulata da Kant a partire dall'ipotesi di mancanza d'inizio del mondo: non pare esservi alcuna impossibilità logica nel pensare il mondo come non-iniziato.

<sup>8</sup>Si noti che esiste un'importante tradizione di pensiero risalente forse a Filone di Alessandria secondo cui Dio avrebbe creato, insieme con il mondo, anche il tempo (Philo: 26). Tale tradizione (che, comunque, distingue fra Dio e il mondo come due entità distinte e non intende 'mondo' come 'tutto ciò che esiste'; vedi sotto) trova in Agostino d'Ipbona il suo più influente rappresentante (Augustinus 1997, p. 6).

<sup>9</sup>Cfr. anche Fried (1940, p. 216): "Ne segue che 'un tempo in cui il mondo non era' difficilmente è un tempo vuoto. Al più, è un tempo non contenente il mondo sensibile. Non è necessario, come egli [Kant] suggerisce, che sia vuoto d'ogni contenuto. I metafisici, infatti, potrebbero sostenere che ha un contenuto temporale [...] che potrebbe aver funzionato originariamente come sorgente creatrice del mondo".

### 3 Conclusione: il problema della fine del mondo

La dimostrazione della tesi è, dunque, proprio una “dimostrazione avvocatessa” del genere che Kant voleva evitare perché si basa su di “un concetto difettoso dell’infinità” (Kant 1976, B298) ed è quindi interamente da rigettare. Viceversa la dimostrazione dell’antitesi merita, per quanto detto, di essere presa seriamente in considerazione anche dalla riflessione dei contemporanei: partendo da ipotesi relativamente deboli (il principio logico *ex nihilo nihil*) essa giunge a conclusioni indubbiamente forti: il tutto (o almeno una parte di esso) è sempre esistito.

Ci chiediamo ora brevemente, in conclusione, se lo stesso ragionamento applicato da Kant nella dimostrazione dell’antitesi<sup>10</sup> possa essere applicato anche alla domanda (simmetrica rispetto a quella da cui siamo partiti):

Il mondo avrà una fine oppure esisterà sempre?

Da un punto di vista strettamente logico, la risposta è certamente positiva: come dal nulla non si può generare un qualcosa, anche da un qualcosa non può generarsi un nulla. Dunque, la fine del mondo intesa come l’ingresso in uno stato di non-esistenza di ciò che prima era esistente è anch’essa un’impossibilità. Questo era ad esempio il parere degli antichi epicurei<sup>11</sup> che ponevano il principio di impossibilità di generazione di un qualcosa dal nulla e, simmetricamente, del nulla da un qualcosa (neanche per effetto di un intervento divino) come base della loro filosofia<sup>12</sup>. Si deve dunque pensare che essi credessero in un universo increato ed eterno, senza inizio né fine temporale. Tale era sicuramente anche la dottrina degli stoici, alcuni dei quali sostenevano la tesi di un universo senza inizio né fine: ciclico<sup>13</sup> come l’universo immaginato nella da Friedrich Nietzsche, il quale si spinse addirittura a darne una dimostrazione matematica dal forte sapore kantiano (Nietzsche, 1888: p. 376). Schopenhauer, invece, pur esplicitamente favorendo l’ipotesi di un mondo senza principio né fine nel tempo<sup>14</sup> sostenne tuttavia la possibilità che il mondo possa, in un certo senso, avere una fine. Il passaggio de *Il Mondo come Volontà e Rappresentazione* citato nella prima sezione, infatti, conclude che:

<sup>10</sup>Inutile chiedersi la stessa cosa per il fallace ragionamento della tesi, basato sulla impossibilità di una serie infinita già trascorsa, e che dunque non preserverebbe in questo caso neppure la sua efficacia retorica.

<sup>11</sup>Ma anche di alcuni presocratici, come ad esempio Melisso (Diels e Kranz 1985, p. 30), Empedocle (Diels e Kranz 1985, pp. 31; 11), Anassagora (Diels e Kranz 1985, pp. 59; 17) e Democrito (Laertius 1999, pp. IX; VII; 4).

<sup>12</sup>Cfr. Lucretius (2013, pp. 149-150) e Persius (2007, pp. III, 83-84).

<sup>13</sup>Cfr. Cicero (2011, p. 118).

<sup>14</sup>Almeno nella rappresentazione. Cfr. Schopenhauer (2006, pp. I325, I330, I335, I366, I416, I432-433).

[...] si può sempre pensare la fine di una serie senza principio, senza recare pregiudizio alla sua mancanza di principio; come anche si può pensare, viceversa, il principio di una serie senza fine.

Tuttavia, continua Schopenhauer,

[...] niente viene portato contro l'argomento veramente giusto dell'antitesi, che i mutamenti del mondo all'indietro presuppongono assolutamente e necessariamente una serie di mutamenti infinita

Dunque Schopenhauer è convinto della contraddittorietà dell'esistenza di uno stato iniziale di esistenza del mondo rispetto al tempo. Viceversa, nel passaggio immediatamente successivo Schopenhauer sembra ammettere che possa esistere una sorta di stato finale nella serie di stati di cose del mondo:

Possiamo pensare la possibilità che la serie causale finisca una volta per tutte in una stasi assoluta, ma nient'affatto la possibilità di un cominciamento assoluto (Schopenhauer 2006, p. I587).

Schopenhauer tuttavia non fornisce un argomento per questa asserzione, la quale pare affermare che sia possibile pensare a una fine del mondo, che tuttavia non coinciderebbe con il totale annullamento del mondo (come passaggio, cioè, dall'essere al non-essere).

Un metodo (non l'unico) per spiegare l'affermazione di Schopenhauer e per trovarne una singolare assonanza con la fisica contemporanea (o di poco successiva), potrebbe essere quello di estendere come segue l'ontologia kantiana con la terminologia della fisica newtoniana. Se da un lato è chiaro, come abbiamo rilevato, che un qualcosa (un sistema di particelle in movimento) non può evolversi in un nulla (un sistema vuoto, ossia privo di particelle), tuttavia esso potrebbe invece evolversi verso un altro genere di stato finale (un sistema di particelle in completa quiete). Tale stato sarebbe a pieno titolo una 'fine del sistema', in quanto dall'assoluta assenza di moto non potrebbe mai generarsi una successiva quantità di moto, se non per effetto di forze esterne al sistema<sup>15</sup> In effetti i risultati della termodinamica classica sembrano confermare l'ipotesi di una fine dell'universo (del mondo) per assenza di moto. Come scrisse William Thomson, uno dei padri fondatori della disciplina,

La seconda grande legge della termodinamica coinvolge un certo principio di azione irreversibile nella Natura. Infatti si può dimostrare che, anche se l'energia meccanica è indistruttibile, c'è una tendenza universale alla sua dissipazione, che produce un graduale allarga-

<sup>15</sup>Qui usiamo di nuovo esplicitamente il ragionamento kantiano per l'antitesi adattato alla terminologia della meccanica newtoniana (da una quantità nulla di moto non può generarsi una quantità non nulla di moto). Tale ragionamento ci viene in soccorso per provare che lo stato raggiunto è effettivamente uno stato finale.

mento e una graduale diffusione del calore, la fine del moto e l'esaurimento dell'energia potenziale nell'universo materiale. Il risultato di tutto questo sarà inevitabilmente uno stato di riposo universale e di morte, nell'ipotesi che l'universo sia finito e soggetto alle leggi considerate.<sup>16</sup> (Thomson Kelvin 1862, p. 388)

In definitiva, dunque, per Schopenhauer e per una visione filosofica informata da questa interpretazione del secondo principio della termodinamica, la fine dell'universo potrebbe configurarsi come uno stato finale, senza tuttavia costituire necessariamente un nulla dal punto di vista ontologico: una 'fine' che si configurerebbe cioè come l'ingresso del mondo in una fase definitiva di stasi<sup>17</sup>.

---

<sup>16</sup>Come è chiaro alla fine del passaggio citato, la fine dinamica dell'universo dipende da due ipotesi in un certo senso 'metafisiche', ossia al di là delle possibilità di ricerca della scienza fisica: 1) che l'universo sia finito 2) che la seconda legge della termodinamica valga sull'universo tutto e non solo su ciascuna delle sue parti.

<sup>17</sup>A mia conoscenza non discendono invece tradizionalmente considerazioni sul cominciamento dell'universo nel tempo a partire dai risultati della termodinamica classica. Tuttavia ci si potrebbe chiedere: dato che l'energia meccanica tende irrimediabilmente a dissiparsi fino ad uno stato di morte termica, allora perché tale stato non è già stato raggiunto, se l'universo esiste da sempre?

## Riferimenti bibliografici

- Alexandrinus, Philo (2013). *De Opificio Mundi Legum Allego-riarum*. A cura di L. Cohn. Berlin: de Gruyter.
- Augustinus (1997). *De Civitate Dei*. A cura di C. Horn. Berlin: de Gruyter.
- Cicero (2011). *Vom Wesen der Götter/De Natura Deorum*. Trad. da O. Gigon e L. Straume-Zimmerman. Berlin: de Gruyter.
- Diels, H. e W. Kranz (1985). *Die Fragmente der Vorsokratiker*. Zurich: Weidmann.
- D'Iorio, P. (1995). *La linea e il Circolo: Cosmologia e Filosofia dell'Eterno Ritorno in Nietzsche*. A cura di M. Miroslav. Genova: Pantograf.
- Drake, F. R. (1974). *Set Theory: An Introduction to Large Cardinals*. Amsterdam: North Holland.
- Fried, M. (1940). "Kant's First Antinomy: A Logical Analysis". In: *Mind* 49.194, pp. 204–218.
- Gram, M. (1967). "Kant's First Antinomy". In: *The Monist* 51.4, pp. 499–518.
- Guyer, P. (1992). *The Cambridge Companion to Kant*. A cura di P. Guyer. Cambridge: Cambridge University Press.
- Kant, I. (1976). *Critica della Ragion Pura*. Milano: Adelphi.
- Kaye, R. (1991). *Models of Peano Arithmetic*. Oxford: Clarendon Press.
- Laertius, Diogenes (1999). *Vitae Philosophorum*. A cura di M. Miroslav. Berlin: de Gruyter.
- Loparic, Z. (1990). "The Logical Structure of the First Antinomy". In: *Kant-Studien* 81.3, pp. 280–303.
- Lucretius (2013). *De Rerum Natura*. A cura di A. Fellin. Torino: UTET.
- Nietzsche, F. (1988). *Nachgelassene Fragmente: 1887-1889*. A cura di G. Colli e M. Montinari. Berlin: de Gruyter.
- Persius (2007). *Saturarum Liber*. A cura di W. Kissel. Berlin: de Gruyter.
- Russell, B. (1915). *Our Knowledge of the External World*. London: The Open Court Publishing Company.
- Schopenhauer, A. (2006). *Il Mondo come Volontà e Rappresentazione*. A cura di S. Giammetta. Milano: Bompiani.
- Smith, K. (1918). *A Commentary to Kant's Critique of Pure Reason*. Milano.
- Strawson, P.F. (1966). *The Bounds of Sense: An Essay on Kant's Critique of Pure Reason*. London: Methuen.

Thomson Kelvin, W. (1862). "On the Age of the Sun's Heat". In: *Macmillan's Magazine* 5.1, pp. 388–393.

Walsh, W.H. (1975). *Kant's Criticism of Metaphysics*. Edinburgh: Edinburgh University Press.





# Sul rapporto tra il sistema di riferimento cosale e gli altri sistemi di riferimento in "Empiricism, Semantics and Ontology"<sup>1</sup>

*Alberto Bergamini*

**Abstract.** Il veto anti-metafisico posto da Carnap in *Empiricism, Semantics, and Ontology* postula e allo stesso tempo implica che tutti i sistemi di riferimento linguistici siano sullo stesso livello, ossia che tutti siano sistemi composti da tipi di entità e regole di composizione deliberatamente introdotti nel linguaggio per scopi pragmatici e che nessuno possa vantare differenze o particolarità di statuto rispetto agli altri. Nel presente articolo cercherò di dimostrare che ci sono buone ragioni per dubitare che le cose stiano così, in particolare sosterrò che il sistema di riferimento cosale gode di un'eccellenza funzionale rispetto agli altri, che lo mette in una posizione asimmetrica e che rende il suo rifiuto più problematico degli altri.

**Keywords.** Carnap, Meta-ontologia, Sistemi di riferimento, Sistema di riferimento cosale.

<sup>1</sup>Desidero ringraziare Paolo Valore per le preziose osservazioni e suggerimenti.

## 1 Introduzione

Negli ultimi anni si è assistito alla ripresa di un vivace dibattito intorno al valore della critica di Rudolf Carnap alla possibilità e al significato della ricerca metafisica (Blatti e Lapointe 2016), e segnatamente intorno al valore della critica espressa nel celeberrimo *Empiricism, Semantics, and Ontology* (Carnap 1956, pp. 205-221). Da più parti e con argomenti diversi si contesta la legittimità della riabilitazione della metafisica operata, consapevolmente o meno, da W.V.O. Quine (1980a, pp. 1-19), al punto che Huw Price, in conclusione a un suo recente articolo, in cui si sosteneva l'attualità della critica carnapiana e la sostanziale inefficacia della riabilitazione quineiana, dichiarava: «The challenge of “Empiricism, Semantics, and Ontology” remains unanswered» (Price 2009, pp. 320-346). Tuttavia anche nel caso che il giudizio di Price sulla riabilitazione quineiana fosse corretto, c'è forse motivo di dubitare che questo comporti un ritorno alla prospettiva carnapiana: l'articolo di Carnap potrebbe infatti presentare ulteriori debolezze, diverse da quelle messe in luce da Quine.

L'obiettivo del presente articolo è mostrare che i vari sistemi di riferimento, in cui secondo Carnap si articola il linguaggio, non sono tutti sullo stesso piano: il sistema di riferimento cosale gode di un'eccellenza funzionale che lo pone in una posizione anomala e asimmetrica rispetto agli altri. Se ciò fosse vero, potrebbero esserci conseguenze a cascata sulla tenuta dell'intera argomentazione carnapiana, almeno per come è formulata in *Empiricism, Semantics, and Ontology*.

## 2 La tesi di Carnap

L'argomento di Carnap è noto: le proposizioni dell'ontologia sono prive di significato perché la domanda circa l'esistenza di entità o di un tipo di entità ha senso solo all'interno di un dato sistema di riferimento linguistico e dal momento che l'ontologia pretende di porre questioni di esistenza in assoluto, ossia indipendentemente da qualsiasi sistema di riferimento di linguistico, le sue proposizioni sono prive di significato.

I vari sistemi di riferimento linguistici sono costruiti mediante l'introduzione di tipi di entità e regole di composizione, e vengono accettati o rifiutati in base a considerazioni pragmatiche per il conseguimento di certi scopi: se, per esempio, riterremo utile perseguire gli scopi della ricerca matematica, introdurremo tipi di entità come numeri, il termine generale “numero” per le entità che abbiamo introdotto, variabili numeriche, proprietà come pari e dispari, assieme alle regole di composizione per gestire le entità e le proprietà che abbiamo introdotto.

In questa prospettiva le uniche questioni esistenziali che ha senso porsi sono interne ai vari sistemi di riferimento e riguardano l'esistenza di entità particolari del tipo ammesso dal sistema di riferimento, per esempio: esiste un numero primo maggiore di 14597? Qualora si ponga la questione circa l'esistenza stessa del tipo di entità, se la domanda è posta internamente al sistema di riferimento sarà analiticamente vera o falsa: se siamo all'interno del sistema dei numeri è analiticamente vero che esistono i numeri perché proposizioni come "cinque è un numero" sono vere, mentre è analiticamente falso che esistano gli orologi perché gli orologi non fanno parte dei tipi di entità ammessi dal sistema di riferimento dei numeri. Se invece la questione circa l'esistenza dei numeri è posta esternamente al sistema di riferimento dei numeri (e a qualsiasi altro sistema di riferimento) i casi sono due: o la si pone come questione teorica e allora è priva di senso in quanto che, come si è detto, fuori da qualsiasi sistema di riferimento non ha senso interrogarsi circa l'esistenza di qualsivoglia entità o tipo di entità; oppure la si pone come questione pragmatica il cui significato è: ci conviene ammettere l'esistenza dei numeri e cioè adottare il sistema di riferimento dei numeri?

A questo punto è bene notare che, per prima cosa, presupposto e assieme conseguenza dell'argomentazione di Carnap è che tutti i sistemi di riferimento siano sullo stesso piano, nel senso che nessuno di essi è altro da un insieme di tipi di entità e regole di composizione deliberatamente introdotti nel linguaggio per scopi pragmatici; e, secondo, affinché tutti i sistemi di riferimento siano sullo stesso piano nessuno deve mostrare caratteristiche peculiari nel suo statuto: posto che i sistemi sono incommensurabili, sono tutti sistemi che funzionano allo stesso modo, nessuno può fondare gli altri e nessuno può vantare un'eccellenza funzionale (uso qui il termine "eccellenza" nel senso in cui si dice che qualcosa è qualcosa "per eccellenza") o una precedenza logica rispetto agli altri. Chiediamoci ora se questo valga anche per il sistema di riferimento cosale.

### **3 La «semplicità» del sistema di riferimento cosale**

Si può notare, *en passant*, come sia curioso che in un articolo scritto per dimostrare l'indipendenza dell'utilizzo dei termini astratti dall'adozione di un'ontologia di tipo platonico, l'analisi della suddetta indipendenza cominci proprio dai termini concreti, ossia dal sistema di riferimento per le *cose*, dove per "cose" si intendono gli oggetti concreti e gli eventi fisici che li inglobano. Stando a Carnap, le ragioni della scelta di cominciare proprio dal sistema di riferimento cosale sarebbero da rintracciare nella semplicità delle entità concrete:

Let us consider as an example the simplest kind of entities dealt with in the everyday language: the spatio-temporally ordered system of

observable things and events. Once we have accepted the thing language with its framework for things, we can rise and answer internal questions, e.g., “Is there a white piece of paper on my desk?”, “Did King Arthur actually live?”, “Are unicorns and centaurs real or merely imaginary?”, and the like. (Carnap 1956, pp. 206-207)

Dobbiamo concludere che la precedenza accordata al sistema di riferimento cosale è meramente espositiva, nel senso che, vista la sua semplicità, conviene partire da questo per illustrare la natura e il funzionamento dei sistemi di riferimento in generale, ma, tolta questa precedenza espositiva il sistema di riferimento cosale è del tutto simmetrico rispetto agli altri nel senso che il suo statuto e il suo funzionamento sono del tutto identici allo statuto e al funzionamento del sistema, poniamo, dei numeri o della logica, ecc. Conviene dunque partire dal sistema di riferimento cosale per illustrare il funzionamento dei sistemi di riferimento in generale, ma al di fuori di questa convenienza non c'è nessuna ragione di partire da questo perché tutti i sistemi di riferimento sono sullo stesso piano, identici nel loro statuto e indipendenti l'uno dall'altro per quel che riguarda la loro fondazione.

È tuttavia lecito domandarsi da che cosa derivi questa agevolezza espositiva: per quale ragione gli oggetti concreti e gli eventi osservabili sono il tipo più semplice di entità con cui abbiamo a che fare? Si ribatterà che questo è appunto un caso in cui la quantificazione deve essere ristretta al linguaggio considerato: non si tratta del tipo più semplice di entità in assoluto, ma del tipo di entità più semplice all'interno del linguaggio quotidiano. Ma allora quali sarebbero le altre entità disponibili nel linguaggio quotidiano, diverse da oggetti concreti ed eventi? E più in generale: in cosa consisterebbe questo “linguaggio quotidiano”? Carnap non ne fornisce alcuna descrizione o caratterizzazione, eppure la rigida partizione del linguaggio in sistemi di riferimento mutualmente esclusivi non ammette zone franche o aree di intersezione in cui possa non valere la suddetta partizione: il linguaggio quotidiano deve essere un sistema di riferimento come gli altri e dal momento che non vengono fornite ulteriori delucidazioni mi pare plausibile assumere che il linguaggio quotidiano coincida col sistema di riferimento cosale, e siccome gli elementi che lo compongono, cioè oggetti concreti ed eventi, sono «the simplest kind of entities», ne segue che è il sistema di riferimento cosale in toto a essere il sistema di riferimento più semplice; e questa in fondo è precisamente la ragione per cui Carnap sceglie di iniziare l'analisi proprio da questo.

Ora, anche volendo non considerare questa semplicità espositiva come un primo indizio di asimmetria del sistema di riferimento cosale rispetto agli altri sistemi, è però lecito domandare in cosa consista tale semplicità. Preso in se stesso, cioè al livello del suo statuto, ossia indipendentemente da qualsiasi considerazione a posteriori, non c'è nessuna ragione per cui il sistema di riferi-

mento cosale dovrebbe essere più semplice di quello delle entità matematiche o di quello della fisica: stando alla caratterizzazione che ne dà Carnap, in tutti i casi si tratta di complessi di termini e regole di utilizzo introdotti nel linguaggio mediante deliberazione. Per quale ragione dunque entità come i tavoli o le sedie dovrebbero essere più “semplici” di entità come la radice quadrata di due? A chi ribattesse che si tratta di una semplicità *intuitivamente* molto chiara ma difficilmente formulabile, risponderi che è lo stesso tipo di chiarezza intuitiva ma difficilmente formulabile che ci rende intuitivamente chiaro ma difficilmente formulabile che le sedie esistono mentre i numeri no. Tuttavia tale chiarezza intuitiva del concetto di esistenza è precisamente quello che Carnap si propone di combattere, dunque la domanda resta.

Un modo di uscire dall'*impasse* è rispondere che il sistema di riferimento cosale è il più semplice dal punto di vista espositivo perché è il più *familiare*, ovvero quello con cui abbiamo più spesso a che fare e con cui siamo di conseguenza più a nostro agio. D'altra parte è proprio a questo che sembra alludere Carnap quando parla di «everyday language». Ma a questo punto si potrebbero domandare le ragioni di questa familiarità: se, di nuovo, i sistemi di riferimento sono identici nel loro statuto, per quale ragione è con quello relativo agli oggetti concreti che abbiamo a che fare più spesso?

La risposta più immediata che può venire in mente, e cioè che il sistema di riferimento cosale è quello con cui abbiamo a che fare più spesso perché è quello della vita di tutti i giorni, si risolve ovviamente in un argomento circolare. Un'altra soluzione, più raffinata, e più nello spirito di Carnap, è sostenere che il sistema di riferimento cosale è quello con cui abbiamo più spesso a che fare perché è il più utile per perseguire gli scopi pratici della vita di tutti i giorni (e così si spiegherebbe l'ambigua identificazione tra sistema di riferimento cosale e «everyday language», implicitamente asserita da Carnap): se infatti è solo occasionalmente e in certe situazioni che a qualcuno capita di perseguire gli scopi per cui è necessario ricorrere al sistema di riferimento della fisica, è invece quotidianamente che a tutti capita di perseguire gli scopi in rapporto ai quali il sistema di riferimento cosale è il più utile. Ecco dunque spiegato l'apparente mistero: la semplicità meramente espositiva del sistema di riferimento cosale si basa sulla sua maggiore familiarità rispetto agli altri sistemi di riferimento, e questa a sua volta sul suo impiego vantaggioso nel perseguire gli scopi della vita quotidiana, che sono perseguiti più spesso di quelli relativi agli altri sistemi di riferimento, e da tutti.

Tuttavia, siccome ora l'accento è posto sugli scopi per i quali il sistema di riferimento cosale è adottato, per meglio metterli in luce confrontiamoli con quelli perseguiti attraverso altri sistemi di riferimento, per esempio quello della fisica. È chiaro che nel caso di quest'ultimo gli scopi saranno gli ordinari scopi scientifici comuni a tutte le scienze e quindi, sotto questo punto di vista, a tut-

ti i sistemi di riferimento relativi a discipline scientifiche: prevedere e spiegare fenomeni, e accertare l'esistenza di entità particolari del tipo ammesso dal sistema di riferimento in questione. Ora, fino a un certo punto questo è anche il caso del sistema di riferimento cosale. Tuttavia, di fianco a tali scopi *lato sensu* conoscitivi, mi sembra vadano contati anche scopi pratici e interattivi che negli altri sistemi di riferimento sono assenti. Per esempio, dal momento che nel sistema di riferimento cosale sono giustamente presenti anche gli individui umani, in quanto oggetti concreti occupanti una posizione spazio-temporale, come dimostra uno dei quesiti che secondo Carnap è possibile porsi all'interno del sistema di riferimento cosale: "Did King Arthur actually live?", allora è chiaro che il sistema di riferimento cosale è largamente impiegato per la funzione di avere a che fare con quegli oggetti concreti che sono gli individui, interazione che il più delle volte coinvolge gli altri oggetti concreti che ci circondano e con cui abbiamo a che fare tutti i giorni. Di conseguenza si potrebbe concludere che gli scopi perseguiti attraverso il sistema di riferimento cosale sono rubricabili sotto l'etichetta "avere a che fare con gli oggetti concreti". È possibile immaginare una replica<sup>2</sup>: anche le conoscenze fisiche, matematiche e scientifiche in generale hanno applicazioni pratiche, lo sviluppo dei nostri dispositivi meccanici e tecnologici, dagli aerei agli smartphone, dipende dall'incremento di conoscenze scientifiche che ne rendono possibile la costruzione, e infatti la ricerca scientifica è finanziata in grossa parte per le sue ricadute pratiche. Di conseguenza il sistema di riferimento cosale non sarebbe l'unico a permetterci di "avere a che fare con le cose" dal momento che anche gli altri sistemi di riferimento si prestano a questi scopi pratici. Non c'è dubbio che le cose stiano così, tuttavia penso che questa obiezione non colga nel segno. È senz'altro vero che anche gli altri sistemi di riferimento hanno ricadute pratiche e per nulla marginali: le conoscenze acquisite nell'ambito di sistemi di riferimento come fisica, chimica e matematica sono quotidianamente sfruttate per scopi pratici e modificano la nostra interazione con *gli oggetti concreti*. Ma proprio qui sta il punto: i risvolti e le conseguenze pratiche di sistemi di riferimento come quello della fisica o della chimica si misurano all'interno del sistema di riferimento cosale, ossia quando, attraverso la formulazione di leggi, i risultati conoscitivi ottenuti nel contesto di sistemi di riferimento come quello della fisica o della chimica o della matematica, sono trasferiti all'interno del sistema di riferimento cosale e impiegati per avere a che fare con cose come i computer o gli aerei o i prodotti dolciari industriali, che sono tutti oggetti concreti. In altre parole è solo una volta che si trova modo di applicare le conoscenze acquisite in altri sistemi di riferimento alle entità del sistema di riferimento cosale, ossia agli oggetti concreti, che quei sistemi di riferimento si prestano a scopi pratici. Non a caso, come si

<sup>2</sup>Ringrazio l'anonimo revisore del presente articolo per aver sollevato questa obiezione, alla quale non avevo pensato.

diceva, spesso e volentieri ciò che fa la differenza nell'assegnazione di fondi per la ricerca sono i possibili risvolti pratici delle eventuali scoperte, i quali si giocano tutti sul terreno del sistema di riferimento cosale: è comunque quest'ultimo a stabilire l'utilità pratica degli altri sistemi di riferimento.

Se dunque mi si concede la caratterizzazione dello scopo perseguito attraverso il sistema di riferimento cosale nei termini di "interazione pragmatica con gli oggetti concreti" in opposizione agli altri sistemi di riferimento, dove non avviene un'analoga interazione con le relative entità, possiamo riprendere il filo della nostra indagine.

Abbiamo detto che la maggiore familiarità del sistema di riferimento cosale rispetto agli altri sistemi di riferimento si basa sul suo impiego vantaggioso nel perseguire gli scopi della vita quotidiana, ossia sul suo impiego vantaggioso nell'avere a che fare con gli oggetti concreti. Ma basta poco per rendersi conto che, dal punto di vista di Carnap, questo è un paradosso, perché è precisamente il sistema di riferimento cosale a introdurre gli oggetti concreti. Come è dunque possibile che il sistema di riferimento cosale risulti vantaggioso nell'interazione con gli oggetti concreti, se prima dell'adozione del sistema di riferimento cosale non c'è nessun oggetto concreto con cui interagire? Ribattere che gli oggetti concreti erano qui da molto prima della comparsa dell'uomo, e dunque da molto prima della comparsa di qualunque sistema di riferimento, è una posizione di realismo metafisico che Carnap rigetterebbe al pari di qualsiasi altra posizione metafisica: di nuovo, parlare dell'esistenza di un tipo di entità prima dell'introduzione del relativo sistema di riferimento è, nella prospettiva di Carnap, privo di senso.

Abbiamo corso troppo, abbiamo posto l'accento sul complemento piuttosto che sul verbo: abbiamo detto che gli scopi della vita quotidiana sono rubricabili sotto l'etichetta "avere a che fare con gli oggetti concreti" oppure "interagire con gli oggetti concreti", ma questo non può essere perché prima dell'adozione del sistema di riferimento cosale non c'è nessun oggetto concreto con cui avere a che fare o con cui interagire. Ne segue che gli scopi della vita quotidiana sono rubricabili sotto l'etichetta "avere a che fare con" oppure "interagire con", ed è proprio il sistema di riferimento cosale che ci fornisce la sponda di questa interazione, rendendocela più semplice e fruttuosa.

Di nuovo sembreremmo essere al capolinea: al pari degli altri sistemi di riferimento, il sistema di riferimento cosale risulta vantaggioso per gli scopi per cui è adottato, nella fattispecie l'interazione o "l'avere a che fare", e la familiarità che ce lo rende tanto semplice è dovuta al fatto che, a differenza degli scopi puramente conoscitivi che ci si propone di perseguire attraverso il sistema di riferimento della fisica o della matematica o della logica, lo scopo dell'interazione o dell' "avere a che fare" è più frequente perché è lo scopo della vita di tutti i giorni.

## 4 Il rifiuto del sistema di riferimento cosale

Eppure che ne è di tutta questa faccenda se la si capovolge e la si guarda dalla prospettiva del rifiuto dei sistemi di riferimento piuttosto che da quello della loro adozione?

A ben guardare ci possono essere due ordini di ragioni per cui i sistemi di riferimento possono essere rifiutati. Il primo è che un dato sistema di riferimento, poniamo quello della fisica aristotelica, non è utile per soddisfare gli scopi che ci si è prefissi. In tal caso il sistema in questione sarà rifiutato in favore di un sistema di riferimento più adatto al perseguimento dei medesimi scopi, per esempio quello della fisica newtoniana. Il secondo livello concerne invece un più radicale rifiuto dello scopo per cui adottare un determinato sistema di riferimento: qualora qualcuno decidesse che la conoscenza della fisica non rientra tra gli scopi che desidera perseguire può tranquillamente rifiutare il linguaggio della fisica senza sostituirlo con un altro, ed è anzi normalmente quello che accade, dato che solo una piccolissima porzione dell'umanità si dedica allo studio della fisica.

Ora, la frequenza del rifiuto di gran parte dei sistemi di riferimento elencati da Carnap (quello della fisica, quello della matematica, quello della logica, ecc.) trova la sua ragione nel fatto che rifiutarli non comporta nessun particolare problema, dato che il loro impiego risulta vantaggioso solo in rapporto a scopi conoscitivi il cui perseguimento non pare essere una necessità inaggirabile. Ma in questa prospettiva che dire del sistema di riferimento cosale? Se il sistema di riferimento cosale fosse, dal punto di vista del suo statuto, perfettamente identico agli altri, la sua adozione o il suo rifiuto dovrebbero seguire le medesime modalità e comportare (o non comportare) i medesimi problemi degli altri. Ma è davvero così?

Un primo indizio ce lo fornisce lo stesso Carnap quando, dopo aver spiegato che le questioni esterne non sono di natura conoscitiva bensì pragmatica e vertono sull'opportunità o meno di adottare il sistema di riferimento in questione, ammette che:

In the case of this particular example [quello del sistema di riferimento cosale], there is usually no deliberate choice because we all have accepted the thing language early in our lives as a matter of course. (Carnap 1956, p. 207)

Innanzitutto si potrebbe notare, con un po' di malizia, come l'espressione «In the case of this particular example» tenti forse di ribadire in maniera, diciamo così, subliminale che il sistema di riferimento cosale è solo un esempio scelto in virtù della sua comodità, senza alcuna reale peculiarità rispetto agli altri; e tenti quindi, sempre subliminalmente, di minimizzare la differenza, per nulla

minima, che si sta per enunciare tra il sistema di riferimento cosale e gli altri sistemi di riferimento.

Ciononostante a me il punto sembra tutt'altro che pacifico. Carnap ha perfettamente ragione a dire che nel caso dell'accettazione del sistema cosale di solito non si dà una scelta «deliberata» (e quindi non si dà una scelta punto). Il problema è che manca di sottolineare come il sistema di riferimento cosale sia l'unico per cui non si dia tale scelta. Per quale altro sistema di riferimento è infatti immaginabile un'adozione talmente automatica da non lasciare spazio a una scelta? C'è di solito ampio margine di scelta circa l'opportunità di adottare il sistema di riferimento della fisica o quello della matematica o quello della logica o quello della semantica, per lo meno al secondo livello, ossia al livello di adozione o rifiuto degli scopi per cui adottare il sistema di riferimento della fisica o della matematica, ecc. Solo in quello cosale invece l'adozione è a tal punto considerata «a matter of course» da non essere neppure, di fatto, una scelta. E questa è davvero una grande asimmetria del sistema di riferimento cosale rispetto agli altri, i quali in effetti si comportano invece tutti nella stessa maniera.

A questo punto una mossa potrebbe consistere nel sostenere che, se l'accettazione del sistema di riferimento cosale avviene in maniera tanto ovvia da sottrarsi di fatto al criterio della scelta deliberata, è già dimostrato che è il sistema di riferimento cosale a sottrarsi al funzionamento dei sistemi di riferimento: se una caratteristica di questo funzionamento è la possibilità di adozione e rifiuto, qualora un sistema di riferimento si sottragga a questa possibilità è chiaro che non può più essere considerato un sistema di riferimento, per lo meno nello stesso senso in cui gli altri sistemi di riferimento sono sistemi di riferimento. Può essere che il sistema di riferimento cosale sia un primo livello di linguaggio che precede logicamente i vari sistemi di riferimento, in quanto piattaforma della loro costruzione o per la loro introduzione?

In ogni caso Carnap si rende conto che se la possibilità di adozione/rifiuto non fosse garantita anche nel caso del sistema di riferimento cosale, quest'ultimo non sarebbe più assimilabile a tutti gli altri sistemi di riferimento, per i quali tale possibilità è effettivamente garantita, e infatti subito dopo tenta di emendare l'asimmetria:

Nevertheless, we may regard it as a matter of decision in this sense: we are free to choose to continue using the thing language or not; in the latter case we could restrict ourselves to a language of sense data and other "phenomenal" entities, or construct an alternative to the customary thing language with another structure, or, finally, we could refrain from speaking. (Carnap 1956, p. 207)

Si potrebbe innanzitutto sostenere che c'è una bella differenza tra rifiutare nel senso di non adottare qualcosa e rifiutare nel senso di abbandonare qual-

cosa che era stato precedentemente (e automaticamente) adottato. Ribattere che anche nel caso degli altri sistemi di riferimento il rifiuto può avvenire nella forma dell'abbandono, per esempio nel passaggio dal sistema di riferimento dell'alchimia a quello della chimica, serve a poco dal momento che anche in questo caso si è comunque data in precedenza quella libera adozione che nel caso del sistema di riferimento cosale non si dà. Questo inoltre porta con sé una constatazione non meno interessante né, a mio giudizio, meno ricca di significato: il sistema di riferimento cosale non conosce mutamenti. Mentre è un fatto che l'attuale sistema di riferimento della fisica sia stato adottato dopo l'abbandono di sistemi di riferimento precedenti per il perseguimento dei medesimi scopi, e così quello della logica o della matematica, tant'è vero che negli attuali sistemi sono disponibili tipi di entità che nei precedenti sistemi di riferimento adottati per il perseguimento dei medesimi scopi non erano disponibili, il sistema di riferimento cosale è immutato dalla notte dei tempi. Ci sono oggetti concreti nuovi, vero: ma sempre oggetti concreti sono, cioè entità che occupano una posizione spazio-temporale. I tipi di entità disponibili non sono mutati: clava e shuttle appartengono entrambi al sistema di riferimento cosale; spiriti vitali e molecole non appartengono entrambi allo stesso sistema di riferimento. Questo, sia detto per inciso, sembra valere in senso sia storico che geografico: il mondo e la storia sono pieni di sistemi di riferimento alternativi a quelli scientifici per il perseguimento dei medesimi scopi, sebbene siano sistemi di minor successo (mito, alchimia, divinazione sono solo alcuni esempi), ma sarei pronto a scommettere che da nessuna parte si è mai data una popolazione che non disponesse, per gli scopi della vita quotidiana, di un sistema di riferimento di entità in relazione spazio-temporale le une rispetto alle altre, a prescindere da come questo sistema fosse concretamente organizzato (ribattere che non tutte le popolazioni operano tra gli oggetti concreti le stesse distinzioni concettuali che operiamo noi non toglie forza alla considerazione: si tratta comunque di entità spazio-temporalmente localizzate, cioè di oggetti concreti secondo la definizione di Carnap).

In ogni caso anche tralasciando non solo il fatto, comunque significativo, che c'è una grande differenza tra il rifiuto nella forma della non-adozione e il rifiuto nella forma dell'abbandono, ma anche il fatto che, di nuovo, solo nel caso del sistema di riferimento cosale il rifiuto nella forma dell'abbandono precede necessariamente quello nella forma della non-adozione, resta comunque dubbio che le cose stiano come dice Carnap. Nella fattispecie resta dubbio che la sostituzione del sistema di riferimento cosale con un sistema di riferimento di dati di senso o con un'alternativa all'attuale linguaggio cosale o infine con il silenzio (sono le tre alternative elencate da Carnap, e a me non ne vengono in mente altre) equivalga ad un rifiuto del sistema di riferimento cosale nello stesso senso in cui si rifiuta il sistema di riferimento della fisica o della matematica.

Per quanto riguarda il rifiuto del sistema di riferimento cosale attraverso l'adozione di un sistema di riferimento di dati di senso, mi pare molto dubbio che si tratti di una possibilità reale. Innanzitutto si potrebbe notare che, qualora si riconosca che lo scopo perseguito attraverso l'adozione del sistema di riferimento cosale è "l'averne a che fare con", un eventuale sistema di riferimento di dati di senso mal si presta a questo scopo, dal momento che un sistema di riferimento di entità fenomeniche conduce facilmente a forme più o meno esplicite di solipsismo, nel contesto del quale non c'è nulla con cui avere a che fare.

Ci sono poi altri problemi. Uno dei punti di forza meno evidenti del sistema di riferimento cosale è che noi, in quanto oggetti concreti, ne facciamo parte; di conseguenza, come agente del linguaggio cosale, non ho difficoltà a riconoscere me e i miei interlocutori come entità dello stesso tipo delle entità di cui parliamo: quando mia madre mi chiede se so dov'è il suo cellulare e io le dico dove l'ho visto l'ultima volta, attraverso il sistema di riferimento cosale posso rappresentarmi la situazione come quella in cui l'oggetto concreto che sono interagisce con l'oggetto concreto che è mia madre a proposito di un terzo oggetto concreto, cioè il cellulare, il tutto in conformità alle regole e alle possibilità del sistema di riferimento cosale. Ora, immaginando di dover far fronte a una circostanza analoga attraverso un ipotetico sistema di riferimento di dati di senso, la cosa si fa problematica già a partire dalla nozione che il soggetto del dato di senso deve avere di se stesso: in quanto agente del linguaggio deve sapere se stesso come agente del linguaggio e il suo interlocutore come interlocutore, e ciò implica che difficilmente potrà considerare entrambi dati di senso perché manifestamente i dati di senso non parlano né hanno coscienza di se stessi. Questo in fondo è solo una conseguenza del fatto che la grammatica stessa dei dati di senso esige un soggetto di quei dati di senso che non sia a sua volta un dato di senso. Un oggetto concreto come un individuo umano può parlare un linguaggio di dati di senso ma solo a partire dal presupposto che lui non è un dato di senso bensì un oggetto concreto<sup>3</sup>, e se si considera a fondo la possibilità proposta da Carnap, ossia la sostituzione del sistema di riferimento cosale con un linguaggio di dati di senso, l'impressione è che ci si troverebbe di fronte a oggetti concreti, che sanno benissimo di essere oggetti concreti (e che dunque pensano se stessi nei termini del sistema di riferimento cosale)<sup>4</sup> che si sforzano di tradurre ciò che per loro sono e saranno sempre oggetti concreti in un linguaggio di dati di senso (mentre una cosa analoga non accade nel caso dell'abbandono del sistema di

<sup>3</sup>Si potrebbe invocare qui l'idealismo berkeleyano e sostenere che il soggetto potrebbe concepirsi come una sostanza mentale, tuttavia non mi pare che questa replica eminentemente metafisica sia nelle possibilità di Carnap.

<sup>4</sup>Anche perché è dubbio che la semplice sostituzione del sistema di riferimento cosale basterebbe a modificare la nostra percezione dello spazio e del tempo, che è ciò che in buona sostanza determina la nostra percezione degli oggetti concreti come oggetti concreti spazio-temporalmente localizzati.

riferimento dell'alchimia per quello della chimica). Non sono sicuro che il mio argomento scardini la proposta di Carnap, tuttavia è ragionevole domandarsi se la possibilità di sostituire il sistema di riferimento cosale con un sistema di riferimento di dati di senso per il perseguimento dei medesimi scopi pratici sia una possibilità reale o non piuttosto apparente.

Nemmeno la possibilità che il rifiuto del sistema di riferimento cosale avvenga nella forma di una sostituzione con «an alternative to the customary thing language with another structure» mi pare fondata. Ciò che infatti distingue i vari sistemi di riferimento sono i tipi di entità disponibili in ognuno di essi. Ora, nel caso di una sostituzione del sistema di riferimento cosale con «an alternative to the customary thing language with another structure», i casi sono due: o i tipi di entità disponibili sono gli stessi, e cioè entità in una rete di reciproche relazioni spazio-temporali, e allora si tratterà comunque di un sistema di riferimento di oggetti concreti ed eventi, solo organizzato diversamente dal punto di vista logico-sintattico, mentre il sistema di riferimento della chimica non è il sistema di riferimento dell'alchimia riformulato dal punto di vista logico-sintattico (e così quello della fisica newtoniana rispetto alla fisica aristotelica); oppure il nuovo sistema di riferimento conterrà tipi di entità del tutto diversi dagli oggetti concreti ed eventi, ma in questo caso mi pare che Carnap sia in debito di un esempio dal momento che, come nel caso del sistema dei dati di senso, sembra molto difficile immaginare un sistema di riferimento adatto ai medesimi scopi del sistema di riferimento cosale diverso dal sistema di riferimento cosale. Certo, difficile non significa impossibile, e senz'altro prima dell'invenzione del sistema di riferimento della fisica newtoniana nessuno avrebbe probabilmente immaginato di poter perseguire la conoscenza fisica del mondo con strumenti diversi da quelli messi a disposizione dalla fisica aristotelica; tuttavia mi pare che un'ipotesi così estrema necessiterebbe di alcune delucidazione e precisazioni ulteriori per essere presa seriamente e non essere considerata, di nuovo, una possibilità apparente. L'impressione è che Carnap presupponga quello che deve dimostrare: dal momento che è convinto che il sistema di riferimento cosale non sia altro che un complesso di entità e regole di composizione deliberatamente introdotto nel linguaggio al pari del sistema di riferimento della fisica, allora come nel caso della fisica è senz'altro possibile immaginare un'alternativa. Tuttavia mi pare che questo sia appunto rovesciare i termini del discorso: bisognerebbe invece dire che solo a patto che sia concretamente possibile mettere in campo un'alternativa al sistema di riferimento cosale allora si potrà considerare quest'ultimo un sistema di riferimento come tutti gli altri. A questo proposito è importante notare che quando parlo di un'alternativa concreta al sistema di riferimento cosale, non sto dicendo che questa alternativa dovrebbe funzionare meglio o anche solo tanto bene quanto il sistema di riferimento cosale per essere considerata una valida alternativa: se anche infatti funzionasse peggio sarebbe

comunque un'alternativa e a quel punto la preferenza accordata al sistema di riferimento cosale sarebbe meramente pragmatica, e questo è precisamente ciò che sostiene Carnap. Quello che sto sostenendo è che non sia concretamente possibile un'alternativa che funzioni punto in relazione ai medesimi scopi per i quali il sistema di riferimento cosale è adottato. Ho esposto le ragioni per cui le due possibilità elencate da Carnap mi paiono apparenti. In ogni caso penso che la possibilità di tale sostituzione sia tutt'altro che ovvia e che l'onere della prova spetti a Carnap.

Diversa, ma ancora più ricca di conseguenze per il nostro discorso, è la terza eventualità, ossia l'astensione dal parlare. La possibilità della scelta deliberata di accettare il sistema di riferimento cosale (possibilità che lo renderebbe omogeneo agli altri sistemi di riferimento) starebbe nella sua possibilità di rifiutarlo astenendosi dal parlare. Ma questo, lungi dal rendere il sistema di riferimento cosale omogeneo agli altri, ne decreta la definitiva peculiarità. Infatti nel caso di quale altro sistema di riferimento, il rifiuto comporta l'astensione dal parlare? È chiaro che né il rifiuto del sistema di riferimento della fisica né di quello della matematica né di quello della logica comporta tale astensione, dal momento che il rifiuto di tutti questi, nel senso del rifiuto degli scopi perseguibili attraverso questi, rende comunque possibile qualcosa che assomiglia molto a un "passo indietro" in quell'«everyday language» che Carnap stesso ci ha suggerito coincidere con il sistema di riferimento cosale. Ma una volta rifiutato questo, pare necessario astenersi dal parlare. Qualora si obiettasse che questo non è vero dal momento che è sempre possibile adottare un altro sistema di riferimento, come quello della fisica o della matematica, risponderei che ciò è vero ma, primo, tale adozione di un sistema come quello della fisica o della matematica a seguito del rifiuto del sistema di riferimento cosale non avviene in maniera tanto automatica e disinvolta come avviene nel caso contrario: si può smettere di fare fisica quando si vuole e continuare comunque a parlare senza che ciò comporti alcuna reale deliberazione, a parte quella di smettere di fare fisica, mentre nel caso del rifiuto del sistema di riferimento cosale, se non si vuole smettere di parlare, pare necessario decidersi a impegnarsi in un'attività come "fare fisica". Secondo, l'adozione di un sistema di riferimento come quello della fisica o della matematica per scongiurare l'afasia derivante dal rifiuto del sistema di riferimento cosale comporta una decisa uscita dallo spettro degli scopi perseguibili attraverso il sistema di riferimento cosale per limitarsi agli scopi perseguibili attraverso il sistema di riferimento della fisica o della matematica; il che significa che il rifiuto del sistema di riferimento cosale comporta la sostituzione della vita con la ricerca scientifica, e ciò non pare essere una possibilità nemmeno ideale.

A dispetto degli sforzi di Carnap per dimostrare il contrario, pare proprio che lo statuto del sistema di riferimento cosale non sia equiparabile allo statuto di tutti gli altri. L'impressione è che ci sia qualcosa che ci obbliga insistentemente

alla sua adozione in modo da renderla materia di fatto più che di scelta.

In questa prospettiva la caratterizzazione di tutti i sistemi di riferimento come funzionalmente omogenei fallisce: il sistema di riferimento cosale è quantomeno asimmetrico rispetto a tutti gli altri e in questa asimmetria trova la sua reale giustificazione la precedenza che Carnap gli accorda al momento dell'esposizione.

Questa precedenza, che Carnap non può non accordare al sistema di riferimento cosale, ma che tenta di deflazionare presentandola come espositiva, è in realtà un'eccellenza funzionale a tutti gli effetti: il sistema di riferimento cosale possiede un grado di necessità maggiore rispetto agli altri, dal momento che potremmo benissimo fare a meno del sistema di riferimento della fisica, qualora la conoscenza delle entità della fisica non rientrassi negli scopi che liberamente decidiamo di perseguire, ma non potremmo mai fare a meno di un sistema di riferimento cosale perché l'interazione con gli oggetti concreti pare non essere uno scopo che possiamo liberamente scegliere se accettare o meno, e qualora il sistema di riferimento cosale non fosse mai stato disponibile nella forma in cui lo è, il sistema di riferimento che avrebbe tenuto il suo posto sarebbe stato per noi un sistema di riferimento cosale a tutti gli effetti.

## Riferimenti bibliografici

- Bird, Graham (2003). "Carnap's Internal and External Questions". In: *Language, Truth and Knowledge: Contributions to the Philosophy of Rudolf Carnap*. A cura di Thomas Bonk. Dordrecht: Kluwer Academic Publishers, pp. 97–131.
- Blatti, Stephan e Sandra Lapointe, cur. (2016). *Ontology after Carnap*. Oxford: Oxford University Press.
- Carnap, Rudolf (1956). "Empiricism, Semantics and Ontology". In: *Meaning and Necessity: A study in Semantics and Modal Logic*. A cura di Rudolf Carnap. Chicago: University of Chicago Press, pp. 205–221.
- Creath, Richard (2005). "Carnap's Program and Quine's Question". In: *Carnap Brought Home: The View from Jena*. A cura di S. Awodey e K. Carsten. Chicago: Open Court, pp. 279–293.
- Dorr, Cian (2005). "What We Disagree About When We Disagree About Ontology". In: *Fictionalism in Metaphysics*. A cura di M.E. Kalderon. Oxford: Clarendon Press, pp. 234–286.
- Eklund, Matti (2009). "Carnap and Ontological Pluralism". In: *Metametaphysics*. A cura di D. Chalmers, D. Manley e R. Wasserman. Oxford: Oxford University Press, pp. 130–156.
- Price, Huw (1992). "Metaphysical Pluralism". In: *Journal of Philosophy* 89.8, pp. 387–409.
- (1997). *Carnap, Quine and the Fate of Metaphysics*. Electronic Journal of Analytic Philosophy. URL: <http://ejap.louisiana.edu/EJAP/1997.spring/price976.html>.
- (2009). "Metaphysics after Carnap: the Ghost who Walks?" In: *Metametaphysics*. A cura di D. Chalmers, D. Manley e R. Wasserman. Oxford: Oxford University Press, pp. 320–346.
- Quine, Willard Van Orman (1966). "On Carnap's Views on Ontology". In: *The ways of paradox and other essays*. A cura di W.V.O. Quine. New York: Random House, pp. 126–134.
- (1980a). "On What There Is". In: *From a logical point of view*. A cura di W.V.O. Quine. Harvard: Harvard University Press, pp. 1–19.
- (1980b). "Two Dogmas of Empiricism". In: *From a logical point of view*. A cura di W.V.O. Quine. Harvard: Harvard University Press, pp. 20–46.
- Soames, Scott (2009). "Ontology, Analyticity and Meaning: the Quine-Carnap Dispute". In: *Metametaphysics*. A cura di D. Chalmers, D. Manley e R. Wasserman. Oxford: Oxford University Press, pp. 424–443.



## Alcune considerazioni sulla possibilità metafisica dei supercompiti

*Matteo Casarosa*

**Abstract.** In questo articolo viene descritto un nuovo supercompito e viene difesa un'argomentazione atta a mostrare che esso è metafisicamente impossibile. Vengono considerate ed oppuguate alcune obiezioni alla suddetta argomentazione. Nel rispondere alle obiezioni, viene sviluppato un confronto tra questo supercompito ed altri già presenti nella letteratura, e vengono sottolineate alcune analogie. La conclusione è che anche tali supercompiti sono metafisicamente impossibili. In accordo con questa linea argomentativa, vengono sviluppate alcune considerazioni a favore della conclusione che il supercompito descritto in questo articolo pone delle difficoltà per la possibilità metafisica dei supercompiti in generale.

**Keywords.** Supercompiti, Paradosso di Littlewood-Ross, Topologia, Ricettacoli, Filosofia del tempo.

## 1 Introduzione

I supercompiti sono processi fisici aventi un numero infinito di passaggi successivi e compientisi in un tempo finito<sup>1</sup>. Il tema dei supercompiti è profondamente legato ad altre questioni come quella della discretezza o continuità del tempo<sup>2</sup>: ad esempio, la possibilità di un numero infinito di stati successivi in un tempo finito richiede un numero infinito di istanti in un tempo finito. Il tema dei supercompiti ha quindi grande rilevanza sia in filosofia della fisica che in metafisica.

I primi esempi di supercompiti descritti nella filosofia occidentale paiono essere alcuni scenari appartenenti alla collezione dei paradossi di Zenone, come quello di Achille e della tartaruga. Per una discussione più approfondita di simili scenari bisognerà aspettare la filosofia del novecento.

Nel 1954, il filosofo inglese J. F. Thomson propose un esperimento mentale, detto della Lampada di Thomson, volto a negare la possibilità metafisica dei supercompiti. Thomson chiedeva di immaginare una lampada che possa essere accesa o spenta, ed il cui stato venga cambiato dopo un certo lasso di tempo, poi nuovamente dopo metà del tempo precedente, poi dopo un quarto, e così via *ad infinitum*<sup>3</sup>. L'intero processo terminerebbe dopo un tempo doppio rispetto a quello del primo passaggio, e vi sarebbero infiniti cambi di stato. L'idea di Thomson è che in una situazione del genere lo stato finale della lampada sembra essere indeterminato, situazione che il filosofo in questione rigetta come assurda.

La letteratura successiva ha trovato inefficaci simili argomenti, principalmente perché lo scenario descritto da Thomson è parso a molti come vago ed indeterminato nella sua stessa formulazione, e si è generalmente creduto che l'indeterminatezza della conclusione sia soltanto un risultato di quella della formulazione<sup>4</sup>. Per cercare di superare tale indeterminatezza, Earman (1996), ad esempio, hanno proposto uno scenario in cui l'accensione della lampada dipende dal contatto con un pavimento di una palla rimbalzante. Dispersa completamente in calore la propria energia dopo infiniti rimbalzi progressivamente più brevi, la palla si arresterà sul pavimento, lasciando la lampada accesa. L'indeterminazione dello scenario iniziale scompare.

In tempi recenti, la discussione dei supercompiti si è allontanata dal terreno prettamente metafisico scivolando nel campo della filosofia della fisica, all'interno del quale è stata principalmente la loro possibilità fisica<sup>5</sup>, e non metafisi-

<sup>1</sup>Manchak e Roberts (2016).

<sup>2</sup>Per una breve discussione generale della topologia del tempo si veda ad esempio Markosian (2016).

<sup>3</sup>Thomson (1954).

<sup>4</sup>Benacerraf (1962).

<sup>5</sup>Generalmente accettata: si veda Romero (2014); per un'eccezione si veda Burke (2000).

ca<sup>6</sup>, ad essere oggetto di dibattito.

In questo articolo cerco di evidenziare una difficoltà per la possibilità metafisica dei supercompiti dando un esempio di supercompito metafisicamente impossibile. L'articolo continua con una lista di possibili obiezioni che un oppositore potrebbe porre ed una risposta ad ognuna di esse. Nel rispondere ad alcune delle possibili obiezioni, colgo l'occasione di sottolineare alcune analogie tra questo nuovo supercompito ed altri già presenti nella letteratura, come quello di Littlewood-Ross. Vi è infine una sezione nella quale discuto ulteriormente i legami tra il supercompito descritto in questo articolo ed altri supercompiti, oltre ad evidenziare dei legami tra alcune caratteristiche di questo particolare supercompito ed alcuni elementi costitutivi della definizione stessa di supercompito, ponendo così una difficoltà per la possibilità metafisica dei supercompiti in generale.

## 2 Il supercompito

Si consideri un macchinario contenente una quantità  $\aleph_0$  di mattoni numerati. Il processo eseguito dal macchinario è il seguente: esso seleziona tra i mattoni quello numerato con il numero 0, eventualmente lo deforma o raschia per dargli un'altezza di mezzo metro e lo pone a terra all'istante  $t = 0$ . Successivamente fa la stessa cosa con il mattone numerato con il numero 1 rendendolo alto un quarto di metro, e lo impila a parallelepipedo sopra il primo all'istante  $t = 0.5$  secondi; prende il mattone 2, lo rende alto un ottavo di metro e lo impila all'istante  $t = 0.75$ , e così via *ad infinitum*.

A questo punto ci si chieda quale sarà la situazione al termine del supercompito, ovvero a  $t = 1$  secondi. La mia risposta è che si avrebbe una pila di mattoni alta un metro, che esporrebbe come sua faccia superiore, una delle facce di uno degli infiniti mattoni che lo compongono. Questa affermazione può essere motivata come segue.

Scelgo di suddividere la mia linea argomentativa in due parti, inframezzate da un commento. In questo come in altri argomenti che compariranno in seguito nell'articolo le premesse sono indicate con una "P" e le deduzioni da una "D".

Chiamando "S" il supercompito in questione si ha:

- 1P) Al termine di S viene a prodursi una pila di altezza finita.
- 2P) Una pila di altezza finita costruita impilando mattoni come in S ha una faccia superiore.
- 3P) La faccia superiore di una pila costruita impilando mattoni come in S è faccia superiore di uno dei mattoni che compongono la pila.

<sup>6</sup>Generalmente accettata; per un'eccezione si vedano Gwiazda (2012) e Groarke (1982).

1D) (Da 1P & 2P). La pila risultante alla fine di  $S$  ha una faccia superiore.

2D) (Da 1D & 3P). La pila risultante alla fine di  $S$  espone come faccia superiore una faccia di uno dei mattoni che la compongono.

Ma la proposizione 2D sembra assurda; infatti, qualunque sia il mattone emergente, esso sarà numerato con un numero naturale  $n$ , ma con qualunque numero naturale  $n$  esso sia numerato, ciò implicherebbe che tale mattone sia stato impilato dalla macchina al passaggio  $n$ -esimo del supercompito, per essere poi coperto dal mattone numero  $n + 1$ , da quello numero  $n + 2$ , e dagli infiniti seguenti. Si ha perciò una contraddizione: deve esserci un mattone emergente che mostra la sua faccia in cima alla pila, ma un qualsiasi siffatto mattone sarebbe numerato con un numero naturale, e come tale uscirebbe all' $n$ -esimo passaggio e sarebbe poi coperto da infiniti altri successivi, così da non poter mostrare la propria faccia. Il fatto che questo supercompito abbia implicazioni contraddittorie, viene usato come *reductio ad absurdum* contro la possibilità stessa del supercompito.

Questa seconda parte dell'argomento può essere strutturata nel seguente modo:

4P) Ogni mattone viene coperto dai mattoni numerati con numeri naturali maggiori.

5P) Se un mattone è coperto da altri mattoni non può esporre la propria faccia in cima alla pila finale.

3D) (Da 4P & 5P). Nessun mattone espone la propria faccia in cima alla pila.

4D) (Da 2D & 3D). Contraddizione.

Le premesse 1P e 2P sono motivate a partire da considerazione matematiche e geometriche, la 3P può essere motivata a partire da considerazioni mereologiche, 4P e 5P derivano direttamente dalla descrizione fisica del supercompito.

### 3 Obiezioni e possibili risposte

**1** Tutto l'argomento si basa proprio sull'idea che ci sia un ultimo mattone della pila, che alla fine sarà visibile alla sommità della stessa. Ma questo è falso, proprio alla luce del modo in cui il supercompito è definito. La conseguenza contraddittoria, dunque, è il risultato di una assunzione che, per quanto intuitivamente plausibile, rappresenta una esplicita contraddizione della descrizione stessa del supercompito, e quindi non può essere usata per argomentare contro la possibilità di quest'ultimo, perché si tratterebbe di una forma di ragionamento circolare.

**R.1** Nell'argomento non si sostiene direttamente l'esistenza di un *ultimo* mattone, bensì di un mattone *emergente*. Con "mattone emergente" deve intendersi un mattone che al termine del supercompito si trova collocato sopra a tutti gli altri, mentre con l'espressione "ultimo mattone" deve intendersi un mattone che viene aggiunto alla pila dopo tutti gli altri. Dal fatto che la prima definizione è in termini spaziali mentre la seconda è in termini temporali, dovrebbe essere chiaro che le due espressioni non hanno lo stesso significato, né segue soltanto da queste definizioni che un blocco emergente debba essere anche un ultimo blocco. Quindi, argomentare a partire da alcune ulteriori premesse che un blocco emergente dovrebbe essere anche blocco finale non è un caso di ragionamento circolare. Inoltre, la stessa esistenza di un blocco emergente non è assunta come premessa, ma dedotta da alcune plausibili premesse.

**2** La premessa 1 non è efficacemente difendibile, ed in particolare la prima difesa che di essa verrebbe da addurre prevede un passaggio immotivato da un infinito potenziale ad uno attuale. Più precisamente, non è detto che solo poiché la serie delle lunghezze dei mattoni converge ad un metro allora a  $t = 1$  secondi la pila sarà alta un metro: infatti una serie non è altro che un limite, ma nulla impone che la funzione  $A(t)$ , che indica l'altezza della pila in funzione del tempo, valga in 1 secondi quanto il suo limite in quel punto; la funzione potrebbe infatti presentare una discontinuità in corrispondenza del valore 1 secondi. In effetti, nella letteratura sono già presenti supercompiti come quello di Littlewood-Ross<sup>7</sup>, che prevedono delle discontinuità simili. Se la pila, pur avendo lunghezza tendente ad 1 metro per  $t$  che tende 1 secondi assumesse lunghezza infinita in  $t = 1$  secondi, allora non ci sarebbe più un mattone emergente, e l'intero argomento cadrebbe.

**R.2** Si può confutare l'obiezione nel modo seguente.

Come fatto poco fa, indichiamo con  $A(t)$  la funzione che indica la lunghezza (altezza) della pila in funzione dell'istante  $t$ . Inoltre indicheremo gli istanti in base al tempo trascorso dall'inizio del supercompito, sottintendendo come unità di misura i secondi.

1aP) Se esiste una lunghezza  $l > 0$  tale che per ogni istante  $x < 1$  vale  $A(x) + l < A(1)$  allora la pila subisce un allungamento istantaneo di lunghezza almeno  $l$  all'istante 1.

2aP) Se la pila è infinitamente lunga all'istante 1 allora esiste una lunghezza  $l > 0$  tale che per ogni istante  $x < 1$  vale  $A(x) + l < A(1)$ .

3aD) (Da 1aP & 2aP). Se la pila è infinitamente lunga all'istante 1 allora essa subisce un allungamento istantaneo all'istante 1.

<sup>7</sup>Si vedano a proposito Littlewood (1953) e Ross (1988).

4aP) Per ogni allungamento istantaneo della pila si ha che esso avviene in un istante strettamente precedente a 1.

5aD) (Da 3aD & 4aP). La pila non è infinitamente lunga all'istante 1.

La premessa 1a segue dal significato stesso di "allungamento istantaneo" cioè un aumento di lunghezza che avviene in un certo istante, facendo sì che ad un certo istante un oggetto sia più lungo di una certa misura rispetto a quanto era in ogni istante precedente. La premessa 2a si giustifica su basi puramente matematiche, considerando che la pila ha effettivamente lunghezza finita per ogni istante precedente a 1. La premessa 4a, infine, deriva dalla descrizione del supercompito, la quale prevede che tutti i mattoni vengano aggiunti in istanti precedenti l'istante finale del processo.

Si osservi che la considerazione fatta qui per il supercompito  $S$  può essere applicata anche al paradosso di Littlewood-Ross. Infatti, poiché nel paradosso il contenitore contiene un numero non nullo di palline in ogni stato del supercompito precedente alla sua supposta conclusione, e nell'istante finale del supercompito esso è vuoto, bisogna affermare che esso perde istantaneamente un certo numero di palline nell'istante finale. Ma per come è descritto il supercompito, ogni variazione nel numero di palline contenute avviene in un istante strettamente precedente a quello finale, e nell'istante finale del supercompito non avviene affatto alcuna variazione. Quindi, se ad esempio si avesse che il primo passaggio in cui vengono aggiunte dieci nuove palline e viene sottratta quella numerata con il numero più basso avviene dopo 0.5 secondi, il secondo passaggio analogo dopo altri 0.25 secondi e così via, nessuno degli infiniti passaggi del supercompito avverrebbe nell'istante finale. Questo perché, data la legge matematica appena espressa, l' $n$ -esimo passaggio del supercompito avviene a  $\frac{1}{2^n}$  secondi dall'istante finale. Ma  $\frac{1}{2^n}$  è diverso da 0 per ogni  $n$ , quindi non c'è alcun passaggio nell'istante finale del supercompito. Eppure in tale istante il contenitore diventa vuoto, pur avendo contenuto palline in ogni istante precedente. Si ha quindi una contraddizione.

**3** La premessa numero 2 dell'argomento principale è falsa nel caso dell'istante  $t = 1$  secondi. L'idea dell'esistenza di una faccia emergente in cima al blocco è basata su un'assunzione topologica errata. Lo spazio occupato in altezza dalla pila finale dovrebbe essere rappresentato da un intervallo semi-aperto a destra, del tipo  $[0, 1)$ . Infatti, se consideriamo lo spazio occupato in altezza dall' $n$ -esimo mattone come rappresentabile dall'intervallo chiuso:

$$1 - \left(\frac{1}{2}\right)^n, 1 - \left(\frac{1}{2}\right)^{n+1}$$

che chiameremo  $A_n$ <sup>8</sup>, allora lo spazio occupato in direzione verticale dalla pila finale sarebbe uguale all'unione di tutti questi intervalli, ovvero:

<sup>8</sup>Il primo blocco è qui considerato lo 0-esimo per semplicità notazionale.

$$\bigcup_{n \in \mathbb{N}} 1 - \left(\frac{1}{2}\right)^n, 1 - \left(\frac{1}{2}\right)^{n+1}$$

Ma questa unione infinita è uguale proprio all'intervallo aperto a destra  $[0, 1)^9$ . Quindi, per quanto possa sembrare difficile da immaginare che una pila di altezza finita possa mancare di una faccia superiore, sembra che questo sia proprio il caso dell'oggetto con cui abbiamo a che fare, poiché di fatto la regione bidimensionale di spazio che delimita superiormente la pila non è occupata dalla stessa. In altre parole, non esiste una sezione orizzontale della pila che abbia la proprietà di trovarsi sopra tutte le altre.

**R.3** L'interpretazione<sup>10</sup> topologica offerta in questa obiezione va incontro a notevoli difficoltà.

Infatti, si consideri il seguente principio:

*P*: Dato un oggetto  $X$  e un insieme di oggetti  $A$ , e detta  $Y$  la somma mereologica degli elementi di  $A$ ,  $X$  tocca  $Y$  se e solo se tocca almeno uno degli elementi di  $A$ .

Credo che si tratti di un principio molto plausibile ed intuitivo. Ad ogni modo, ne darò una difesa più avanti nel corso di questa risposta.

Adesso si consideri di appoggiare un cubo, la cui collocazione esatta<sup>11</sup> è rappresentata da una regione chiusa, sulla sommità della pila. Anche ammettendo che la regione di spazio occupata dalla pila sia un aperto e non comprenda lo strato che delimita superiormente la pila, si può comunque affermare, a partire da una plausibile definizione di contatto, che il cubo e la pila si toccano. Si consideri ad esempio la seguente definizione di contatto, proposta in Hudson (2001):

“Necessarily,  $x$  touches  $y$  if and only if  $\exists R1, \exists R2, \exists w, \exists v, \exists p$  (i)  $w$  is a part of  $x$ , whereas  $v$  is a part of  $y$ , (ii)  $w$  exactly occupies  $R1$ , whereas  $v$  exactly occupies  $R2$ , (iii)  $p$  is a boundary point of both  $R1$  and  $R2$  (iv)  $p$  is a member of at least one of  $R1$  and  $R2$ , and (v)  $w \neq v$ .”

A questo punto, tenendo conto del principio *P* e di questa definizione, è possibile introdurre la seguente *reductio ad absurdum*.

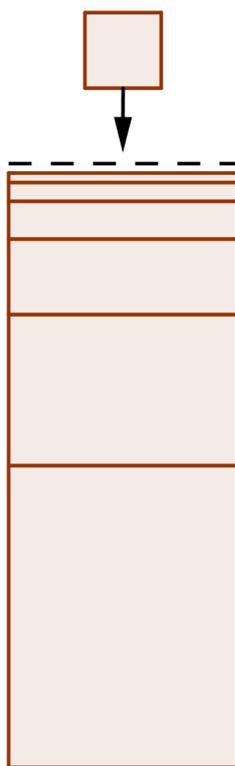
<sup>9</sup>La dimostrazione è semplice: si nota che ogni valore strettamente minore di 1 appartiene ad uno degli intervalli, e quindi all'unione, ma che allo stesso tempo 1 non appartiene ad  $A_n$  per alcun  $n$ , e quindi non appartiene all'unione.

<sup>10</sup>Parlo qui di interpretazione perché chiaramente gli oggetti fisici non sono, in sé stessi, astrazioni matematiche, e non è scontato che ogni astrazione topologica possa applicarsi a degli oggetti concreti. In effetti, è oggetto di dibattito tra i filosofi quali regioni di spazio possano essere *collocazione esatta* di oggetti fisici (si veda a tale riguardo Uzquiano 2006). Hudson, in Hudson (2008), chiama *C-theory* la teoria per la quale solo le regioni chiuse possono essere collocazione esatta di oggetti fisici.

<sup>11</sup>Per la definizione di collocazione esatta si veda la sezione “Which location relation is fundamental?” della pagina “Location and mereology”, Gilmore (2018). D'ora in poi utilizzerò le espressioni “collocazione esatta” e “regione occupata” come equivalenti.

- 1bP) Se X tocca la somma mereologica di un insieme di oggetti, allora tocca almeno un elemento dell'insieme.
- 2bP) La pila è la somma mereologica di tutti i mattoni.
- 3bP) Il cubo tocca la pila e non tocca alcuno dei mattoni.
- 4bD) (da 1 & 3) Se un oggetto tocca la pila, esso tocca almeno uno dei mattoni.
- 5bD) (da 4 & 5) Contraddizione.

La prima premessa corrisponde ad una delle due implicazioni previste dal principio *P*, e la seconda può essere presa come definizione della pila. Quindi, a patto di affermare il principio *P*, bisogna negare la terza premessa. Ma la terza premessa è implicata dal fatto che lo strato che delimita superiormente la pila non faccia parte della pila stessa. Quindi, con un' inferenza *modus tollens*, è possibile dedurre che tale regione deve essere occupata dalla pila.

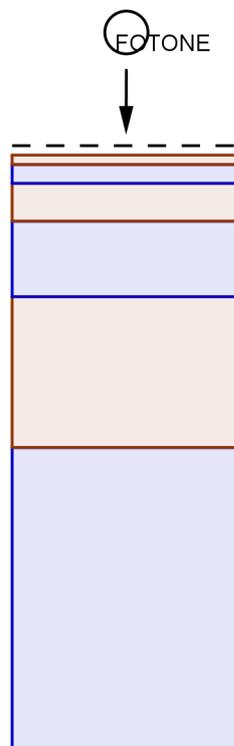


**Figura 1**

A questo punto propongo uno scenario che ritengo utile ad illustrare la plausibilità del principio *P*.

Si immagini che il macchinario che impila i mattoni, prima di posizionarli, li colori alternatamente di blu e di rosso. Ci si chieda di che colore apparirebbe la pila guardandola da sopra. Abbiamo stabilito in precedenza che la pila, anche

dalla parte della sua estremità superiore, può essere toccata; quindi anche un fotone può toccare la sua estremità superiore e riflettersi su di essa. Se però lo strato che delimita superiormente la pila non appartenesse alla pila stessa vista da sopra, essa non potrebbe apparire né blu né rossa. Infatti ogni mattone sarebbe coperto da infiniti mattoni sovrastanti, e quindi nessun mattone, né blu né rosso, sarebbe visibile da sopra alla pila. La pila sarebbe visibile, ma nessuno dei suoi mattoni lo sarebbe. Ma vedere la sommità della pila come blu oppure rossa richiederebbe appunto di vedere alla sua sommità la faccia superiore di un mattone rispettivamente blu oppure rosso.



**Figura 2**

C'è un senso in cui lo scenario appena descritto sembra presentare non soltanto una situazione controintuitiva e difficilmente immaginabile, ma un' autentica contraddizione. Infatti, sembra impossibile che un oggetto visibile, cioè luce-riflettente, sia al tempo stesso incolore. La luce è un'onda, e come tale deve avere in ogni istante una determinata lunghezza d'onda, la quale corrisponde ad un colore.

La apparente contraddizione appena evidenziata nasce dall'ammettere che un fotone possa toccare la pila senza toccare alcuno dei suoi mattoni. Sembra quindi di dover negare una tale possibilità, e questo va a corroborare il principio *P*.

## 4 Legami con altri supercompiti

La risposta all'obiezione numero 2 ha già dato l'occasione per presentare un argomento contro la possibilità del supercompito di Littlewood-Ross. Al di là di questo, si potrebbe pensare che il fatto che *S* sia impossibile non ponga alcuna difficoltà per la possibilità degli innumerevoli altri supercompiti.

A mio parere è però possibile sostenere che il supercompito *S* sia rilevante rispetto alla definizione stessa di supercompito. Infatti, ciò che è comune a tutti i supercompiti è l'esistenza dello *stato finale* di un processo, che infatti ha durata finita, pure in mancanza di *passaggio finale*, che manca proprio perché i passaggi sono infiniti. Ma come la faccia emergente in *S* deve appartenere necessariamente ad un mattone, così lo stato finale deve appartenere ad un tratto temporale del processo, cioè un intervallo di tempo che inizia con un passaggio del supercompito e che termina con il passaggio successivo, qualora ve ne sia uno. Questo tratto temporale a cui appartiene lo stato finale non può che essere l'ultimo, ma questo implica a sua volta l'esistenza di un ultimo passaggio.

L'analogia appena espressa è ancora più evidente se si accetta la teoria delle parti temporali<sup>12</sup>: in tale prospettiva i tratti temporali successivi sarebbero appunto come dei blocchi successivi sovrapposti l'uno agli altri, di lunghezza sempre minore, al pari dei mattoni della pila descritta nello scenario *S*.

Sempre in questa prospettiva, si può notare che l'esperimento mentale della Lampada di Thomson diventa del tutto analogo allo scenario dei mattoni a colorazione alterna: basta far corrispondere i tratti temporali in cui la lampada è spenta ai mattoni blu, e quelli in cui essa è accesa ai mattoni rossi.

Suggerendo questa analogia tra il supercompito *S* ed i supercompiti in generale sto esprimendo una linea argomentativa non troppo diversa da quella di Gwiazda (2012), secondo il quale il problema dei supercompiti è che in realtà non è concettualmente possibile giungere al termine di una serie di passaggi senza passare per un ultimo passaggio, il quale, appunto, non esiste nei supercompiti.

## 5 Conclusioni

In questo articolo si è mostrato come, a partire dalla descrizione del particolare supercompito preso qui in considerazione, è possibile dedurre proposizioni contraddittorie, sintomo dell'impossibilità del supercompito in questione. Si è poi mostrato come le stesse considerazioni possano essere usate per ottenere risultati contraddittori nel caso del paradosso di Littlewood-Ross e del paradosso della Lampada di Thomson e si sono discussi brevemente i legami del supercompito qui descritto con altri già presenti nella letteratura. In conclusione è

<sup>12</sup>Per una esposizione di questa teoria si veda Sider (2001).

opportuno notare anche che, nonostante in questo articolo si sia seguita la linea abbracciata da gran parte della letteratura sui supercompiti e sulla filosofia del tempo in generale, che prevede di descrivere gli intervalli di tempo a partire dagli istanti e di parlare senza particolari chiarimenti di eventi o cambiamenti che sono detti avvenire *in un certo istante*, alcuni autori<sup>13</sup> ritengono che sia più corretto e rispettoso della struttura del tempo definire gli istanti come dei concetti-limite a partire dagli intervalli, e analogamente parlare dei cambiamenti istantanei come di un limite di cambiamenti su intervalli di tempo concentrici progressivamente più piccoli. La scelta fatta in questo articolo di parlare di cambiamenti che avvengono *in un certo istante* deve intendersi come una preferenza stilistica ai fini della semplicità dell'esposizione, e non come una presa di posizione sulla reale natura dei cambiamenti istantanei o della questione più generale di quali siano più fondamentali nella struttura del tempo tra istanti ed intervalli<sup>14</sup>.

## Ringraziamenti

Ringrazio Agostino Pigozzi per aver proposto l'obiezione numero 2, Paolo Bussotti per l'aiuto fornitomi nella revisione di questo scritto ed uno dei referee per i suoi utili commenti.

<sup>13</sup>Vedasi ad esempio Oderberg (2006).

<sup>14</sup>Per una discussione approfondita di questo tema si veda Koons e Pickavance (2017, p. 415).

## Riferimenti bibliografici

- Benacerraf, P. (1962). "Tasks, super-tasks, and the modern eleatics". In: *Journal of Philosophy* 59.24, pp. 765–784.
- Burke, M. B. (2000). "The Impossibility of Superfeats". In: *Southern Journal of Philosophy* 38.2, pp. 207–220.
- Earman J. Norton, J. (1996). "Infinite pains: the trouble with supertasks". In: *Benacerraf and His Critics*. A cura di Adam Morton e Stephen P. Stich. Oxford: Blackwell, pp. 11–271.
- Gilmore, Cody (2018). *Location and Mereology*. A cura di Edward N. Zalta. The Stanford Encyclopedia of Philosophy. URL: <https://plato.stanford.edu/archives/spr2018/entries/location-mereology>.
- Groarke, L. (1982). "Zeno's Dichotomy: Undermining the Modern Response". In: *Auslegung* 9, pp. 67–75.
- Gwiazda, J. (2012). "A Proof of the Impossibility of Completing Infinitely Many Tasks". In: *Pacific Philosophical Quarterly* 93.1, pp. 1–7.
- Hudson, H. (2001). "Touching". In: *Nous* 35.15, pp. 119–128.
- (2008). *The Metaphysics of Hyperspace*. Oxford: Oxford University Press.
- Koons, R.C. e T. Pickavance (2017). *The Atlas of Reality: A Comprehensive Guide to Metaphysics*. Oxford: Wiley-Blackwell.
- Littlewood, J.E. (1953). *A Mathematician's Miscellany*. Londra: Methuen Co. Ltd.
- Manchak, John e Bryan W. Roberts (2016). *Supertasks*. A cura di Edward N. Zalta. The Stanford Encyclopedia of Philosophy. URL: <https://plato.stanford.edu/archives/win2016/entries/spacetime-supertasks>.
- Markosian, Ned (2016). *Time*. A cura di Edward N. Zalta. The Stanford Encyclopedia of Philosophy. URL: <https://plato.stanford.edu/archives/fall2016/entries/time/>.
- Oderberg, D.S. (2006). "Instantaneous Change Without Instants". In: *Analytical Thomism: Traditions in Dialogue*. A cura di C. Paterson e M.S. Pugh. Aldershot: Ashgate, pp. 101–118.
- Romero, G. E. (2014). "The Collapse of Supertasks". In: *Foundations of Science* 19.2, pp. 209–216.
- Ross, S.M. (1988). *A First Course in Probability*. USA: Macmillan.
- Sider, T. (2001). *Four Dimensionalism: An Ontology of Persistence and Time*. Oxford: Oxford University Press.
- Thomson, J.F. (1954). "Tasks and super-tasks". In: *Analysis* 15.1, pp. 1–13.

Uzquiano, G. (2006). "Receptacles". In: *Philosophical Perspectives* 20.1, pp. 427–451.





## Il progetto NOC (Natural Origins of Content): un bilancio critico

Marco Facchin

**Abstract.** Recentemente Hutto e Satne hanno proposto un nuovo programma di ricerca (NOC), al cui interno le strategie di naturalizzazione dell'intenzionalità neo-cartesiana, neo-behaviorista e neo-pragmatista vengono utilizzate come tre steps per comprendere il passaggio da una forma di *ur-intenzionalità* minimale e priva di contenuti alle forme di cognizione intenzionale che comprendono contenuti intensionali. In questo modo, NOC gioca un ruolo di rilievo nel permettere a REC, l'enattivismo radicale di Hutto, Myin e Satne, di fare i conti coi *representation hungry problem* e le forme complesse di cognizione. In questo *paper*, cercherò di valutare criticamente NOC. Proverò quindi a far emergere i punti di forza di NOC, intesa come una strategia, funzionale a REC, per risolvere i *representation hungry problem* da un punto di vista antirappresentazionalista. Al contempo, evidenzierò alcuni punti di debolezza, che dipendono da un'insufficiente caratterizzazione dell'*ur-intenzionalità*, leggibile come un ritorno al comportamentismo; da una visione troppo semplice del contenuto mentale rilevante nelle pratiche ascrivitive e da alcune ambiguità nella gestione della continuità darwiniana che intercorre tra *ur-intenzionalità* ed intenzionalità contenutistica. Appoggiandomi al *framework* della *skilled intentionality*, quindi, cercherò di proporre un modo di sviluppare l'*ur-intenzionalità* in termini non comportamentistici; e cercherò di mostrare che questa proposta, integrata con la *tesi della manipolazione* (la tesi che la manipolazione abile di strutture ambientali sia costitutiva di alcune capacità cognitive), può aiutare a comprendere il passaggio dalla cognizione minimale *ur-intenzionale* a forme cognitive più sofisticate e ricche di contenuto. Concluderò quindi indicando alcuni problemi aperti che NOC deve ancora fronteggiare. Il piano di lavoro è come segue: nel

primo paragrafo introdurrò NOC; nel secondo proporrò un bilancio critico che ne metta in luce i pro ed i contro, nel terzo avanzerò la mia proposta per separare l'*ur-intenzionalità* dal comportamentismo, e nell'ultimo concluderò con una breve rassegna dei problemi ancora aperti.

**Keywords.** Enattivismo, Intenzionalità, Representation Hungry Problems, Skilled Intentionality, Contenuto Mentale.



## 1 Il progetto NOC

NOC (*Natural Originis of Content*) è il nome del programma di ricerca proposto da Hutto e Satne (2015b) che mira ad individuare il modo in cui i contenuti mentali intensionali possono essere emersi a partire da semplici coordinazioni sensorimotorie. In questo *paper*, intendo proporre un bilancio critico, individuandone i pro ed i contro. In particolare, sosterrò che NOC fallisce nell'indicare in modo chiaro in cosa consista l'intenzionalità originaria, e cercherò di proporre una soluzione per questo problema.

NOC è collocato in un progetto (REC) di ampia portata, che mira a radicalizzare le posizioni enattiviste in filosofia della mente (Hutto e Myin 2013, 2017). L'assunto centrale delle posizioni enattive consiste nel sostenere che la cognizione non è in primo luogo una rappresentazione interna manipolabile secondo regole; ma è qualcosa che emerge dalle interazioni incorporate tra agente ed ambiente. In un senso generalissimo, un processo cognitivo è *enattivo* se è, del tutto od in parte, costituito dall'accoppiamento dinamico di processi interni all'organismo con appropriate strutture ambientali (Rowlands 2018, in pubblicazione). Hutto e Myin propongono di radicalizzare questa posizione eliminando il ricorso esplicativo a conoscenze tacite e/o rappresentazioni implicite (quali quelle in Noe 2004) e ponendo il *problema difficile dei contenuti* (HPC, Hutto e Myin 2013, cap. 4) che, in linea di principio, è supposto evitare che le posizioni enattiviste possano essere lette come mutazioni di posizioni cognitiviste (per un'interpretazione di questo tipo, vedi Dennett 2011).

HPC si basa su due semplici osservazioni: a) l'unica nozione scientificamente rispettabile di informazione è quella di *informazione come covarianza nomica* e b) la covarianza non è sufficiente a costituire i contenuti mentali nel senso rilevante (intensionale) del termine (Hutto e Satne 2015a). Perciò, il lessico rappresentazionale è, al più, un modo metaforico di comprendere ciò che la cognizione è. HPC non intende negare che *alcune* cognizioni, altamente derivate, possano comprendere dei contenuti. Intende negare che, *epistemologicamente*, il modo migliore per comprendere la cognizione, anche nei casi minimali, sia di fare riferimento ai contenuti; e *metafisicamente* che la cognizione sia essenzialmente un fenomeno di natura rappresentazionale. I contenuti mentali passano così dall'essere un elemento esplicativo, qualcosa nei cui termini spiegare qualcosa d'altro, ad essere un *target* delle spiegazioni: qualcosa *da spiegare* nei termini di qualcosa d'altro.

NOC ha come oggetto di studio proprio il passaggio dalle forme minimali e a-contenutistiche di cognizione a quelle più impegnative entro cui le rappresentazioni ed i contenuti hanno posto. Il punto di partenza di NOC è l'analisi di Haugeland (1990); in particolare l'individuazione di tre strategie (neo-cartesiana<sup>1</sup>,

<sup>1</sup>Nell'uso di Hutto e Satne, l'etichetta indica ogni posizione che da priorità esplicativa

neo-behaviorista e neo-pragmatista) di gestione dei contenuti mentali. Abbandonato l'assunto per cui "avere intenzionalità è avere un contenuto semantico" (Haugeland 1990), che è ovviamente una posizione inconsistente con REC, le tre diverse strategie vengono interpretate come tre *step graduali* che segnano il passaggio dalle forme cognitive minimali e prive di contenuto a quelle più impegnative e ricche di contenuto. In questo modo, Hutto e Satne intendono mostrare l'emergenza dei contenuti mentali a partire dalle interazioni basilari di un organismo con il suo ambiente.

La prima mossa di NOC è una re-interpretazione della migliore strategia di naturalizzazione dell'intenzionalità neo-cartesiana, individuata nella teleosemantica di Millikan (1984), in chiave non contenutistica. In questo modo, Hutto e Satne intendono fornire l'*ur-intenzionalità*, la forma minimale di rapporto agente-mondo tipica dei casi minimali e a-contenutistici di cognizione. La proposta di Millikan è attraente per molte ragioni: incorpora un concetto di normatività biologica, in virtù del ruolo della selezione naturale nel fissare alcune Funzioni Proprie, permettendo così subito di considerare normativamente i casi minimali di cognizione. La distinzione tra Funzioni Proprie (e spiegazioni) distali e prossimali permette anche di caratterizzare l'*ur-intenzionalità* in modo relativamente flessibile ed aperto all'apprendimento, non solo come una serie di tropismi meccanici innati. Questo permette di presentare l'*ur-intenzionalità* come un fenomeno intrinsecamente cognitivo, in grado di rendere conto dell'apprendimento e della flessibilità rilevante del rapporto agente-mondo.

Data l'*ur-intenzionalità*, si passa all'analizzare le pratiche di ascrizione rilevanti entro cui i contenuti mentali hanno posto. In questo modo, Hutto e Satne ritengono di poter utilizzare il lavoro fatto dai neo-behavioristi. L'atteggiamento intenzionale (Dennett 1987) viene ampliato distinguendo tra *agenti* e *pazienti* (*Patients*) intenzionali. I primi eseguono l'ascrizione, mentre i secondi ne sono i *target*. Agli agenti, e solo a loro, è chiesto di padroneggiare le pratiche rilevanti di manipolazione ascrivibile dei contenuti; mentre un paziente intenzionale è un qualsiasi sistema che soddisfi i requisiti per il possesso dell'*ur-intenzionalità*. Con l'introduzione dell'*ur-intenzionalità*, una proprietà reale del paziente intenzionale, Hutto e Satne intendono evitare di essere *meri* ascrivisti: intendono cioè cercare di rendere l'atteggiamento intenzionale qualcosa di *più* di un semplice "come se" euristicamente utile, rendendo ragione del fatto che l'intenzionalità ascritta non è *creata* dall'atto ascrivibile, ma è una reale proprietà del suo *target*. In questo modo, viene chiarito come i pazienti intenzionali siano intenzionali indipendentemente dall'ascrizione, risolvendo un problema dell'atteggiamento intenzionale (Gallagher 2017, cap. 4.1). Un secondo motivo di serietà dell'ascrizione dipende dal fatto che il contenuto ascritto può essere ragio-

ai contenuti mentali sulle norme socioculturali. Non implica necessariamente una visione *consciousness-centered* come nell'uso di Gallagher (2017).

nevolmente determinato osservando sia ciò che per *design* il sistema è supposto fare (Millikan 2000) sia la storia rilevante di interazioni che il sistema ha avuto con l'ambiente (Hutto e Myin 2017, cap. 5). L'ascrizione quindi è seria sia perché l'intenzionalità non è metaforica (Searle 1984) sia perché il contenuto ascritto è determinabile in modo chiaro dal *design stance* e dalla storia ontogenetica di relazioni rilevanti che il paziente intenzionale ha sviluppato con l'ambiente.

Il terzo *step* (neo-pragmatista), consiste nella collocazione delle pratiche di ascrizione nelle forme di vita socioculturali in cui esse sono realmente situate. L'*ur-intenzionalità* in questo contesto è supposta fornire la forma di direzionalità che alimenta i meccanismi di conformazione sociale alla base dell'emergenza delle norme proprie della forma di vita comune. In questo modo, Hutto e Satne intendono eliminare alcune tensioni presenti nella strategia neo-pragmatista. Ad esempio, il problema della legittimità del lessico intenzionale per descrivere gli *outgroup* (di cui fanno parte anche i neonati, vedi Haugeland 1990, Gallagher 2017, cap. 4.3) viene risolto dal fatto che gli *outgroup* sono comunque pazienti intenzionali. L'*ur-intenzionalità* fornisce anche una soluzione al problema di come le pratiche sociali (che per il neo-pragmatista permettono l'emergenza di una parte rilevante dell'intelligenza) possono essere costituite senza dover presupporre l'intelligenza che esse dovrebbero, per il neo-pragmatista, costituire.

Questo è, in breve, il progetto NOC. Una prima considerazione che mi sembra importante fare è che non c'è nessuna tensione tra NOC e HPC (Hutto e Satne 2015a, 2018). Per comprendere come mai, mi sembra agevole introdurre una distinzione tra due modi in cui i contenuti possono essere *originari* o *intrinseci*, cioè essere incorporati in un veicolo. I contenuti possono essere S-originari o H-originari. Un contenuto è S-originario nei casi in cui l'averne un contenuto è una proprietà *intrinseca* di un veicolo (come in Searle 1980) o nei casi in cui il possesso di un contenuto è una proprietà *necessitata, causata* o *costituita* da proprietà di più basso livello ma intrinseche al veicolo (come in Searle 1992). Viceversa, un contenuto è H-originario se l'incorporazione nel veicolo è *indipendente da una stipulazione* (Haugeland 1990). Hutto, Satne e Myin presentano HPC come un problema riguardo ai contenuti presenti *solo in virtù* delle covarianze nomiche tra stati funzionalmente rilevanti del macchinario cognitivo e i *target* ambientali (Hutto e Myin 2013, 2017, Hutto e Satne 2015b, 2018). Come tale, HPC colpisce solo i veicoli che sono supposti avere un contenuto S-originario, come ad esempio i processi neurali. HPC non colpisce invece i contenuti H-originari, di cui NOC si impegna a ricostruire l'origine.

La compatibilità tra NOC e HPC può essere espansa anche osservando il tipo di strategia naturalista che Satne, Hutto e Myin intendono utilizzare. Hutto e Satne propongono un naturalismo *rilassato* (Hutto e Satne 2015b, 2017, 2018) che accetta, tra i suoi strumenti, non solo quelli messi a disposizione dalle scienze

dure, ma anche quelli dell'etologia e della archeologia cognitiva, della psicologia culturale e della psicologia dello sviluppo. In questo modo, Myin, Satne e Hutto possono comprendere l'emergenza dei contenuti non solo come emergenza *fisica*, ma anche *storica* (a livello ontogenetico e filogenetico), aprendo uno spazio per i contenuti H-originari. Al contrario, i naturalisti *stretti*, che ricorrono esclusivamente alle risorse delle "scienze dure", sono costretti a comprendere l'emergenza dei contenuti *solo* come emergenza *naturale*, e quindi comprendere il contenuto solo come S-originario.

Sembra importante anche notare che neppure l'*ur-intenzionalità* è S-originaria, cioè non è una proprietà base dell'"inventario del mondo" né è necessitata, causata o costituita dalle proprietà di più basso livello che costituiscono i pazienti intenzionali. Modellando l'*ur-intenzionalità* sul modello di Millikan, infatti, Hutto e Satne ereditano il problema dello *swampman*. Se una coincidenza cosmica smaterializzasse la mia gatta e poi ne ri-materializzasse l'esatta replica, la seconda gatta *non* sarebbe un paziente intenzionale, dal momento che non rispetterebbe i requisiti che l'*ur-intenzionalità* eredita da Millikan (1984); in particolare l'appartenenza ad una Famiglia Riproduttivamente Stabilita. Oltre a chiarire la compatibilità tra NOC e HPC, questa discussione sul contenuto originario è importante poiché a REC, come altre posizioni *estese*, può venire mossa l'obiezione dei *contenuti non derivati* (Adams e Aizawa 2001, 2010). A seconda che quest'obiezione sia letta vertere su contenuti H-originari o S-originari può 1) non costituire (nel caso H-originario) affatto un'obiezione, oppure 2) può chiedere (caso S-originario) di sposare una posizione la cui insostenibilità è resa manifesta da HPC. In questo modo, REC può difendersi da un'importante obiezione alle posizioni estese.

NOC è un progetto ambizioso, che si propone di *ribaltare* i rapporti tra enativismo e cognitivismo. Alla luce di HPC non è il primo a dover giustificare il suo antirappresentazionalismo davanti al secondo, ma è viceversa quest'ultimo a dover giustificare il suo ricorso pervasivo al concetto di rappresentazione. Ci sono diverse ragioni per apprezzare NOC. In primo luogo, è un programma di ricerca che mette al suo centro la costruzione di un modo per affrontare i *representation hungry problems* (Clark e Toribio 1994); cioè è teso a spiegare come si possa passare dai casi minimali di cognizione intesa come semplice coordinamento sensomotorio ai casi più complessi, che sembrano coinvolgere contenuti mentali e concetti socialmente condivisi. Una seconda ragione per apprezzare NOC è che esso propone un *framework* rispettoso della continuità darwiniana (ma su questo tornerò in seguito) entro cui comprendere il pensiero non concettuale ed il suo rapporto con il pensiero concettuale. In questo modo, esso ci offre uno strumento, l'*ur-intenzionalità*, che è possibile utilizzare per comprendere i modi diretti, non verbalmente e/o concettualmente mediati, con cui un organismo può essere a contatto in modi significativi e *specificamente* co-

gnitivi con il suo ambiente. Inoltre, la differenza tra *ur-intenzionalità* ed intenzionalità semantica non mappa la dicotomia tra animali umani ed animali non umani. Gli infanti che non sono ancora stabilmente entrati in una forma di vita culturale socialmente condivisa possiedono *solo l'ur-intenzionalità*; così come gli agenti intenzionali che pure possiedono le pratiche culturali rilevanti con un certo grado di maestria possono a volte essere impegnati in rapporti cognitivi con l'ambiente che sono *ur-intenzionali*. Non è inoltre escluso *a priori* che sia impossibile per gli animali non umani, purché debitamente acculturati ed inseriti in una forma di vita in modo rilevante, maneggiare almeno alcuni contenuti semantici. In questo modo, la disputa tra pensiero concettuale e non concettuale, così come tra cognizione rappresentazionale e diretta, è *svicolata* dalla dicotomia animali umani/animali non umani in cui spesso si trova costretta, e può essere meglio affrontata. Un'ultima ragione per cui NOC può essere attraente ed ammirevole è il tipo di atteggiamento *ecumenico* che sembra adottare nei confronti delle scienze naturali. Anche chi persegue un "agenda strettamente unificatrice" (Hutto e Satne 2015b) dovrebbe confrontarsi con tutte le evidenze rilevanti disponibili.

## 2 Un bilancio critico per NOC

Purtroppo però ci sono valide ragioni per pensare che NOC, almeno nella forma in cui è stato per ora sviluppato, abbia poche possibilità di successo, dovute sia a problemi a livello concettuale che al mancato sviluppo di alcuni punti rilevanti nello svolgimento del progetto.

Un primo problema è che è difficile concepire l'*ur-intenzionalità* in termini non comportamentistici. L'ispirazione biologica e la concezione della cognizione come processo *on-line* di risposta a stimoli ambientali richiamano direttamente il comportamentismo, che è come REC "un approccio externalista, in primo luogo interessato agli stimoli ambientali ed alle loro risposte extraneurali ed extramentali" (de Oliveira e Chemero 2015). Ci sono varie ragioni per cui un collasso di REC (e NOC) nel comportamentismo sarebbe fatale. Anche supponendo che Hutto, Myin e Satne riescano a rispondere alle obiezioni che hanno messo in crisi il paradigma comportamentista (per una loro sintetica esposizione, vedi Piattelli Palmarini 2008, parte II cap. 1 e Kim 2011, cap. 3), il comportamentismo rimarrebbe una cattiva teoria psicologica, segnata da profondi limiti epistemici. Ad esempio, il comportamentismo non sembra poter spiegare i casi di *apprendimento osservativo*, cioè i casi in cui viene appreso un comportamento dopo *una sola* osservazione ed in assenza di rinforzo (Mason 2006, cap. 1, §2.3). Inoltre, se REC risultasse essere una forma di comportamentismo, REC e NOC sarebbero concettualmente incompatibili, dal momento che NOC metterebbe a disposizione quei contenuti mentali intenzionali che REC, come teo-

ria comportamentista, si impegna ad eliminare. Questo sarebbe un grave fallimento per REC, che, reintroducendo il comportamentismo, fallirebbe nell'indicare un'alternativa *percorribile* al cognitivismo rappresentazionalista. Ciò sembrerebbe anche restituire legittimità al rappresentazionalismo della scienza cognitiva classica, che è il bersaglio principale di REC. Inoltre, l'enattivismo, una corrente mentalista nata per superare il funzionalismo computazionale (Varela, Thompson e Rosch 1991), verrebbe ridotto alla posizione eliminativista ed antimentalista *par excellence*. In questa prospettiva, la risposta di Hutto, Myin e Satne sembra essere troppo debole. Insistere che l'*ur-intenzionalità* spieghi "fenomeni paradigmaticamente cognitivi" (Hutto e Myin 2017, pp. 14-17) e non solo il comportamento semplicemente non è sufficiente per distanziare questa proposta dal comportamentismo: anche il comportamentismo studiava alcuni fenomeni paradigmaticamente cognitivi; ad esempio l'apprendimento, che era comportamentisticamente compreso come condizionamento. L'osservazione che l'*ur-intenzionalità* rende anche conto dell'attività *autogenerata* dell'animale e non solo di quella che viene elicitata dagli stimoli (presente sostanzialmente di sfuggita in Hutto e Satne 2017) sembra una buona osservazione, ma insufficiente per distanziare REC dal comportamentismo. Anzi, potrebbe essere vista come un *raffinamento* del comportamentismo. Questo è un problema pressante, ma non credo debba essere esagerato. HPC, se preso seriamente, impone di trovare delle ragioni a sostegno della tesi che la cognizione debba essere caratterizzata come un fenomeno *primariamente* rappresentazionale; e "altrimenti torna Skinner" non è una ragione sufficiente (Chemero 2009, cap. 1 presenta un'analisi simile).

Io non credo che il problema sia il ritorno dei comportamentisti, ma il modo insoddisfacente in cui l'*ur-intenzionalità* è caratterizzata. Consideriamo la strategia per naturalizzare l'intenzionalità di Millikan (1984) (per comodità condenserò gli ultimi due punti in un unico punto. I termini tecnici compariranno con la maiuscola):

1. L'Icona Intenzionale è parte di una Famiglia Riproduttivamente Stabilita.
2. È Normalmente frapposta tra *Producer* e *Consumer*, reciprocamente standardizzati, la cui reciproca presenza è condizione Normale rispetto funzionamento.
3. Normalmente l'Icona Adatta il *Consumer* alle condizioni sotto le quali esegue la sua Funzione Propria.
4. Nel caso sia Imperativa, il *Consumer* produce Normalmente le condizioni che l'Icona Intenzionale Mappa secondo una regola; nel caso sia *Descrittiva* il modo in cui l'Icona Intenzionale Adatta il *Consumer* fa Normalmente riferimento al fatto che l'Icona Intenzionale Mappa secondo una regola.

Alcuni aspetti di questa strategia sono facili da interpretare in modo non rappresentazionale. Ma com'è possibile comprendere in modo non rappresentazionale "Mappa secondo una regola"? Credo che Myin, Satne ed Hutto non sviluppino a sufficienza questo punto, che, se sviluppato, potrebbe allontanare in modo significativo REC dal comportamentismo. Nel seguito, proverò ad indicare una soluzione.

Un secondo problema che NOC deve affrontare è che non sembra essere sufficientemente chiaro *come* il contenuto mentale sia collegato alle pratiche di ascrizione. Non intendo accusare Hutto, Myin e Satne di non aver ancora proposto una spiegazione positiva, dal momento che si sono per ora semplicemente impegnati a delineare NOC e non ancora svilupparlo. Piuttosto, mi pare che essi falliscano nell'operare una distinzione concettuale. Nell'ascrizione di un contenuto mentale ad un paziente intenzionale sono presenti due intenzionalità rilevanti e *due* contenuti rilevanti. Per comodità, userò la notazione di Searle (1983), con qualche modifica. Un'ascrizione intenzionale ha la forma:  $A_{paz}S(xy)$ ; dove  $A$  è l'atto ascrivito;  $paz$  è il paziente intenzionale,  $S$  lo stato ascritto e  $(xy)$  il contenuto che si ascrive. Ma se, come sembra essere (Hutto 2006, p. 20, Hutto e Myin 2013, pp. 79-82, Hutto e Myin 2017, cap. 5), il problema con i contenuti è che nessuna strategia naturalistica fissa l'intensione, anche in *questo caso c'è* un'intensione che vaga libera senza nulla che la fissi. Sono d'accordo con Hutto e Satne quando, seguendo Millikan (2000), sostengono che il *design stance* e la storia di interazioni rilevanti siano sufficienti per determinare i contenuti ascritti  $(xy)$ , e quindi la descrizione *sotto cui* il contenuto viene ascritto. Ma questo non è sufficiente per fissare il modo in cui il paziente intenzionale appare, cioè la descrizione *sotto cui* eseguiamo l'ascrizione, che è in grado di determinare in modo rilevante il contenuto che viene ascritto. Pensiamo ad un'ascrizione come  $A_{Marco}desiderio(dimangiarelapizza)$ : l'ascrizione a Marco del desiderio di mangiare la pizza. Poniamo anche il caso che sia fatta vedendo Marco aprire un cartone per la pizza d'asporto. Ma se scambiamo Marco per Mirco (che è il fratello gemello), famoso per il suo odio viscerale per la cucina italiana, difficilmente gli ascriverò un desiderio di mangiare la pizza – quanto piuttosto il desiderio di vedere se il cartone è vuoto così da poterlo buttare il prima possibile. Le intensioni rilevanti in gioco sono due; quella che determina  $xy$  e quella che determina  $paz$ . Quindi NOC ha, per ora, tematizzato in modo troppo semplice gli atti ascriviti: non solo essi ascrivono un contenuto a un paziente intenzionale, ma pure ne contengono un contenuto che caratterizza *intensionalmente* il paziente intenzionale, a cui siamo diretti *come*<sup>2</sup> (Hutto 2006); in un modo che coinvolge rilevantemente i contenuti.

In modo connesso ma logicamente indipendente, credo si presenti un secondo problema. Capisco che l'essere umano abbia una nicchia ecologica emi-

<sup>2</sup>In analogia a "vedere come".

nementemente socializzata. Capisco anche che le pratiche discorsive, entro cui la concettualità umana è forgiata, sono sorrette dalla *divisione sociale del lavoro linguistico* (Putnam 1975). Può anche essere vera la tesi *empirica* per cui ciò che contraddistingue le capacità cognitive umane è principalmente la capacità di agire impegnandosi in azioni condivise (Tomasello 2014). Nondimeno, mi pare che comprendere l'emergere dei contenuti mentali, e l'accesso umano alla sfera dei significati, *solo* a partire dalle prassi sociali di *ascrizione* e dal gioco sociale e pubblico del dare e chiedere ragioni sia restrittivo. Anche concedendo, come credo sia giusto fare, che il significato è basato in un senso importante sulla divisione del lavoro linguistico ed è perciò culturalmente stipulato, mi sembra abbastanza evidente che *almeno alcuni* contenuti mentali non nascano dalle prassi sociali *di ascrizione*. Consideriamo il concetto di numero (senza sofisticazione matematica); e supponiamo che Malafouris (2013) fornisca la ricostruzione giusta della sua origine. In quest'ottica, il concetto di numero si è formato lentamente, all'interno di una prassi sociale legata alla tassazione, a partire da un associazione 1:1 tra un certo bene posseduto (come una cesta di grano) ed una particolare forma d'argilla (come un piccolo cono). Si è poi passati ad archiviare le forme d'argilla utilizzate, specifiche per un bene, in delle sfere di creta, che sono in seguito state sostituite con delle tavolette d'argilla, su cui le forme venivano impresse, sempre in un rapporto 1:1 con lo specifico bene che contavano. Solo in seguito si passò dal distinguere il segno per il bene (come il cono che indica la cesta di grano) dal segno per la sua quantità (ad es, un cerchio per 10, una linea per 1). In questo modo, per contare 10 ceste di grano non era più necessario imprimere nell'argilla 10 volte il cono, ma era sufficiente imprimerlo una volta e farlo seguire da un cerchio. In questo modo, il numero era separato dalla cosa che veniva contata, e lo stesso segno poteva indicare la stessa quantità di oggetti diversi. Lentamente, i segni impressi dalle forme vennero sostituiti con fonogrammi, usati per indicare sia il bene che la sua quantità, come facciamo anche oggi quando scriviamo "dieci ceste di grano". Questa è una ricostruzione stilizzata di come *potrebbe* essere nato il nostro contenuto mentale di numero, in modo distinto dagli oggetti numerati. Se questa ricostruzione è giusta, è vero che il numero è nato in una prassi sociale, ma non è vero che, come afferma NOC, è nato in una prassi sociale *ascrittiva*. In questo senso, l'attenzione specifica che NOC riserva alle prassi sociali *ascrittive* come produttrici di contenuto sembra essere ingiustificatamente stretta. Ma se il *focus* esplicativo si allarga alle prassi sociali in questo senso più generico, l'ipotesi di NOC diventa una riproposizione, allargata al mondo sociale, della *tesi della manipolazione* (l'affermazione che alcuni processi cognitivi sono costituiti anche da pratiche manipolatorie pubbliche, vedi Menary 2007, 2010). REC è compatibile con questa tesi; ma la sua adozione modifica NOC in modo rilevante.

Ho anche sostenuto che uno dei punti attraenti di NOC è il modo in cui riesce

a gestire la *continuità darwiniana*, ed il modo in cui riesce a rendere ragione della dicotomia tra pensiero concettuale e non-concettuale in termini che non sono quelli di “umani *vs* non umani”. Questi sono due aspetti positivi, ma la posizione di REC su questi temi oscilla in modo abbastanza confuso. In Hutto e Satne (2015b) leggiamo: “This assumes that there are two kinds of intentionality”; in Hutto e Satne (2017) è sostenuto che la continuità *biologica* tra l’uomo ed il resto degli animali non implica necessariamente la continuità *psicologica*. Questa è una lettura *moderatamente* saltazionista. Sono saltazioniste quelle posizioni che, data la dicotomia pensiero umano/non-umano, ritengono che la loro differenza sia di genere (le continuiste ritengono sia di grado). NOC è *moderatamente* saltazionista per due motivi: 1), Hutto e Satne (2017) sono chiari nel sostenere che l’avvento dei contenuti non trasforma la cognizione in qualche cosa di diverso da un rapporto dinamico (anche se contenutisticamente mediato) con l’ambiente; 2) essendo la dicotomia tra *ur-intenzionalità* ed intenzionalità concettualmente mediata neutra rispetto alla divisione tra umani e non umani, non pone *necessariamente* una differenza di genere tra i pensieri della nostra specie e quelli di tutte le altre. In particolare, per come NOC è delineato, non è escluso *a priori* che ci siano pratiche culturali a cui gli animali non umani possono prendere parte in cui almeno alcuni rapporti intenzionali con l’ambiente siano concettualmente mediati. Questa possibilità è tanto più reale quanto meno i contenuti sono legati alle pratiche *specificatamente* ascrivibili, che sono unicamente umane. Tuttavia, Hutto e Satne (2017) articolano diversamente questo passaggio: come certe conversazioni coinvolgono del caffè, senza essere basate su o orientate al caffè e rimanendo dello stesso *genere*, così certi rapporti intenzionali coinvolgono dei contenuti senza diventare basate sui contenuti o orientate ai contenuti, e rimanendo dello stesso *genere*. Queste due letture sono ovviamente incompatibili. Entrambe possono essere sostenute, ed entrambe rispettano la continuità biologica darwiniana; anche se una sola rispetta la continuità *psicologica*. Credo che la versione psicologicamente *continuista* sia da preferirsi per almeno due ragioni. La prima è che se un solo *genere* di intenzionalità è sufficiente, non è necessario, per parsimonia, postularne un altro. Questa è una ragione banale. Una ragione più specifica è che se effettivamente l’intenzionalità fosse di *due generi*, NOC farebbe i conti con i *representation hungry problem* in modo meno convincente. Questo tipo di obiezione sfida l’antirappresentazionalista a trovare un modo per gestire certi fenomeni cognitivi (riguardanti situazioni controfattuali, assenti o astratte) senza fare ricorso a rappresentazioni interne all’agente. Per rispondere, l’antirappresentazionalista deve trovare un meccanismo *essenzialmente non-rappresentazionale* che permetta quel genere di cognizioni. Ad esempio, il ricorso a situazioni surrogate (Clark 2005) può essere una risposta: il *target* assente, controfattuale o astratto, è rimpiazzato da un appropriata struttura ambientale che lo surroga, permettendo all’agente cogni-

tivo un impegno *on-line* che risulta in una cognizione *off-line* sul *target* assente, astratto o controfattuale. In questo caso, il centro della spiegazione è la dinamica attiva che accoppia agente e struttura ambientale, ed è qualcosa di *essenzialmente* non rappresentazionale. Se l'intenzionalità semantica è solo un reale rapporto *ur-intenzionale* con qualcosa di concreto che incorpora un contenuto semantico di qualche tipo, allora NOC ha (potenzialmente, perché la spiegazione positiva deve ancora essere prodotta) una risposta<sup>3</sup>. È meno chiaro se l'abbia sostenendo che ci siano due generi di intenzionalità – entrambi *essenzialmente* attivi, ma differenti per il coinvolgere o non coinvolgere contenuti.

Se questa fosse la posizione, allora, almeno dal punto di vista metafisico, le posizioni rappresentazionaliste vincerebbero, dal momento che questo è un modo “*soft*” per affermare che alcune pratiche cognitive *devono necessariamente* coinvolgere, per essere spiegate, dei contenuti, che quindi devono essere postulati. Questa è una vittoria dei rappresentazionalisti, che *non* hanno problemi ad includere nei loro schemi esplicativi casi in cui la cognizione dipende da semplici dinamiche sensomotorie: è sufficiente affermare che quei casi sono casi di cognizione minimale, e che la cognizione *davvero interessante* richiede delle rappresentazioni. Perciò dividere “in due” l'intenzionalità è una vittoria per i rappresentazionalisti: perché concede loro il monopolio sulla “cognizione interessante”. Uno dei compiti che l'antirappresentazionalista è *anche* quello di mostrare che c'è un modo più economico, privo di contenuti mentali, per spiegare i casi di cognizione *davvero interessanti* (Chemero 2009, cap. 5).

### 3 Un modo per chiarificare l'*ur-intenzionalità*

Ho sostenuto sopra che un problema rilevante per il progetto NOC è quello di una insufficiente tematizzazione dell'*ur-intenzionalità*, dipendente dall'assenza di un modo ovvio per comprendere in termini non semantici il “Mappare secondo una regola” che le Icone Intenzionali sono supposte fare. Ora desidero provare a fornire una soluzione. Ovviamente, come NOC, non mi impegno in un'ipotesi empirica di individuazione del macchinario cognitivo che costituisce l'*ur-intenzionalità*, ma solo in una sua chiarificazione concettuale utile a sviluppare NOC coerentemente.

Kiverstein e Rietveld (2015) hanno, in opposizione alla proposta dell'*ur-intenzionalità*, proposto una visione dell'intenzionalità minimale basata sulle abilità dell'agente. Ci sono tre motivi preliminari per cui la loro proposta è interessante per sviluppare in modo più completo NOC. Il primo è che anch'essi

<sup>3</sup>Questo è un modo di leggere l'*ur-intenzionalità* (e NOC) secondo le linee della *tesi della manipolazione* (Menary 2007, 2010). Su questo punto tornerò in seguito. In ogni caso, questa lettura è compatibile con REC – che anzi sembra sostenere che *ogni cognizione comprende una manipolazione* (Hutto e Myin 2013, cap. 7).

muovono all'*ur-intenzionalità* l'accusa di essere fondamentalmente definita in termini comportamentistici. In questo senso, la loro proposta dovrebbe incorporare almeno alcuni strumenti che permettano di comprendere la cognizione minimale in modo non comportamentistico. Il secondo motivo è che essi ritengono che il ricorso esplicativo alle abilità dell'organismo permetta di comprendere i fenomeni cognitivi minimali in modo non *environment-driven*. Questo permette di comprendere l'agente cognitivo come un attore con fini ed interessi propri, e non solo come un meccanismo che seleziona ed effettua delle risposte agli stimoli che l'ambiente gli somministra. L'ultimo motivo di interesse, almeno preliminarmente, è che anche Kiverstein e Rietveld ritengono che la differenza tra *ur-intenzionalità* ed intenzionalità semantica sia meglio compresa come una differenza di *tipo* e non di *genere*; anche se il loro argomento a proposito è diverso da quello che ho presentato.

Ora, in che modo il suggerimento di Kiverstein e Rietveld di dare alle abilità possedute dall'organismo un ruolo importante nella definizione dell'intenzionalità minimale, aiuta a comprendere meglio l'*ur-intenzionalità*? Un primo modo è che il ricorso esplicativo alle abilità incorporate dell'agente ci può permettere di comprendere il Mappaggio che l'Icona Intenzionale è supposta compiere in termini non semantici. Ci sono ragioni per sostenere i movimenti fisici di cui l'esercizio di un'abilità è costituito siano *veicoli informativi*. Becchio e al (2012), ad esempio, hanno mostrato come il *pre shaping* che la mano effettua durante un'azione di afferramento sia un latore di informazioni rispetto all'intenzione che anima la presa e all'oggetto a cui la presa è diretta. Questo però è ancora troppo poco per giustificare una lettura delle azioni come Icone Intenzionali. Dopotutto, un cielo nuvoloso porta informazioni sulla pioggia che cadrà, ma non *rappresenta* la pioggia che cadrà, né ha la *funzione* di farlo.

Perché un'azione sia *leggibile* come una rappresentazione, il semplice fatto di incorporare dell'informazione non è sufficiente; è richiesto anche il rispetto dei vincoli tipici a cui i fenomeni rappresentazionali sono sottoposti. In particolare, i movimenti che costituiscono l'esercizio di un'abilità sono soggetti a *vincoli informativi* (covariano regolarmente con i target rilevanti dell'abilità, come una rappresentazione covaria regolarmente rispetto a ciò che rappresenta); sono soggetti a *vincoli normativi* (ci sono modi *giusti* e *sbagliati* di esercitare le abilità rilevanti, come ci sono modi giusti o sbagliati di rappresentare). I movimenti possono *mal rappresentare* (nel caso in cui non siano debitamente accoppiati con i *target* ambientali rilevanti, esattamente come una rappresentazione mal rappresenta nel caso in cui alcune sue caratteristiche non siano debitamente accoppiate al fenomeno rappresentato), esibiscono una *struttura compositiva* (come una rappresentazione rappresenta solo per mezzo del modo in cui i suoi elementi sono arrangiati, così un'abilità seleziona un *target* solo per mezzo del modo in cui i movimenti di cui il suo esercizio è composto sono svolti) e so-

no *separabili* rispetto al loro obiettivo, almeno nel senso minimale che possono fallire. Questi sono *esattamente* i vincoli a cui sono soggette le rappresentazioni (Rowlands 2006).

Il rispetto di questi vincoli, però, non ci costringe a leggere l'esercizio di un'abilità semanticamente. La prima cosa da notare è che l'asse normativo entro cui si dispone l'esercizio di un'abilità è orientato secondo le polarità del *successo/insuccesso*, non quelle del *vero/falso* (Haugeland 1990). Questa è una prima ragione per cui il possesso e l'esercizio di un'abilità sono differenti in modo rilevante dalle rappresentazioni semantiche standard, di tipo descrittivo-enunciativo, nonostante rispettino gli stessi vincoli. Inoltre, l'esercizio di un'abilità, anche assicurato il rispetto di questi vincoli, non è un fenomeno intensionale. La mia abilità rilevante di afferrare al volo un pallone non implica che io lo debba afferrare (o riconoscere) *come*, ad esempio come il pallone che mi ha autografato Totti. L'esercizio di un'abilità è in questo senso slegato in modo importante dalla presentazione dei suoi *target* sotto una descrizione. Questo è vero anche delle Icone Intenzionali: esse possono intenzionare un *target* senza *identificarlo come*, cioè possono "prenderlo di mira" senza presentarlo sotto una descrizione (Millikan 1984). Va anche notato che l'esercizio di un'abilità è posto tra un *producer* ed un *consumer*, identificabili in prima battuta rispettivamente come i meccanismi di produzione e monitoraggio del movimento. In questo senso, al pari della familiare nozione di rappresentazione mentale, anche l'esercizio di un'abilità è *interveniente* all'interno della percezione e dell'azione, intese in un senso ampio (Haugeland 1990).

Certo, è interveniente in un senso diverso rispetto al modo in cui le familiari rappresentazioni mentali sono intervenienti; ma questo non costituisce un problema, dal momento che NOC è un progetto di ricerca che opera all'interno del *framework* dell'esternalismo attivo o dei veicoli (Clark e Chalmers 1998, Hurley 2010). Per ristabilire l'aria di familiarità è sufficiente che sia rispettato il *principio di parità* (Clark e Chalmers 1998) e, dal momento che l'esercizio di un'abilità 1) appartiene a una Famiglia Riproduttivamente Stabilita, 2) è frapposto tra *Consumer* e *Producer* standardizzati e che sono condizioni Normali del reciproco funzionamento, 3) adatta il *Consumer* alle condizioni sotto il quale esegue la sua Funzione Propria e 4) esegue un Mappaggio secondo una regola in un senso rilevante, mi pare che il principio di parità sia rispettato. Questo non è sorprendente per chi ricorda l'insistenza di Millikan (1984) sul fatto che *Producer* e *Consumer* non debbano necessariamente essere nello stesso organismo. In questo senso, la proposta di Millikan è amica dell'esternalismo attivo, dal momento che è *intrinsecamente* neutra rispetto ai vincoli somatici "della pelle e del cranio" (Clark e Chalmers 1998).

Quindi l'esercizio di un'abilità rispetta pienamente le condizioni che caratterizzano l'*ur-intenzionalità* e ci permette di comprendere non intensionalmente

te (e non semanticamente) il “Mappaggio”, che i veicoli informativi sono supposti compiere. L’esercizio di un’abilità è strutturalmente isomorfo ad una rappresentazione, ma non perciò è una rappresentazione: non è cioè un oggetto con la Funzione Propria di presentare un *target* secondo una descrizione. Piuttosto, l’esercizio di un’abilità può essere *letto come* una rappresentazione, è accessibile attraverso un *atteggiamento rappresentazionale* rilevante (Chemero 2009, cap. 4). Per capire cosa sia l’atteggiamento rappresentazionale, possiamo immaginare di vedere un cielo scuro e nuvoloso. Ora, non c’è nessun senso *legittimo* in cui il cielo ci “dice” qualcosa: non *dichiara* che sta per piovere, né *comanda* (per far giustizia alle Icone Intenzionali imperative) di cercare l’ombrello più vicino. Nondimeno, rispetto ad esso c’è un atteggiamento rilevante per cui si può guardare al cielo *come se* dicesse queste cose – esattamente come l’atteggiamento intenzionale ci permette di guardare al comportamento di sistema intenzionale *come se* dichiarasse i suoi stati mentali (Dennett 1987); anche se in entrambi i casi non è una Funzione Propria (né del comportamento né del cielo) veicolare quelle informazioni, è possibile fare *come se* lo fosse.

È importante notare che l’atteggiamento rappresentazionale è compatibile con REC in un senso importante<sup>4</sup>. REC si impegna a sostenere che non è necessario utilizzare il concetto di rappresentazione per *spiegare* in che modo il macchinario cognitivo produce il comportamento di un agente. REC, quindi, nega che la *spiegazione* di come il macchinario cognitivo svolge il suo lavoro debba essere rappresentazionale. Presentando HPC, nega anche che ci sia un senso legittimo in cui è possibile sostenere che la covarianza nomica tra *target* ambientale e stati del macchinario cognitivo sia *sufficiente* per permetterci di considerare questi ultimi come *rappresentazioni semantiche* dei primi. Tuttavia l’interpretazione rappresentazionale di un’abilità non si dispone sull’asse semantico del vero e del falso, ed è quindi al riparo di HPC. Inoltre, l’interpretazione rappresentazionale non riguarda il modo in cui un’abilità è prodotta o esercitata: non riguarda il modo in cui funziona il macchinario cognitivo necessario per esercitare un’abilità. In altri termini, non *spiega* (né pretende di spiegare) cosa succede quando un agente esercita la sua abilità rilevante di afferrare al volo un pallone. Piuttosto, si concentra sui movimenti che compongono l’esercizio di un’abilità e chiarifica *in virtù di che cosa* questi movimenti costituiscono un legame intenzionale tra l’agente ed il suo *target*.

Quest’atteggiamento è di per sé *neutro* rispetto all’esistenza di rappresentazioni (nel senso familiare del termine) come struttura interne all’agente su cui delle computazioni sono performati, e permette di parlare di rappresentazioni (purché se ne specifichi il senso) ogni qual volta può esserci euristicamente utile (Gallagher 2017, cap. 5.6). L’interpretazione “rappresentazionale” dell’esercizio di un’abilità è anche in grado di produrre contenuti compatibili HPC e,

<sup>4</sup>Ringrazio i *referee* per avermi stimolato a sviluppare questo punto.

quindi, anche con NOC e REC<sup>5</sup>. HPC si basa sull'osservazione che l'unica nozione scientificamente *accettabile* di informazione che abbiamo a disposizione è la nozione di informazione come *covarianza* e che essa è insufficiente a costituire i contenuti mentali. Ma questo tipo di obiezione colpisce *solo* i contenuti informativi S-originari, cioè i contenuti intenzionali supposti essere presenti indipendentemente da pratiche culturali rilevanti (ad esempio la *representational stance*). Il contenuto delle abilità è tuttavia H-originariamente presente (in modo compatibile con il terzo *step* di NOC): pur non essendo *costituito, causato* o *necessitato* dai movimenti che costituiscono l'esercizio di un'abilità è nondimeno *impermeabile* rispetto alle stipulazioni. Una prassi sociale ascrittiva sta alla base della lettura del mio prendere al volo la palla come l'estrinsecazione del mio desiderio di non prendere pallonate in faccia e, allo stesso modo, una prassi sociale sta alla base del vedere i movimenti fatti *come* covarianti, normativamente regolati, composizionalmente composti, separabili e potenzialmente mis-rappresentatori; ma questo non implica che il loro significato sia *liberamente contrattabile*. Non sembra essere *concretamente* possibile accordarsi per vedere quei movimenti come rappresentanti l'orbita di Venere, nello stesso modo in cui non sembra essere concretamente possibile interpretare in quel modo un romanzo di Joyce.

Il riferimento all'esercizio delle abilità rilevanti dell'agente permette anche di poter comprendere l'*ur-intenzionalità* in termini non comportamentistici, per diverse ragioni. In primo luogo, il riferimento alle abilità dell'agente ci permette di comprendere l'ambiente come un "ricco paesaggio di *affordances*" (Kiverstein e Rietveld 2014). Ci sono due sensi importanti in cui questo ci distanzia dal comportamentismo. Il primo è che le *affordances* non sono stimoli: sia perché sono descrivibili in termini mentalistici (come il "significato" che qualcosa ha per un agente) sia perché "*affordance*" non designa l'energia che un recettore capta. In secondo luogo, il ricco paesaggio di *affordances* non somministra *uno* stimolo a cui segue *una* risposta, ma è un luogo in cui diverse salienze hanno contemporaneamente un peso (variabile) a cui l'agente cognitivo è sensibile. È vero che il rapporto *ur-intenzionale* è costituito dall'*esercizio* di un'abilità e non solo dal suo possesso; ma questo non deve spingerci ad una visione ciclopica del rapporto originario tra organismo ed ambiente. Al di là del fatto che è possibile esercitare più di un'abilità per volta e quindi avere un rapporto *ur-intenzionale* con più di un aspetto ambientale rilevante contemporaneamente, l'ambiente, anche se non direttamente *ur-intenzionato* dall'organismo, è comunque presente per lui nelle possibilità di azione rilevanti che offre: è *virtualmente* presente nella sua interezza (Noe 2004). Questo mi sembra riuscire a gestire – almeno

<sup>5</sup>È importante ricordare che REC nega solo il ricorso al contenuto *semantico*. Altri generi di contenuto non sono minacciati. In particolare, Hutto e Myin sembrano felici di accettare i contenuti fenomenici ed intenzionali (Hutto e Myin 2017, pp. 10-13).

in linea di principio – la tensione presente nell'enattivismo tra *esercizio* e *possessione* delle abilità rilevanti (Rowlands 2010, Maye e Engel 2015). Inoltre, le contingenze sensomotorie che l'organismo sfrutta nell'ambiente possono essere “di alto livello”, cioè essere utilizzate per individuare proprietà astratte o categoriali (Beer 2003, Pfifer e Bongard 2006, cap. 4). L'introduzione di dinamiche di questo tipo è un altro passo che distanzia l'*ur-intenzionalità* dal semplice rapporto stimolo-risposta del comportamentismo, dal momento che fa riferimento alle capacità dell'agente di sviluppare ed impiegare comportamenti esplorativi di *determinazione* degli oggetti che non sono la risposta a nessuno specifico stimolo ambientale.

Questi sono modi importanti in cui l'*ur-intenzionalità* si distacca dal comportamentismo e soprattutto dall'obiezione basata sull'argomento di povertà dello stimolo. In effetti, a differenza degli stimoli dei comportamentisti, le *affordances* permettono la creazione di un *bootstrapping paradise* (Clark 2016, cap. 1.3) per l'acquisizione di informazione. Sia dal punto di vista filogenetico che ontogenetico, la presenza di un'*affordance* può elicitare un'azione, che può aumentare o modificare le abilità rilevanti possedute da un agente, a cui appaiono quindi *nuove affordances* e nuove possibilità di azione (che possono risultare in nuove *affordances* e così via). Questa è una possibilità reale, sebbene la sua portata sia ancora da determinare con precisione.

Questo tipo di osservazione si sposa bene con la *tesi della manipolazione* (Menary 2007, 2010). Essa sostiene che almeno alcune capacità cognitive umane dipendono in modo cruciale dalla manipolazione, per mezzo di abilità rilevanti, di appropriate strutture ambientali; o, in modo equivalente, che almeno alcune capacità cognitive umane sono costituite dall'accoppiamento dinamico di strutture ambientali e processi interni. In questa seconda formulazione, essa è chiaramente compatibile con l'enattivismo ed in generale con le prospettive *action-oriented* nelle scienze cognitive (Menary 2015). Per quanto riguarda REC (e quindi NOC, che è una parte di REC), la tesi della manipolazione è non solo *accettabile*, ma anche da generalizzare (Hutto e Myin 2013, cap. 7). Per REC la cognizione, anche nei casi minimali, è *estensiva*, cioè è sempre almeno parzialmente costituita dalla manipolazione abile di alcuni aspetti ambientali. Questa tesi, in fondo, non è che la logica conseguenza della refutazione della cognizione come computazione su di una rappresentazione.

Il ricorso alla *tesi della manipolazione* permette anche di comprendere come l'*ur-intenzionalità*, chiarificata per mezzo del ricorso alle abilità rilevanti possedute dall'agente, possa coinvolgere contenuti semantici senza subire un mutamento di *genere*. In questi casi, i contenuti semantici sono *fattualmente incorporati* da adeguati oggetti esterni e strutture ambientali (come ad esempio dei segni grafici<sup>6</sup>), che sono coinvolti, coll'agente, in un rapporto *ur-intenzionale*

<sup>6</sup>Ma non solo. Per vedere la tesi della manipolazione in opera su cose diverse da segni grafici, vedi

di manipolazione abile, socioculturalmente regolata da norme manipolatorie ed interpretative. L'utilizzo abile di queste norme rilevanti permette all'agente di essere in contatto in modo significativo con i contenuti rilevanti, all'interno di un rapporto che è essenzialmente *ur-intenzionale*.

Anche in questo caso, come è per le abilità, il contenuto è H-intrinseco (non essendo le norme sociali che governano la manipolazione e l'interpretazione *contrattate in loco*, ma semplicemente *tramandate* dalla tradizione) ma non S-intrinseco. In questo modo, la tesi della manipolazione ci permette di avere un tipo di contenuti che resiste all'obiezione HPC, cioè compatibile con REC. Ovviamente, questo tipo di contenuto è esattamente il contenuto di cui NOC intende mostrare l'origine. Inoltre, il fatto che nell'esercizio di un'abilità il contenuto venga performato sensomotoricamente (*Enacted*, Menary 2006) mentre nei simboli è codificato (*Encoded*; Menary 2006) non sembra essere un motivo *sufficiente* per introdurre una differenza di genere, precisamente perché la *codifica* che “mette dentro” un contenuto ad un simbolo è dipendente dall'*enazione* di quel contenuto per mezzo di un'abilità rilevante. È appunto l'esecuzione delle abilità rilevanti appropriate ai rilevanti oggetti fisici a renderli H-originariamente significativi. È vero che davanti al segno “STOP!” non devo *impormi* di utilizzare la *representational stance*, mentre una simile imposizione è richiesta per guardare all'esercizio di un'abilità come ad un veicolo informazionale; ma anche questa osservazione non è decisiva. Può darsi che in certi casi il *representational stance* sia diventato semplicemente una “seconda natura”, esattamente come lo è l'*intentional stance* quando abbiamo a che fare con le persone (ma non con altri sistemi intenzionali).

Il ricorso alla tesi della manipolazione solleva il problema di come l'agente possa seguire le norme che regolano la manipolazione *senza* doverle rappresentare internamente<sup>7</sup>. Questa è un'obiezione importante, perché sembra minare la capacità di NOC di illustrare il passaggio dai casi minimali a quelli massimali di cognizione senza fare ricorso a rappresentazioni interne. Un primo passo per rispondere è notare che è possibile *conformarsi* ad una regola senza *seguirla esplicitamente*. Un pesce potrebbe mantenere il suo posto nel banco grazie ad un meccanismo che gli faccia seguire, ad una velocità determinata dalla distanza percepita, il pesce di fronte. Lo stesso meccanismo potrebbe anche sfruttare la distanza percepita rispetto ai pesci “a lato” per mantenere una rotta ottimale. Questo semplice meccanismo *sarebbe* sufficiente per far conformare il nostro pesce ad una norma in modo del tutto interattivo ed *ur-intenzionale*. Oltre questi meccanismi tipici della robotica situata (Brooks 1999, Clark 1997, §1.2) NOC permette, col suo terzo *step*, di fare ricorso ad un meccanismo di *conformazione sociale*. Come nel caso dei pesci, non è importante ora *quale* sia questo mec-

Malafouris (2013, cap. 5).

<sup>7</sup>Di nuovo, devo ringraziare i *referee* per avermi spinto a sviluppare questo punto.

canismo; quello che è rilevante è che la semplice responsabilità sociale che esso implica permette di *seguire norme culturali locali* senza doverle rappresentare, semplicemente agendo come gli *ingroup* o in modo coordinato ad essi. Questo è un comportamento regolato da una norma in un senso *più forte* di quello dei pesci, dal momento che la norma seguita dipende dalla cultura locale in cui l'agente è immerso; ma non sembra *di per sé* sufficiente per esibire il comportamento normativamente regolato caratteristico della manipolazione di una struttura esterna per fini cognitivi. Perché questo avvenga, la manipolazione deve avvenire in accordo con le *affordances* rilevanti dell'oggetto manipolato. Consideriamo un abaco: in teoria può avere molti usi (lo si può bruciare o usare come arma impropria); ma in pratica offre principalmente un'*affordance* rilevante: spostare gli anellini lungo la colonna. *Sfruttare* quest'*affordance* non è però sufficiente per *utilizzare* l'abaco: si può sfruttare l'*affordance* muovendo a caso gli anelli. Tuttavia, il meccanismo di conformazione sociale può spingere l'agente ad interagire in un modo specifico con l'*affordance*, selezionando, tra i diversi modi per *sfruttarla*, quello che corrisponde al suo *utilizzo* socialmente regolato. Questo è *sufficiente* per mostrare come un agente possa seguire una norma senza *rapresentarla internamente*. Tuttavia, è presupposto che la comunità dell'agente disponga delle norme rilevanti, ed in questo senso sembra solo spostare indietro il problema. È però possibile che una comunità possieda delle norme senza che queste debbano essere rappresentate. Perché ciò avvenga, sembra essere sufficiente che il modo peculiare di utilizzare un'*affordance* (ad esempio, il modo peculiare di interagire con l'abaco) sia scoperto e cominci ad essere replicato dagli agenti (grazie ai meccanismi *ur-intenzionali* di responsabilità sociale) fino a diventare la norma sociale tipicamente diffusa in un gruppo.

#### 4 Problemi aperti

Nel paragrafo precedente ho cercato di fornire un modo in cui l'*ur-intenzionalità* può essere sviluppata in modo compatibile con NOC e possa quindi contribuire alla comprensione dell'emergere naturale dei contenuti mentali. Rimangono però dei problemi aperti. Non solo alcuni problemi che ho indicato in precedenza non sono stati affrontati; ma anche l'introduzione delle abilità rilevanti all'interno dell'*ur-intenzionalità* apre problemi nuovi. Desidero ora farne una rassegna veloce, indicando in modo generale delle possibili soluzioni.

Un primo problema riguarda l'*informazione*. In effetti, il *contenuto* che i veicoli rilevanti si trovano ad avere, che ho cercato mostrare essere dipendente dalla loro manipolazione sulla base di abilità rilevanti, non dipende *sempre* da covarianze nomiche. Ad esempio, il significato di "5€" non sembra dipendere da una covarianza di questo tipo. Questo è vero, ma sembra possibile porvi rimedio modificando la nozione di informazione in un senso più vicino a quello della psi-

ecologia ecologica. In questo senso, l'informazione dipenderebbe dalle covarianze affidabili all'interno della nicchia ecologia (che comprende anche gli aspetti culturali, vedi Laland, Olding-Smee e Feldman 2000, Laland e O'Brien 2012) anche nel caso in cui quelle covarianze non siano nomiche (Chemero 2009, cap. 6). In questo modo, l'informazione generalizzata (covariazioni affidabili) avrebbe un primato epistemologico, nel senso che sarebbe quella più cospicuamente utilizzata dagli agenti cognitivi, anche se l'instaurazione di questo tipo generalizzato di covariazione dipende per la propria esistenza dalla presenza naturale di covarianze nomiche, cioè di informazione nel senso *non* ecologico.

Un altro problema rilevante da affrontare riguarda il grado di compatibilità tra tesi della manipolazione e l'idea propria di NOC per cui i contenuti hanno posto nelle pratiche ascrivite proprie di una forma di vita. Io non credo che, *in linea di principio*, le due tesi siano incompatibili e sarei tentato di sostenere che le pratiche di ascrizione siano un tipo particolare di manipolazione, che non riguarda *strutture ambientali*, ma pazienti ed agenti intenzionali incontrati nell'ambiente. Questo può essere vero, ma può mettere in ombra le (importanti) differenze che intercorrono tra la manipolazione normativamente regolata di una struttura ambientale e l'interazione sociale (anche solo limitata al gioco del dare e chiedere ragioni). Per evitare questa (potenzialmente nociva) tendenza ad appiattare sul modello della manipolazione di artefatti l'interazione sociale, può essere importante valorizzare il ruolo che l'intenzionalità condivisa e le pratiche sociali hanno nel mettere a disposizione gli artefatti culturali da manipolare, sia da un punto di vista ontogenetico che filogenetico, e l'importante effetto da dente d'arresto che forniscono, permettendo la creazione del *bootstrapping paradise* che sembra caratteristico della cognizione umana (Tomasello 2014).

Un'altra domanda rilevante è se l'*ur-intenzionalità*, così compresa, permetta davvero di ospitare tutti quegli sviluppi, filogenetici ed ontogenetici, che contraddistinguono la cognizione umana ed animale, nelle scale spaziali e temporali rilevanti (sulle quali vedi Pfifer e Bongard 2006, Maye e Engel 2015). Questa sfida è sia empirica che concettuale, e riguarda sia l'individuazione dei reali processi cognitivi nelle differenti scale spazio-temporali sia l'individuazione di un modo in cui quei processi potrebbero essere compresi nel *framework* dell'*ur-intenzionalità*. Rispondere a questa domanda equivale sostanzialmente a far passare la scienza cognitiva da una fase rivoluzionaria ad una paradigmatica (Chemero 2009, cap. 1) e non è detto che sia qualcosa di possibile (almeno a breve termine). Quello che sembra essere in ogni caso sicuro è che la scienza della mente *matura* dovrà dare uno spazio considerevole al ruolo dell'azione e della manipolazione ambientale nella cognizione.

Un problema connesso è se il *framework* dell'*ur-intenzionalità* possa rendere ragione del rapporto complessivo tra agente cognitivo e mondo in tutta la sua ricchezza di sfaccettature. Nel *paper* mi sono concentrato, come del resto Hut-

to, Myin e Satne, sugli aspetti cognitivi dell'*ur-intenzionalità*; ma non credo che l'ampliamento del *framework ur-intenzionale* agli aspetti emotivi e sociali della mente presenti problemi in linea di principio. In particolare, ci sono numerose proposte *action oriented* per comprendere questi fenomeni (vedi Caruana e Borghi (2016) per una sintetica introduzione) e, sebbene ne manchi una sviluppata nei termini propri di REC, non sembrano esserci ostacoli a priori perché ciò avvenga. Nella misura in cui è possibile, ad esempio, comprendere l'aspetto *situato* ed *esteso* della mente emotiva (Colombetti e Krueger 2015) sembra anche possibile comprendere l'intenzionalità propria dei fenomeni emotivi *prima* in termini *ur-intenzionali* e quindi svilupparla facendo ricorso alla tesi della manipolazione.

L'ultimo problema, che trovo più pressante, lo presentano i concetti astratti, come ad esempio il concetto di "democrazia". Un problema in questi casi è che *non c'è* un *singolo* pattern motorio o serie di pattern motori rilevanti associabili a "democrazia", né una singola abilità rilevante. Ma le nostre mappe motorie possano mappare in una relazione che non è necessariamente 1:1 (Pulvemuller 2015) ed esattamente com'è possibile che nell'ambiente *ecologico* sia presente dell'informazione di "alto livello", così pure è possibile che le nostre mappe motorie siano di *alto livello*, e non siano direttamente legate alla struttura biomeccanica dei movimenti (Umiltà e al 2008). Certo parlare di mappe motorie è fare ricorso alla *representational stance*, ma nulla ci vieta di farlo; ed è anche possibile (anche se difficile) comprendere le mappe motorie in termini non rappresentazionali (Chemero 2009, cap. 4). Questo sembra essere un indizio per sostenere che il rapporto *motorio, ur-intenzionale* in un senso chiaro, sia in grado di esibire dei caratteri di flessibilità ed astrattezza che *sembrano* permettere di alloggiare al suo interno certi gradi di astrattezza. Di certo è vero che solo un rapporto motorio non è sufficiente per avere a che fare con i concetti astratti, ma è anche vero che i concetti astratti *non* esistono in natura, ma solo all'interno di forme di vita altamente specifiche e quindi solo in presenza di una serie rilevante di attrezzature ambientali (nel senso più ampio possibile) manipolabili. In questo senso, un termine "altamente astratto" come *democrazia potrebbe* essere trattato come il più concreto termine "coltivare": chiaramente non esiste un'azione che è il coltivare; ma il coltivare è fatto da diverse azione socio-culturalmente e materialmente unite – e lo stesso *potrebbe* valere per "democrazia".

## Riferimenti bibliografici

- Adams, Fred e Kenneth Aizawa (2001). "The Bounds of Cognition". In: *Philosophical Psychology* 14.1, pp. 43–64.
- (2010). "Defending the Bounds of Cognition". In: *The Extended Mind*. A cura di Richard Menary. Cambridge MA: The MIT press.
- Becchio, Cristina e Et al (2012). "Grasping intentions: from thought experiments to empirical evidence". In: *Frontiers in Human Neuroscience* 6.117.
- Beer, Randall (2003). "The dynamics of active categorical perception in an evolved model agent". In: *Adaptive Behavior* 11.4, pp. 209–243.
- Brooks, Rodney (1999). *Cambrian Intelligence*. Cambridge MA: The MIT Press.
- Caruana, Fausto e Anna Borghi (2016). *Il Cervello in Azione*. Bologna: Il Mulino.
- Chemero, Anthony (2009). *Radical Embodied Cognitive Science*. Cambridge MA: The MIT Press.
- Clark, Andy (1997). *Being There: putting Brain, Body and World Together Again*. Cambridge MA: The MIT Press.
- (2005). "Beyond the Flesh: some lessons from a mole cricket". In: *Artificial Life* 11.1, pp. 233–244.
- (2016). *Surfing Uncertainty*. Oxford: Oxford University Press.
- Clark, Andy e David Chalmers (1998). "The Extended Mind". In: *Supersizing the mind*. A cura di Andy Clark. Oxford 2008: Oxford University Press.
- Clark, Andy e Josefa Toribio (1994). "Doing Without Representing?" In: *Synthese* 101.3, pp. 401–431.
- Colombetti, Giovanna e Joel Krueger (2015). "Scaffoldings of the Affective Mind". In: *Philosophical Psychology* 28.8, pp. 1157–1176.
- De Oliveira, G.S. e A. Chemero (2015). "Against Smallism and Localism". In: *Studies in logic, grammar and rhetoric* 41.54, pp. 9–23.
- Dennett, Daniel (1987). *The Intentional Stance*. Cambridge MA: The MIT Press.
- (2011). "Shall We Tango? No, but Thanks for Asking". In: *Journal of Consciousness Study* 18.5, pp. 23–34.
- Gallagher, Shaun (2017). *Enactivist Interventions*. Oxford: Oxford University Press.
- Haugeland, John (1990). "The Intentionality All-Stars". In: *Having Thought*. A cura di John Haugeland. Cambridge MA: Harvard University Press, Chapter 3.
- Hurley, Susan (2010). "The Varieties of Externalism". In: *The Extended Mind*. A cura di Richard Menary. Cambridge MA: MIT Press.

- Hutto, Daniel (2006). “Unprincipled Engagements: Emotional experience, expression and response”. In: *Radical Enactivism*. A cura di Richard Menary. Amsterdam: John Benjamins Publishing Company.
- Hutto, Daniel e Erick Myin (2013). *Radicalizing Enactivism*. Cambridge MA: The MIT Press.
- (2017). *Evolving Enactivism*. Cambridge MA: The MIT Press.
- Hutto, Daniel e Glenda Satne (2015a). “Introduction: Searching for the Natural Origins of Content”. In: *Philosophia* 43.3, pp. 505–519.
- (2015b). “The natural origins of content”. In: *Philosophia* 43.3, pp. 521–536.
- (2017). “Continuity Skepticism in doubt: a radically enactive take”. In: *Embodiment Enaction and Culture*. A cura di T. Durt e C.T. Fuchs. Cambridge MA: The MIT Press, pp. 107–127.
- (2018). “Naturalism in the Goldilocks Zone: Wittgenstein’s Delicate Balancing Act”. In: *Wittgenstein and Naturalism*. A cura di K. Cahill e T. Raleigh. London: Routledge, Chapter 3.
- Kim, Jaegwon (2011). *Philosophy of Mind (Third edition)*. Boulder Colorado: Westview Press.
- Kiverstein, Julian e Erick Rietveld (2014). “A Rich Landscape of Affordances”. In: *Ecological Psychology* 26.4, pp. 325–352.
- (2015). “The primacy of Skilled intentionality: on Hutto and Satne’s natural origins of content”. In: *Philosophia* 43.3, 701–721.
- Laland, Kevin, John Olding-Smee e Marcus W. Feldman (2000). “Niche construction, biological evolution and cultural change”. In: *Behavioral and brain sciences* 23.1, pp. 146–175.
- Laland, K.N. e M.J. O’Brien (2012). “Cultural niche construction: an introduction”. In: *Biological Theory* 6.3, pp. 22–33.
- Malafouris, Lambros (2013). *How Things Shape The Mind*. Cambridge MA: The MIT Press.
- Mason, Lucia (2006). *Psicologia dell’apprendimento e dell’istruzione*. Bologna: Il Mulino.
- Maye, Alexander e Andreas Engel (2015). “Extending Sensorimotor Contingencies to Cognition”. In: *The Pragmatic Turn*. A cura di A. Engel, K. Friston e D. Kragic. Cambridge MA: The MIT Press.
- Menary, Richard (2006). “What is Radical Enactivism?” In: *Radical Enactivism*. A cura di Richard Menary. Amsterdam: John Benjamins Publishing Company.

- Menary, Richard (2007). *Cognitive Integration, mind and cognition unbound*. Basingstoke: Palgrave MacMillan.
- (2010). “Cognitive Integration and the Extended Mind”. In: *The Extended Mind*. A cura di Richard Menary. Cambridge MA: The MIT Press.
- (2015). “Pragmatism and the Pragmatic Turn in Cognitive Science”. In: *The Pragmatic Turn*. A cura di A. Engel, K. Friston e D. Kragic. Cambridge MA: The MIT Press.
- Millikan, Ruth (1984). *Language, Thought and Other Biological Categories*. Cambridge MA: The MIT Press.
- (2000). “Reading Mother Nature’s Mind”. In: *Dennett’s Philosophy, a comprehensive assessment*. A cura di D. Ross, A. Brook e D. Thompson. Cambridge MA: The MIT Press.
- Noe, Alva (2004). *Action in Perception*. Cambridge MA: The MIT Press.
- Pfifer, Rolf e Josh Bongard (2006). *How The Body Shapes The Way We Think*. Cambridge MA: The MIT Press.
- Piattelli Palmarini, Massimo (2008). *Le scienze cognitive classiche: un panorama*. Torino: Einaudi.
- Pulvemuller, Friedelman (2015). “Language, Action, Interaction”. In: *The Pragmatic Turn*. A cura di A. Engel, K. Friston e D. Kragic. Cambridge MA: The MIT Press.
- Putnam, Hilary (1975). “Il significato di significato”. In: *Mente, Linguaggio e Realtà*. A cura di Hilary Putnam. Milano: Adelphi.
- Rowlands, Mark (2006). *Body Language*. Cambridge MA: The MIT Press.
- (2010). *The New Sciences of The Mind*. Cambridge MA: The MIT Press.
- (2018). *Embodied Consciousness*. In pubblicazione. URL: [https://www.academia.edu/35551655/Embodied\\_Consciousness](https://www.academia.edu/35551655/Embodied_Consciousness).
- Searle, John (1980). “Minds Brains and Programs”. In: *Behavioral and Brain Sciences* 3.3, pp. 417–457.
- (1983). *Intentionality, an essay in the philosophy of mind*. Cambridge: Cambridge University Press.
- (1984). “Intentionality and it’s place in nature”. In: *Synthese* 61.1, pp. 3–16.
- (1992). *La riscoperta della mente*. Torino: Bollati Boringhieri.
- Tomasello, Michael (2014). *Unicamente umano*. Bologna: Il Mulino.
- Umiltà, Maria e Et al (2008). “When Pliers Become Fingers in the Monkey Motor System”. In: *proceedings of the National Academy of Sciences* 105.6, pp. 2209–2213.

Varela, F., E. Thompson e E. Rosch (1991). *The Embodied Mind*. Cambridge MA: The MIT Press.





# Le allucinazioni sono immagini mentali?<sup>1</sup>

*Giorgio Mazzullo*

**Abstract.** Recentemente, una nuova proposta sulla natura delle allucinazioni ha scosso il consolidato dibattito in filosofia della percezione tra teorie disgiuntiviste e teorie del genere comune. Secondo alcuni autori (Nanay 2015, 2016; Allen 2015; Pagondiotis 2013), dovremmo considerare le allucinazioni come appartenenti allo stesso genere mentale delle immagini mentali. In questo articolo valuto la solidità di questa proposta e argomento in favore della tesi che essa, per come è stata delineata finora, non è accettabile. In ogni caso, suggerisco che le allucinazioni potrebbero essere eventi che appartengono a diversi generi mentali. In particolare, potrebbero esistere allucinazioni derivanti dal sistema delle esperienze percettive e allucinazioni derivanti dal sistema delle immagini mentali.

**Keywords.** Allucinazioni, Immagini mentali, Esperienza percettiva, Disgiuntivismo, Genere comune, Percezione.

---

<sup>1</sup>Sono grato a due referee di questa rivista per i preziosi commenti che hanno contribuito sensibilmente a migliorare molte parti del testo. Inoltre, vorrei ringraziare per il sostegno e i confronti sull'oggetto di questo articolo Simone Gozzano, Giorgio Lando, Alessio Santelli e Marco Segala.

## Introduzione

La riflessione sulla natura delle allucinazioni ha un ruolo fondamentale in filosofia della percezione e ha per di più il potenziale di modificare radicalmente il dibattito in filosofia della mente e in epistemologia. Negli ultimi venti anni il dibattito sulla natura delle allucinazioni è stato monopolizzato da due teorie filosofiche della percezione: il disgiuntivismo e la teoria del genere comune. Il nucleo del disaccordo tra queste due teorie ha a che fare con una diversa concezione del rapporto tra allucinazioni e esperienze veridiche<sup>2</sup>.

I filosofi della percezione che riflettono sulle allucinazioni non prendono generalmente in considerazione episodi allucinatori come effettivamente occorrono soprattutto in relazione a noti quadri patologici<sup>3</sup>. Oggetto del contendere sono, piuttosto, delle allucinazioni puramente ipotetiche: le cosiddette allucinazioni dei filosofi o allucinazioni causalmente corrispondenti (*causally matching hallucinations*). Le allucinazioni causalmente corrispondenti sono allucinazioni causate dall'attivazione del medesimo stato cerebrale di una corrispondente esperienza veridica.

I sostenitori del genere comune (*common kind*) ritengono che le allucinazioni e le esperienze veridiche appartengano allo stesso genere mentale. Spesso si argomenta in favore di questa tesi proponendo un Argomento Causale (Robinson 1994) basato sul principio stessa causa prossimale-stesso effetto. La stessa causa prossimale, una certa configurazione neurale, deve produrre lo stesso effetto. Pertanto, almeno esperienze veridiche e allucinazioni causalmente corrispondenti appartengono allo stesso genere mentale. Sulla scorta di considerazioni riguardo la fenomenologia e/o il conto degli stati veridici e allucinatori, i sostenitori del genere comune ritengono, generalmente, che ciò che rende una certa esperienza veridica o allucinatoria è, in un certo senso, indipendente dalla natura dello stato stesso. Le teorie intenzionaliste della percezione (Crane 2006; Searle 2015) sono, solitamente, teorie del genere comune.

Secondo le teorie disgiuntiviste (Fish 2009; Hinton 1967; Martin 2004, 2006; Snowdon 2005) le esperienze veridiche e le esperienze allucinatorie sono, invece, radicalmente differenti. I disgiuntivisti rifiutano il principio stessa causa prossimale - stesso effetto, quando ci sono condizioni non causali necessarie per l'occorrenza di un certo genere di effetto. Nel caso delle esperienze veridiche la

<sup>2</sup> Nell'articolo utilizzerò una terminologia ormai consolidata in filosofia della percezione. Userò "esperienza percettiva" per riferirmi al genere di stato mentale che occorre in esperienze veridiche, illusioni e allucinazioni. Userò invece il verbo "percepire" in senso fattivo, per riferirmi a esperienze veridiche.

<sup>3</sup> Questa strategia è stata recentemente oggetto di critica. Allen (2015) ritiene che sia metodologicamente discutibile il fatto che i filosofi della percezione, nel momento in cui si propongono di decidere quale sia la migliore teoria sulle allucinazioni, diano maggior peso a intuizioni su remote situazioni controfattuali piuttosto che a effettive occorrenze allucinatorie. Concordo con Allen su questo punto, come emergerà in modo chiaro in §5.

condizione ulteriore per l'occorrenza di quello stato è che l'oggetto percepito sia effettivamente presente. Lo stato esperienziale veridico è essenzialmente relazionale, in quanto almeno in parte determinato dalla natura di oggetti, proprietà e relazioni indipendenti dalla mente (Campbell 2002, p. 118). Le esperienze veridiche, a differenza delle allucinazioni, sono una "rivelazione" del mondo circostante.

All'inizio del ventesimo secolo il dibattito in filosofia della percezione vedeva contrapposti i teorici del realismo diretto (Austin 1962) e i teorici dei dati sensoriali (Russell 1912). Il realismo diretto è stato in grado di mostrare come la teoria dei dati sensoriali avesse delle conseguenze ontologiche difficilmente accettabili. La teoria dei dati sensoriali è stata, così, quasi unanimemente accantonata<sup>4</sup>. L'odierno dibattito tra teorie del genere comune e teorie disgiuntiviste non ha visto finora il chiaro prevalere di una posizione sull'altra.

È emersa, piuttosto, la tendenza a proporre teorie ibride (Byrne e Heater 2008), cercando di prendere in considerazione gli aspetti più convincenti delle due proposte teoriche<sup>5</sup>. In particolare, si è cercato di costruire delle teorie intenzionaliste che fossero in grado di dar conto dell'intuizione disgiuntivista a proposito della relazione con il mondo che si verifica durante una percezione. L'impressione generale, tuttavia, è che il dibattito sia ancorato a due visioni mutualmente esclusive sulla natura dell'esperienza percettiva.

Se il disgiuntivismo non è stato in grado di articolare una proposta convincente sulla natura dell'esperienza allucinatoria, le teorie del genere comune non sono riuscite a trovare un accordo su quale debba essere considerato il nucleo della comunanza di genere tra allucinazioni e esperienze veridiche.

Recentemente, tuttavia, probabilmente come tentativo di rispondere a questa tensione, alcuni filosofi della percezione (Allen 2015; Nanay 2015, 2016; Pagoniotis 2013) hanno proposto di considerare le allucinazioni come appartenenti allo stesso genere delle immagini mentali (*mental imagery*).

In questo articolo valuterò la solidità di questa proposta. A tal fine, dopo avere tratteggiato una brevissima ricostruzione dell'analisi filosofica delle immagini mentali e una classificazione delle teorie adottabili nei confronti della relazione tra immagini mentali ed esperienze percettive (§1), analizzerò due teorie della percezione che propongono di considerare le allucinazioni come appartenenti allo stesso genere delle immagini mentali. Cercherò di mostrare come entrambe le teorie non siano in grado di rendere conto in modo soddisfacente di alcuni fenomeni e debbano rispondere a una serie di convincenti controargomenti (§2)

<sup>4</sup>In particolare, i sostenitori della teoria dei dati sensoriali sono stati accusati di postulare l'esistenza di oggetti mentali dal dubbio statuto ontologico. Robinson (1994) è uno dei pochi filosofi rimasti che ritiene che la teoria dei dati sensoriali sia la migliore teoria filosofica della percezione

<sup>5</sup>Si tratta di quelle teorie che secondo Byrne e Heater (2009) possono essere definite appartenenti a un Disgiuntivismo Moderato. Queste teorie accettano che allucinazioni e percezioni siano significativamente differenti in qualche aspetto mentale, pur avendo qualche caratteristica comune.

e §3). Sosterrò, in ogni caso, che l'ipotesi che le allucinazioni appartengano allo stesso genere mentale delle immagini mentali non debba essere rigettata, ma ulteriormente indagata (§4). In questo senso, avanderò l'ipotesi che le allucinazioni siano eventi che possono occorrere in diversi generi mentali (§5). In particolare, potrebbero esistere allucinazioni derivanti dal sistema dell'esperienza percettiva e allucinazioni derivanti dal sistema delle immagini mentali.

## 1 Le immagini mentali e l'esperienza percettiva

Recentemente, alcuni filosofi hanno avanzato una nuova proposta per sbloccare la situazione di impasse in cui sembra trovarsi il dibattito sulla natura dell'esperienza percettiva. L'ipotesi formulata da questi autori (Allen 2015; Nanay 2015, 2016; Pagondiotis 2013), che propongono teorie riconducibili sia al disgiuntivismo sia all'intenzionalismo, è la seguente:

T: Le allucinazioni appartengono allo stesso genere mentale delle immagini mentali.

In questo modo, in maniera più o meno consapevole, questi filosofi hanno posto le basi per l'apertura di un nuovo filone di ricerca interno alla filosofia della percezione. Questa nuova linea di ricerca assume, più o meno dichiaratamente, che una migliore comprensione del fenomeno delle immagini mentali o episodi di immaginazione sensoriale, nella loro relazione con l'esperienza percettiva, possa giocare un ruolo cruciale nel risolvere il dibattito tra teorie disgiuntiviste della percezione e teorie del genere comune.

La tesi che le immagini mentali possano aiutarci a comprendere la natura delle esperienze percettive poggia apparentemente su una solida tradizione teorica. Le immagini mentali sono spesso considerate degli stati quasi-percettivi (Richardson 1969) con rilevanti similarità fenomeniche e neurofisiologiche con l'esperienza percettiva.

Per ciò che riguarda le similarità fenomeniche, il *locus classicus*, nonostante Hopkins (2012) l'abbia recentemente messo in discussione, è l'esperimento<sup>6</sup> di Perky (1910) in cui si mostrerebbe che percezioni e immagini mentali possono risultare fenomenicamente indistinguibili.

<sup>6</sup> L'esperimento funzionava nel modo seguente: si chiedeva ai soggetti di fissare un punto su una parete bianca e al contempo di formare l'immagine mentale di un oggetto colorato come una banana o un pomodoro. Lo sperimentatore, posta la stanza in una condizione di semioscurità, proiettava l'immagine dell'oggetto immaginato sulla parete. Una volta proiettata, per esempio, una banana i soggetti riferivano di aver soltanto immaginato la banana. L'immagine sulla parete era tale che un soggetto non coinvolto nell'esperimento avrebbe certamente notato la parete illuminata da quell'immagine. Secondo Perky i soggetti dell'esperimento scambiavano l'esperienza percettiva dell'immagine di una banana con la visualizzazione (l'immagine mentale visiva) della banana e ciò mostrava in modo del tutto evidente come vedere e visualizzare sono episodi fenomenicamente simili.

Dal punto di vista neurofisiologico, inoltre, ha acquisito sempre maggiore credibilità l'idea che la percezione e l'immaginazione sensoriale condividano, almeno in parte, il medesimo substrato neurale. È riscontrabile, infatti, una quasi completa sovrapposizione tra le regioni cerebrali coinvolte nell'immaginazione sensoriale e quelle coinvolte nell'esperienza percettiva (Kosslyn, Thompson e Ganis 2006; Page, Duhamel e Crognale 2011).

L'interesse per le immagini mentali, anche in riferimento alla loro relazione con l'esperienza percettiva, non è certamente nuovo in filosofia. Sono riscontrabili almeno due filoni all'interno della riflessione intorno alla natura delle immagini mentali.

Il primo, più recente, ha a che fare con un vivace dibattito sorto nell'ambito della psicologia cognitivista, che ha coinvolto ben presto una nutrita cerchia di filosofi. Quello che è stato definito "dibattito sulle immagini mentali", sviluppatosi soprattutto negli anni Ottanta, vedeva contrapposte due ipotesi sulla natura delle immagini mentali. Da un lato, l'ipotesi pittorialista sostenuta, tra gli altri, da Stephen Kosslyn (1980). Secondo i pittorialisti le immagini mentali sono entità pittoriche irriducibili a strutture di livello inferiore. Dall'altro, l'ipotesi proposizionalista, che ha visto Zenon Pylyshyn (1973) come principale sostenitore, secondo cui le immagini mentali hanno natura eminentemente proposizionale. L'aspetto figurale delle immagini, in questa seconda prospettiva, è generato da strutture proposizionali della mente ed è soltanto la superficie esterna del fenomeno.

Il secondo filone è invece quello più prettamente filosofico che ha visto i filosofi interrogarsi, nel corso dei secoli, sulla natura delle immagini mentali e sulla relazione tra queste e l'esperienza percettiva. Il dibattito classico assumeva, generalmente, una precisa caratterizzazione metafisica dell'esperienza percettiva, accettando i presupposti delle teorie del genere comune. La discussione, pertanto, era finalizzata a comprendere se tra le immagini mentali e le esperienze percettive ci fosse soltanto una differenza di grado, oppure una differenza di genere. È nota, in questo senso, la tesi di Hume (1739) secondo cui tra esperienze percettive e immagini mentali c'è soltanto una differenza nel grado di nitidezza e non una differenza di genere.

L'attuale dibattito in filosofia della percezione, tuttavia, comporta una proliferazione di teorie adottabili nei confronti di questo tema. I disgiuntivisti, infatti, rifiutano che le percezioni (intese nell'uso fattivo) e le allucinazioni appartengano a un genere mentale comune. Nell'attualizzare il dibattito si ha, pertanto, la necessità di introdurre almeno un terzo oggetto di indagine: le allucinazioni. Le illusioni, che potrebbero essere prese in considerazione come quarto oggetto di indagine, vengono oggi considerate, specialmente in ambito disgiuntivista, prive di uno statuto autonomo<sup>7</sup>. Esse sono assimilate in alcune teorie alle

<sup>7</sup>Martin (2004, 2006) e McDowell (2013) considerano le illusioni analoghe alle allucinazioni. Al-

esperienze veridiche, in altre alle allucinazioni.

Per comprendere su quali linee si muova l'attuale discussione intorno alle immagini mentali propongo una classificazione delle posizioni possibili che non si rifaccia necessariamente alla letteratura esistente, ma che distingua le possibili combinazioni in merito alla relazione tra percezioni, immagini mentali e allucinazioni. Ciò consentirà di chiarire, date le attuali premesse, le cinque possibili teorie adottabili nei confronti di questo tema. Nonostante sia almeno in parte implicitamente assunta dagli autori che si occupano di immagini mentali, questa classificazione non è stata esplicitata.

**T1:** Allucinazione e immaginazione sensoriale sono lo stesso genere mentale, mentre la percezione è un genere differente (Allen 2015; Pagondiotis 2013).

**T2:** Percezione, allucinazione e immaginazione sensoriale sono tre diversi generi mentali (Martin 2002).

**T3:** Percezione e immaginazione sensoriale sono lo stesso genere, ma l'allucinazione è un genere differente.

**T4:** Percezione, immaginazione sensoriale e allucinazione sono un unico genere mentale (Nanay 2015).

**T5:** Percezione e allucinazione sono un unico genere mentale, mentre l'immaginazione sensoriale è un genere differente.

T1 e T2 sono due teorie disgiuntiviste; entrambe affermano che la percezione è un genere mentale radicalmente differente rispetto agli altri. T4 e T5 sono invece due teorie del genere comune, dato che affermano che percezione e allucinazione appartengono al medesimo genere mentale.

T3 non è invece rappresentata nel dibattito. Questa teoria dovrebbe motivare non solo la problematica identificazione tra immagini mentali e percezione, ma dovrebbe inoltre spiegare perché l'allucinazione appartenga a un genere mentale differente. In effetti, una volta eliminato il confine tra immagini mentali e percezione sembra arbitrario non estendere l'identità di genere anche alle allucinazioni.

Emerge chiaramente da questa classificazione come soltanto le teorie T1 e T4 propongano una chiara identificazione di genere tra allucinazioni e immagini mentali. Nei paragrafi seguenti (§2 e §3) cercherò di valutare la solidità di queste due proposte teoriche.

tri autori, tuttavia, ritengono le illusioni analoghe alle percezioni (Snowdon 2005). La questione è molto complessa. Rimando a Calabi (2009) per un ottimo inquadramento della problematica, a Kalderon (2011) per una teoria delle illusioni che si discosta dalla concezione filosofica tradizionale e a Macpherson e Batty (2016) per un tentativo di ridefinire le nozioni tradizionali di illusione e allucinazione.

## 2 T1: Il trattamento disgiuntivista delle allucinazioni come immagini mentali

Secondo T1 le allucinazioni appartengono allo stesso genere mentale degli episodi di immaginazione sensoriale e questo genere mentale comune deve essere distinto dalla percezione. Si tratta, come già sottolineato, di una teoria disgiuntivista, poiché il nucleo della teoria afferma che esperienze veridiche e allucinazioni appartengono a generi mentali differenti.

Pagondiotis (2013, p. 377) e Allen (2015, p. 288) hanno recentemente proposto questa teoria sostenendo che l'allucinazione è un genere "corrotto" dell'immaginazione sensoriale in cui il soggetto non ha controllo volontario diretto<sup>8</sup>.

Il trattamento delle allucinazioni come immagini mentali sembra, almeno prima facie, ritagliato perfettamente per una strategia disgiuntivista; permette infatti di fornire una caratterizzazione positiva delle allucinazioni e di mostrarne la differenza rispetto alle percezioni.

Tuttavia, ipotizzare una caratterizzazione positiva delle allucinazioni, sostenendo che esse siano episodi di immaginazione sensoriale, sfida lo *screening off problem* (Martin 2006). Secondo Martin, a causa dello *screening off problem*, qualsiasi caratterizzazione positiva dello stato allucinatorio deve essere accettata dal disgiuntivista anche per l'esperienza veridica. Questo metterebbe fuori gioco il ruolo dell'effettivo oggetto percepito dallo spiegare lo stato mentale, una conseguenza che, secondo Martin, il disgiuntivista non può accettare.

Per comprendere la forza dello *screening off problem* è necessario fare un passo indietro. Abbiamo visto come il disgiuntivista rifiuti le conclusioni dell'Argomento Causale e il principio stessa causa prossimale-stesso effetto. Tuttavia, come ha sottolineato Martin (2006, p. 368), c'è un'altra minaccia per il disgiuntivista: il *Reverse Causal Argument*. Il *Reverse Causal Argument* concede al disgiuntivista un principio causale modificato che ammette la necessità di condizioni non causali, affinché si dia un certo genere di stato mentale:

*Principio Causale Modificato*: un evento  $e_1$  è dello stesso genere G di un evento  $e_2$  se l'evento  $e_1$  è prodotto dalla stessa causa prossimale di  $e_2$  in circostanze che non differiscono in nessuna condizione non causale necessaria per l'occorrenza dell'evento del genere G<sup>9</sup>.

Come abbiamo visto, ci sono condizioni non causali necessarie perché si dia un'esperienza veridica che sono assenti nel caso allucinatorio (la presenza dell'oggetto percepito). Cosa succede, però, se analizziamo il caso allucinatorio attraverso il Principio Causale Modificato?

<sup>8</sup> Più precisamente, Pagondiotis (2013, p. 377) sostiene che le allucinazioni siano degli episodi di immaginazione sensoriale in cui viene meno il *sense of agency*.

<sup>9</sup> Questa è la formulazione di Nudds (2009).

Consideriamo l'evento allucinatorio *a* causalmente corrispondente all'evento veridico *v*. Se accettiamo il Principio Causale Modificato dei disgiuntivisti arriviamo a concludere che per avere il genere dell'evento *a* allucinatorio c'è bisogno di un evento prodotto dalla stessa causa prossimale (una data configurazione neurale), che non differisce in nessuna condizione non causale necessaria affinché sia dia un evento di quel genere. Essendo l'evento *a* un evento allucinatorio non c'è nessuna condizione non causale necessaria affinché si dia quell'evento che non sia presente nell'evento *v* di percezione veridica. Questo vuol dire che affinché si dia un evento *v* di percezione veridica è necessario che si dia almeno un evento del genere *a* allucinatorio. Un evento che necessita soltanto dell'attivazione di un certo pattern neurale, una medesima causa prossimale.

Il *Reverse Causal Argument* afferma, in sostanza, che il genere di stato che contraddistingue l'allucinazione deve essere in qualche modo una componente del genere dello stato veridico. Queste considerazioni mettono le basi per il cosiddetto problema dello *screening off*. Ammettere una stessa causa prossimale per lo stato allucinatorio e veridico, infatti, sembra rendere la presenza dell'oggetto percepito del tutto ridondante: il genere dell'allucinazione metterebbe fuori gioco il ruolo esplicativo dell'esperienza veridica nel dar conto delle caratteristiche fenomeniche, comportamentali e dossastiche dell'esperienza percettiva.

In effetti, Martin sostiene che a questo punto il disgiuntivista debba dare una caratterizzazione epistemica puramente negativa dello stato allucinatorio, sostenendo che tutto ciò che si può dire di esso è che è semplicemente fenomenicamente indistinguibile dallo stato veridico nel caso di esperienze causalmente corrispondenti. Secondo Martin, infatti, l'esperienza veridica ha ancora un ruolo esplicativo cruciale da giocare, dato che la nozione di indistinguibilità a cui si aggancia lo stato allucinatorio è basata nondimeno sulla percezione veridica. Ciò renderebbe l'evento allucinatorio parassitario rispetto a quello veridico per quanto riguarda la spiegazione delle caratteristiche psicologiche sopra menzionate. Il disgiuntivista, secondo Martin, non dovrebbe dare alcun tipo di caratterizzazione positiva dell'allucinazione, pena l'impossibilità di render ragione del ruolo esplicativo della percezione veridica a causa dello *screening off problem*.

Una caratterizzazione positiva dello stato allucinatorio come quella proposta da T1 andrebbe incontro al problema dello *screening off*. Infatti, una volta posta l'identità di genere tra allucinazioni e immagini mentali, per il *Reverse Causal Argument* anche lo stato veridico apparterebbe al genere delle immagini mentali e delle allucinazioni.

Si verificherebbe in questo modo il collasso di T1 su T4. L'ipotesi che le immagini mentali e le allucinazioni appartengano al medesimo genere mentale e che la percezione sia un genere mentale differente sembra dovere rispondere, in primo luogo, al problema dello *screening off*.

### 3 T4: L'estensione del genere comune alle immagini mentali

Secondo T4, percezione, immaginazione sensoriale e allucinazione sono un unico genere mentale. Si tratta, pertanto, di un'estensione della tesi del genere comune alle immagini mentali. T4, sostenendo che esista un unico genere, sembra in grado di rendere conto in modo soddisfacente delle similarità fenomeniche e neurofisiologiche tra percezione e immagini mentali già menzionate.

Nanay (2015, 2016) ha recentemente proposto una variante di T4. Secondo Nanay, la similarità fenomenica riscontrabile tra immagini mentali ed esperienze percettive si basa sulla similarità di contenuto. Il contenuto degli stati percettivi e dell'immaginazione sensoriale è costituito dalle proprietà che sono attribuite dal nostro sistema percettivo alla scena percepita o immaginata. È un contenuto non proposizionale e spazialmente organizzato, sulla scorta del contenuto scenario (*scenario content*) di Peacocke (1992).

L'unica differenza, secondo Nanay, tra un episodio di percezione e un'immagine mentale è riscontrabile quando si cerca di mettere a fuoco la scena percepita o immaginata mediante l'attenzione. Nel caso di una percezione la fonte da cui si riescono a trarre proprietà maggiormente determinate è il mondo. Nel caso delle immagini mentali, e quindi delle allucinazioni, l'extra-determinatezza proviene dalla memoria, dalle credenze e dalle aspettative. La proposta di Nanay deve rispondere ad almeno due obiezioni<sup>10</sup>.

In primo luogo, l'immaginazione sensoriale è tradizionalmente considerata un atto volontario, mentre si è generalmente concordi nel ritenere che le allucinazioni e le percezioni non lo siano. A questa obiezione, che deriva da una concezione filosofica generale delle immagini mentali, Nanay ha, tuttavia, ribattuto in modo molto convincente. In primo luogo, egli ha sostenuto (Nanay 2017a,b) che i *flashback*, esperienze che presentano, totalmente o in parte, episodi del passato, sono da considerare casi paradigmatici in cui le immagini mentali si presentano in modo involontario. Inoltre, alcuni studi (Taylor, MacKay e Miguel 2014) hanno stimato che circa il 98% della popolazione occidentale abbia sperimentato in prima persona i cosiddetti *earworms*, ritornelli musicali che vengono

<sup>10</sup> Un ulteriore problema per T4 potrebbe giungere da una serie di studi che sfidano la tesi dell'esistenza di un comune substrato neurale tra percezione e immagini mentali. Degli esperimenti su pazienti che hanno subito danni cerebrali sembrano suggerire che la compromissione dell'attività percettiva non implichi quella riguardante la produzione di immagini mentali, e viceversa. L'ipotesi che è stata avanzata per rendere conto di questo fenomeno è che le immagini mentali e le esperienze percettive abbiano, almeno in parte, distinti correlati neurali. Il dibattito, interno alle neuroscienze cognitive, sul modo in cui vadano interpretati questi studi nel quadro generale dell'ipotesi di una equivalenza funzionale e anatomica tra percezione e immagini mentali è tuttora in corso. Rimando a Bartolomeo (2008) e Bértolo (2005) per una revisione critica del dibattito. Sulla questione del comune substrato neurale di percezione e immagini mentali si tornerà nel §4.

esperiti in modo del tutto involontario. Il fatto che gli *earworms* siano episodi di immagini mentali sembra corroborato dal fatto che la lettura scientifica si riferisca a essi come episodi di *Involuntary Musical Imagery* (INMI). I controesempi di Nanay sembrano mostrare come le immagini mentali possano essere involontarie.

Come secondo elemento c'è da menzionare il fatto che si ritiene che l'esperienza percettiva sia dotata di un senso di realtà o carattere presentazionale (Farakas 2013; Pagondiotis 2013). A differenza di ciò che avviene per atteggiamenti intenzionali come le credenze o i desideri, infatti, durante l'esperienza percettiva gli oggetti appaiono come effettivamente presenti. Tuttavia, quando si immagina un oggetto, sembra non essere compreso questo carattere presentazionale.

Nanay (2017a,b) ritiene di poter rispondere anche a questa obiezione. La soluzione proposta, tuttavia, sembra molto meno convincente rispetto alla precedente. Secondo Nanay i sogni lucidi (*lucid dreams*), in cui il sognatore ha consapevolezza di essere in un sogno, sono degli episodi di immagini mentali che godono di un senso di realtà o carattere presentazionale.

Ritengo che questa sia una posizione abbastanza controversa. In primo luogo, alcuni filosofi rifiuterebbero la tesi che i sogni siano immagini mentali<sup>11</sup>. Inoltre, anche qualora accettassimo questa ipotesi, non risulta così ovvio il fatto che i sogni lucidi siano accompagnati da un carattere presentazionale. Infatti, non è escludibile a priori che il sognatore, oltre ad essere consapevole di essere in un sogno, sia consapevole anche del fatto che il mondo gli si presenti in un modo differente, "costruito". Se così fosse la sua esperienza non avrebbe carattere presentazionale<sup>12</sup>.

Nonostante le premesse apparentemente favorevoli, T4 non riesce a rendere conto in modo soddisfacente del fatto che le immagini mentali possano essere accompagnate da un senso di realtà/carattere presentazionale. La concezione filosofica ordinaria presuppone che il carattere presentazionale sia una caratteristica necessaria perché un evento sia allucinatorio. Pertanto, allo stato attuale, T4 non è in grado di offrire una teoria soddisfacente delle allucinazioni come immagini mentali.

#### **4 Impasse, di nuovo?**

L'analisi delle due proposte teoriche che considerano le allucinazioni appartenenti allo stesso genere delle immagini mentali sembra riportare alla situazione di impasse presente nel dibattito tra disgiuntivismo e teorie del genere comune. Se, come ho sostenuto, T3 non è una proposta credibile, le alternative resta-

<sup>11</sup>Ma vedi Ichikawa (2009).

<sup>12</sup>Le due obiezioni presentate contro T4 in merito alla volontarietà e al senso di realtà sono valide anche per T1.

no T2 e T5, che sono le due teorie genuinamente connesse, rispettivamente, al disgiuntivismo e alle teorie del genere comune.

Nel seguito, tenterò di mostrare come, a partire da T5, si possa giungere a considerazioni interessanti a proposito dell'identificazione di genere tra allucinazioni e immagini mentali. Non offrirò, tuttavia, un argomento decisivo che sia in grado di mostrare l'insostenibilità di T2. Le motivazioni per preferire T5 a T2 sono tutt'altro che decisive e si basano fondamentalmente su una considerazione che riflette un atteggiamento filosofico di fondo<sup>13</sup>. È forse proprio su questo punto che è identificabile la cesura tra due visioni mutualmente esclusive sulla natura dell'esperienza percettiva.

T2 e, in generale, i disgiuntivisti sono generalmente restii a identificare stati mentali in virtù dei sottostanti meccanismi neurofisiologici, dato che, a seconda di come stanno le cose nell'ambiente del soggetto, uno stesso stato cerebrale può dar luogo a eventi appartenenti a differenti generi mentali. Come già sottolineato, però, in ambito scientifico c'è ampio consenso intorno all'idea che le percezioni, le allucinazioni e le immagini mentali condividano, almeno in parte, i medesimi meccanismi neurofisiologici.

La tesi che i fenomeni mentali siano generi naturali che possano essere classificati in base ai meccanismi neurofisiologici sottostanti è certamente dibattuta e controversa. Tuttavia, mentre T2 riflette un atteggiamento filosofico secondo cui accettare una teoria che propone di revisionare i nostri schemi classificatori del senso comune è un costo teorico, T5 riflette un atteggiamento secondo cui le teorie proposte in ambito filosofico devono aderire, in un certo modo, a vari dati empirici e scoperte scientifiche. T5, date queste premesse, è una candidata migliore di T2, che, sostenendo l'esistenza di tre generi mentali distinti, è invece la proposta che si allontana maggiormente dalle ipotesi scientifiche avanzate.

T5 sostiene che esiste un genere comune comprendente esperienze veridiche e allucinatorie e che l'immaginazione sensoriale appartiene a un differente genere mentale. Si tratta di una teoria genuina del genere comune che però, rispetto a T4, si trova in una situazione migliore nel rendere conto della mancanza

<sup>13</sup>Nonostante ciò, T2 non è esente da problemi. T2 propone la tesi che vi siano tre diversi generi mentali, quello percettivo, quello allucinatorio e il genere delle immagini mentali. Martin (2002) considera la percezione uno stato in cui si è in relazione con un oggetto, le allucinazioni come episodi semplicemente indistinguibili rispetto a una corrispondente percezione e gli episodi di immaginazione sensoriale come di natura rappresentazionale, la rappresentazione dell'esperienza di un dato oggetto (tesi della dipendenza). Nell'immaginare una scena, secondo Martin, si immagina l'esperienza di una scena e l'immagine mentale eredita le proprietà fenomeniche dello stato percettivo (Martin 2002). Questa teoria incontra tuttavia delle difficoltà nello spiegare le similitudini neurofisiologiche e comportamentali tra percezione e immagini mentali. Secondo Nanay (2016) non riesce a render conto delle scoperte scientifiche in merito alla somiglianza di movimenti oculari riscontrabili tra esperienze percettive e immagini mentali (Johansson, Holsanova e Holmqvist 2006). Nel caso delle immagini mentali non è chiaro perché il rappresentare un'esperienza debba includere specifici movimenti oculari, dato che è il contenuto di un'esperienza a esigerli.

di senso di realtà o carattere presentazionale delle immagini mentali (§3). Ciò è infatti imputabile, nella cornice teorica offerta da T5, alla differenza di genere mentale tra esperienze percettive e immagini mentali.

Se da un lato sembra in grado di rispondere efficacemente alle critiche a T4 in merito al carattere presentazionale, dall'altro T5 sembra, almeno *prima facie*, non avere le risorse per rendere conto delle similarità fenomeniche e neurofisiologiche tra esperienze percettive e immagini mentali.

Sul fronte delle similarità fenomeniche, Macpherson (2013) ha sottolineato come T5 abbia difficoltà a gestire le immagini mentali, dato che esse possono essere scambiate facilmente per esperienze visive, come l'esperimento di Perky (1910) mostrerebbe chiaramente. L'indistinguibilità fenomenica tra allucinazioni e percezioni è stata a lungo utilizzata dai sostenitori del genere comune, in particolare dai teorici dei dati sensoriali (Russell 1912), per argomentare in favore della comunanza di genere tra allucinazioni e percezioni. Tuttavia, il movimento disgiuntivista ha mostrato come questa tesi sia tutt'altro che scontata (Austin 1962; Putnam 1999). Il fatto che due stati siano fenomenicamente indistinguibili non ci autorizza a postulare un'uguaglianza di genere.

Il sostenitore di una teoria del genere comune potrebbe pertanto sostenere, adottando una strategia disgiuntivista, che l'esperimento di Perky non deve necessariamente essere considerato come una dimostrazione sulla medesima natura degli stati mentali coinvolti nell'esperienza percettiva e nelle immagini mentali. In questo modo si disinnescerebbe questa obiezione a T5.

Sul versante delle similarità neurofisiologiche, la tesi della quasi completa sovrapposizione riscontrabile tra le regioni cerebrali coinvolte nella produzione di immagini mentali e nell'esperienza percettiva (Kosslyn 1980) non sembra essere in contraddizione con T5. Non lo è ancora maggiormente se si menziona che questa tesi è sì accettata, ma dibattuta per quanto riguarda l'effettiva attivazione delle aree visive specializzate durante la produzione di immagini mentali (Ganis, Thompson e Kosslyn 2004; Howard 1998).

Secondo Domyntyche (2013) le immagini visive non comportano, infatti, l'attivazione delle aree visive specializzate della corteccia cerebrale<sup>14</sup>. Le allucinazioni della sindrome di Charles Bonnet sono, secondo Domyntyche, la testimonianza del fatto che le allucinazioni non possono essere trattate *tout court* come esemplificazioni dell'immaginazione sensoriale. Queste allucinazioni sem-

<sup>14</sup>Per aree visive specializzate Domyntyche intende le aree corticali che rispondono in modo preferenziale a un attributo visivo rispetto ad altri. È importante sottolineare, ad ogni modo, come queste aree non rispondano esclusivamente a un attributo visivo. Tra le aree visive specializzate che Domyntyche annovera riporto: la corteccia visiva primaria (V1), specializzata per il contrasto, l'orientamento e la luminosità, l'Area Paraippocampale dei Luoghi (PPA, *Parahippocampal place area*) importante per l'elaborazione dei luoghi, l'area V4, che risponde preferenzialmente ai colori e l'Area Fusiforme Faciale (FFA, *fusiform face area*), specializzata nel riconoscimento dei volti. Per l'elenco completo delle categorie visive e delle aree corticali interessate menzionate da Domyntyche rimando a Domyntyche (2013).

brano in larga parte non riguardare contenuti visivi precedentemente incontrati e sembrano richiedere l'attivazione delle aree specializzate del sistema visivo coinvolte nell'esperienza percettiva (ffytche 2013).

La tesi che vi siano almeno due generi sembra catturare, inoltre, la differenza tra "la mente dell'occhio" e "l'occhio della mente" che, secondo ffitche, emerge da un punto di vista neuro-fenomenologico. La mente dell'occhio (*eye's mind*), comporta l'attivazione di aree specializzate del cervello ed è riconducibile all'esperienza percettiva. L'occhio della mente (*mind's eye*), comporta l'attivazione del lobo frontale, parietale e mediale temporale con assente attivazione di aree visive specializzate (Howard 1998) ed è riconducibile all'immaginazione sensoriale.

In conclusione, T5 sembra avere le risorse per rispondere alle obiezioni riguardanti le similarità fenomeniche e neurofisiologiche tra esperienze percettive e immagini mentali. Adottare T5, tuttavia, significa al contempo invalidare l'ipotesi che le immagini mentali possano contribuire a sbloccare la situazione di impasse in cui si trova il dibattito in filosofia della percezione. Ritengo, invece, che T5 possa essere usata la base per sviluppare una nuova teoria che possa rendere conto degli elementi sin qui emersi. La tesi che intendo proporre è che, oltre all'esistenza di allucinazioni del sistema dell'esperienza percettiva, dovremmo ammettere l'esistenza di allucinazioni del sistema dell'immaginazione sensoriale.

## 5 Le allucinazioni sono immagini mentali?

La concezione filosofica ordinaria considera le allucinazioni come l'attività spontanea dei nostri apparati percettivi, in uno dei cinque sensi, in assenza di stimoli esterni (Allucinazioni<sub>EP</sub>). La percezione dovrebbe metterci in contatto con oggetti del mondo, ma nel caso allucinatorio ciò non accade. Come è emerso, le allucinazioni<sub>EP</sub> sono caratterizzate dai filosofi come involontarie e accompagnate da un carattere presentazionale. Le allucinazioni causalmente corrispondenti (*causally matching hallucinations*), assunte dal dibattito filosofico sulla percezione, rispettano i canoni delle allucinazioni<sub>EP</sub>. Lo stesso sembra valere anche per effettive occorrenze allucinatorie come le già menzionate allucinazioni da sindrome di Charles Bonnet.

Il tentativo di T1 e T4 di mostrare come le immagini mentali possano appartenere allo stesso genere mentale delle allucinazioni sembra non condurre a esiti positivi. Infatti, queste teorie non riescono a giustificare la tesi che le immagini mentali possano essere accompagnate da un senso di realtà/carattere presentazionale. A ogni modo, che esista un unico genere allucinatorio (Allucinazioni<sub>EP</sub>) è un'assunzione filosofica. Un'assunzione, tra l'altro, che sembra confliggere con le prospettive di ricerca più recenti sulle allucinazioni. Nella pratica clini-

ca è noto che le allucinazioni possano avere diverse eziologie (Teeples, Caplan e Stern 2009) ed emerge in misura crescente la necessità di “frammentare” la visione monolitica delle allucinazioni in modo da sviluppare teorie specifiche che possano portare a trattamenti differenziati (ffytche 2013).

Inoltre, una tradizione consolidata nella letteratura neuroscientifica e psichiatrica ha considerato e considera almeno alcune allucinazioni come strettamente connesse al fenomeno delle immagini mentali. È il caso, ad esempio, delle allucinazioni, solitamente uditive, che occorrono frequentemente nella schizofrenia<sup>15</sup> (Bentall 1990; Horowitz 1975). È a questa tradizione che si appoggiano T1 e T4 per difendere la tesi che le allucinazioni appartengano allo stesso genere mentale delle immagini mentali (T).

A questo punto, a mio avviso, è possibile rintracciare l'errore metodologico che commettono T1 e T4. Queste teorie propongono di motivare la tesi che le allucinazioni appartengano al genere delle immagini mentali (T), cercando di mostrare come quest'ultime possano soddisfare tutti i requisiti che sono richiesti dalla nostra concezione filosofica ordinaria delle allucinazioni. Viceversa, ritengo che è solo a partire da una corretta caratterizzazione delle immagini mentali che possiamo arrivare a ipotizzare l'esistenza di allucinazioni che appartengono al genere delle immagini mentali.

Come nel caso dell'esperienza percettiva, possiamo ipotizzare che esista un'attività spontanea dell'apparato immaginativo, in uno dei cinque sensi. Queste allucinazioni (Allucinazioni<sub>IM</sub>), data la caratterizzazione emersa sinora delle immagini mentali, sarebbero involontarie e, presumibilmente, non accompagnate da un carattere presentazionale/senso di realtà. Il funzionamento irregolare delle immagini mentali consisterebbe, pertanto, nella loro produzione involontaria.

Il sostenitore di T5 non è in alcun modo tenuto a far propria questa tesi. Sono state, invece, poste le basi per lo sviluppo di una nuova teoria (T6):

T6: Percezione e allucinazione<sub>EP</sub> sono lo stesso genere mentale, mentre immaginazione sensoriale e allucinazione<sub>IM</sub> sono un genere diverso.

L'ipotesi che esistano differenti generi allucinatori non è del tutto nuova in ambito filosofico. Macpherson (2013) ha recentemente esplorato questa possibilità e le considerazioni che ne ha tratto sembrano andare nella direzione proposta da T6. Macpherson ha distinto quattro differenti concezioni filosofiche delle allucinazioni. Di queste quattro concezioni, tuttavia, ha rintracciato nella letteratura scientifica soltanto le allucinazioni<sub>EP</sub> e le allucinazioni<sub>IM</sub>.

<sup>15</sup>Da una prospettiva filosofica, anche Currie e Ravenscroft (2002) hanno suggerito la tesi che le allucinazioni che si verificano nella schizofrenia siano riconducibili alle immagini mentali.

Prima di concludere prendo in considerazione una delle possibili obiezioni a T6. Se accettiamo la caratterizzazione delle allucinazioni<sub>IM</sub> proposta da T6, fenomeni come gli *earworms* e i *flashback*, episodi involontari di immagini mentali, dovrebbero essere classificati come allucinazioni<sub>IM</sub>. Si tratta di una tesi che può risultare inizialmente controversa: gli *earworms* e i *flashback* sembrano accadere con una certa frequenza e presso una porzione significativa della popolazione, mentre i fenomeni allucinatori sono considerati generalmente indicatori, più o meno affidabili, di una qualche condizione patologica sottostante. Ci sono almeno due strategie per rendere conto di questa tensione.

Una prima possibile risposta, che prende in seria considerazione la bontà dell'obiezione, consiste nell'ammettere che la nozione di involontarietà che utilizza T6 sia a grana troppo spessa. Ad esempio, potremmo essere interessati ad avere una teoria che distingua tra episodi di immagini mentali scatenati, in qualche modo, da eventi del mondo e episodi completamente involontari di immagini mentali. Solo questi ultimi, in questa prospettiva, dovrebbero essere considerati casi di allucinazioni<sub>IM</sub>.

La seconda risposta ha invece un costo maggiore, perché propone una strategia revisionista. L'idea è che le allucinazioni<sub>IM</sub> esistano in un *continuum* che varia da espressioni deboli nella popolazione normale fino ai sintomi caratteristici di condizioni patologiche. A tal proposito, è interessante menzionare il caso della schizofrenia che, come detto, è contraddistinta da allucinazioni generalmente ricondotte alle immagini mentali. Studi condotti su pazienti schizofrenici sembrano mostrare che la schizofrenia è correlata a un significativo incremento nella nitidezza delle immagini mentali (Sack et al. 2005). Ciò potrebbe essere indicativo del perché delle allucinazioni<sub>IM</sub> si presentino con una finezza di grana più simile alle allucinazioni<sub>EP</sub> rispetto a episodi come gli *earworms*. Il fatto che siano maggiormente assimilabili alla nostra concezione filosofica ordinaria delle allucinazioni, in questa seconda prospettiva, non significa che esse non siano allucinazioni<sub>IM</sub>.

Allo stato attuale non è chiaro quale, tra le due, sia la risposta migliore a questa obiezione. Entrambe presentano costi e benefici. Evidente è, invece, il carattere provvisorio della proposta presentata che, per il momento, è soltanto un'opzione teorica da sviluppare.

## 6 Conclusioni

L'analisi delle teorie filosofiche adottabili nei confronti della relazione tra immagini mentali e esperienze percettive ci ha condotti a ipotizzare l'esistenza di almeno due generi allucinatori: il primo riconducibile al sistema dell'esperienza percettiva (allucinazioni<sub>EP</sub>), il secondo al sistema delle immagini mentali (allucinazioni<sub>IM</sub>).

In sostanza, se la prospettiva di considerare le allucinazioni come appartenenti allo stesso genere delle immagini mentali (T) era finalizzata a una semplificazione del dibattito sulle allucinazioni, le considerazioni riportate in questo articolo non sembrano portare ai risultati sperati. Se T6 fosse realmente la miglior candidata a spiegare la relazione tra esperienze percettive e immagini mentali, infatti, saremmo giunti a duplicare il problema originario. In particolare, potremmo dover ammettere l'esistenza di allucinazioni appartenenti a due generi differenti.

Questo però non è necessariamente uno svantaggio. Tale nuova prospettiva potrebbe, infatti, essere utile per rendere conto di alcune caratteristiche delle allucinazioni che ancora non riescono a trovare una spiegazione soddisfacente nel consolidato dualismo disgiuntivismo-teorie del genere comune. Inoltre, abbastanza curiosamente, T6 sembra rendere contemporaneamente vere le intuizioni principali del disgiuntivismo e delle teorie del genere comune. Da un lato, infatti, accoglie l'istanza disgiuntivista per cui alcune allucinazioni, quelle derivanti dal sistema delle immagini mentali (allucinazioni<sub>IM</sub>), sono radicalmente differenti rispetto alle percezioni. Dall'altro, accetta la tesi del genere comune per cui le allucinazioni tradizionali (allucinazioni<sub>EP</sub>) appartengono allo stesso genere mentale delle percezioni. Per concludere, le allucinazioni, ma non tutte, potrebbero essere immagini mentali.

## Riferimenti bibliografici

- Allen, Keith (2015). "Hallucination and Imagination". In: *Australasian Journal of Philosophy* 93.2, pp. 287–302.
- Austin, John L. (1962). *Sense and Sensibilia*. Oxford: Oxford University Press.
- Bartolomeo, Paolo (2008). "The neural correlates of visual mental imagery: An ongoing debate". In: *Cortex* 44.2, pp. 107–108.
- Bentall, Richard (1990). "The Illusion of Reality: A Review and Integration of Psychological Research on Hallucinations". In: *Psychological Bulletin* 107.1, pp. 82–95.
- Byrne, Alex e Logue Heater (2008). "Either/Or". In: *Disjunctivism: Perception, Action, Knowledge*. A cura di Adrian Haddock e Macpherson Fiona. Oxford: Oxford University Press, pp. 57–95.
- (2009). "Introduction". In: *Disjunctivism: Contemporary Readings*. A cura di Alex Byrne e Logue Heater. Cambridge, MA: The MIT Press.
- Bértolo, Helder (2005). "Visual imagery without visual perception?" In: *Psychologica* 26.1, pp. 173–188.
- Calabi, Clotilde (2009). *Filosofia della Percezione*. Bari: Laterza.
- Campbell, John (2002). *Reference and Consciousness*. Oxford: Oxford University Press.
- Crane, Tim (2006). "Is there a Perceptual Relation". In: *Perceptual Experience*. A cura di Tamar S. Gendler e John Hawthorne. Oxford: Oxford University Press, pp. 126–146.
- Currie, Gregory e Ian Ravenscroft (2002). *Recreative Minds: Imagination in Philosophy and Psychology*. Oxford: Oxford University Press.
- Farkas, Katalin (2013). "A sense of reality". In: *Hallucination: Philosophy and Psychology*. A cura di Fiona Macpherson e Dimitris Platchias. Cambridge, MA: The MIT Press, pp. 399–416.
- ffytche, Dominic H. (2013). "The Halucitating Brain: Neurobiological Insights into the nature of Hallucinations". In: *Hallucination: Philosophy and Psychology*. A cura di Fiona Macpherson e Dimitris Platchias. Cambridge, MA: The MIT Press, pp. 45–63.
- Fish, William (2009). *Perception, Hallucination and Illusion*. Oxford: Oxford University Press.
- Ganis, Giorgio, William L. Thompson e Stephen M. Kosslyn (2004). "Brain areas underlying visual mental imagery and visual perception: An fMRI study". In: *Cognitive Brain Research* 20.2, pp. 226–241.

- Hinton, John Michael (1967). "Visual Experiences". In: *Mind* 76.302, pp. 217–227.
- Hopkins, Robert (2012). "What Perky did not Show". In: *Analysis* 72.3, pp. 431–439.
- Horowitz, Mardi (1975). "Hallucinations: An Information Processing Approach". In: *Hallucinations: Behaviour, Experience and Theory*. A cura di Ronald Siegel e Louis West. New York: John Wiley, pp. 163–195.
- Howard Robert, ffytche Dominic et al. (1998). "The functional anatomy of imagining and perceiving colour". In: *Neuroreport* 9.6, pp. 1019–1023.
- Hume, David (1739). *A Treatise of Human Nature*. A cura di Ernest C. Mossner. London: Penguin Books 1969.
- Ichikawa, Jonathan (2009). "Dreaming and imagination". In: *Mind and Language* 24.1, pp. 103–121.
- Johansson, Roger, Jana Holsanova e Kenneth Holmqvist (2006). "Pictures and Spoken Descriptions Elicit Similar Eye Movements during Mental Imagery, Both in Light and in Complete Darkness". In: *Cognitive Science* 30.6, pp. 1053–1079.
- Kalderon, Mark E. (2011). "Color Illusion". In: *Noûs* 45.4, pp. 751–771.
- Kosslyn, Stephen M. (1980). *Image and Mind*. Cambridge MA: Harvard University Press 2006.
- Kosslyn, Stephen M., William L. Thompson e Giorgio Ganis (2006). *The Case for Mental Imagery*. Oxford: Oxford University Press.
- Macpherson, Fiona (2013). "The Philosophy and Psychology of Hallucination: An Introduction". In: *Hallucination: Philosophy and Psychology*. A cura di Fiona Macpherson e Dimitris Platchias. Cambridge, MA: The MIT Press, pp. 1–38.
- Macpherson, Fiona e Clare Batty (2016). "Redefining illusion and hallucination in light of new cases". In: *Philosophical Issues* 26.1, pp. 263–296.
- Martin, Michael G.F. (2002). "The Transparency of Experience". In: *Mind & Language* 17.4, pp. 376–425.
- (2004). "The Limits of Self-Awareness". In: *Philosophical Studies* 120.1, pp. 37–89.
- (2006). "On Being Alienated". In: *Philosophical Studies* 120.1, pp. 37–89.
- McDowell, John (2013). "The disjunctive conception of experience as material for a transcendental argument". In: *Disjunctivism: Perception, Action, Knowledge*. A cura di Fiona Macpherson e Adrian Haddock. Oxford: Oxford University Press, pp. 376–389.

- Nanay, Bence (2015). "Perceptual content and the content of mental imagery". In: *Philosophical Studies* 172.7, pp. 1723–1736.
- (2016). "Imagination and Perception". In: *Routledge Handbook of the Philosophy of Imagination*. A cura di Amy Kind. London: Routledge, pp. 124–134.
- (2017a). *Multimodal mental imagery*. *Cortex*. URL: <http://dx.doi.org/10.1016/j.cortex.2017.07.006>.
- (2017b). "Sensory Substitution and Multimodal Mental Imagery". In: *Perception* 46.9, pp. 1014–1026.
- Nudds, Matthew (2009). "Recent Work in Perception: Naïve Realism and its Opponents". In: *Analysis* 69.2, pp. 334–336.
- Page, Jonathan W., Paul Duhamel e Michael A. Crognale (2011). "ERP Evidence of Visualization at Early Stages of Visual Processing". In: *Brain and Cognition* 75.2, pp. 141–146.
- Pagondiotis, Costas (2013). "Hallucination, mental representation and the presentational character". In: *Hallucination: Philosophy and Psychology*. A cura di Fiona Macpherson e Dimitris Platchias. Cambridge, MA: The MIT Press, pp. 361–380.
- Peacocke, Christopher (1992). *A Study of Concepts*. Cambridge MA: MIT Press.
- Perky, Mary C.W. (1910). "An Experimental Study of Imagination". In: *American Journal of Psychology* 21.3, pp. 422–452.
- Putnam, Hilary (1999). *The Threefold Cord*. New York: Columbia University Press.
- Pylyshyn, Zenon (1973). "What the Mind's Eye tells the Mind's Brain: A critique of Mental Imagery". In: *Psychological Bulletin* 80.1, pp. 1–24.
- Richardson, Alan (1969). *Mental Imagery*. New York: Springer.
- Robinson, Howard (1994). *Perception*. London: Routledge.
- Russell, Bertrand (1912). *I problemi della Filosofia* (t.o. *The problems of Philosophy*). A cura di Oxford University Press. Milano: Feltrinelli 2007.
- Sack, Alexander et al. (2005). "Enhanced vividness of mental imagery as a trait marker of schizophrenia?" In: *Schizophrenia Bulletin* 31.1, pp. 97–104.
- Searle, John (2015). *Seeing Things as They Are: A Theory of Perception*. Oxford: Oxford University Press.
- Snowdon, Paul (2005). "The formulation of disjunctivism: A response to Fish". In: *Proceedings of the Aristotelian Society* 105.1, pp. 129–141.
- Taylor, Steven, Dean MacKay e Euripedes C. Miguel (2014). "Musical obsessions: a comprehensive review of neglected clinical phenomena". In: *Journal of Anxiety Disorders* 28.6, pp. 580–589.

Teeple, Ryan, Jason Caplan e Theodore Stern (2009). "Visual hallucinations: Differential diagnosis and treatment". In: *The Journal of Clinical Psychiatry* 11.1, pp. 26-32.



## Evidentialism and Transparency: a Critical Approach

*Rodrigo Neira*

**Abstract.** With the current paper we will intend i) to elaborate a critical examination of the alleged use of transparency as independent of any epistemological view in favor of evidentialism, ii) to remark the inadequacies of the Absjorn Steglich-Petersen's attempt to replace the concept of belief with that of truth in reflective contexts in response to the claim of some evidentialist who want to use transparency to support their own dialectical purposes, and iii) to suggest a new approach to transparency which allows us to expand the theoretical horizons in which philosophical research on the subject has so far moved.

**Keywords.** Belief, Transparency, Evidentialism, Doxastic Deliberation.

## Introduction

Concerning the examination of one's own doxastic behavior, a preliminary characterisation of the phenomenon of transparency may be as follows: when a reflexive subject raises the question of whether to believe that  $p$  – when he is going to examine his epistemic state with respect to  $p$  –, the considerations he selects are the same as if he has to answer the question of whether  $p$  is the case. The phenomenon of transparency consists in transferring the question about one's own inner doxastic constitution to the objective consideration of the truth value of  $p$ . In this sense, it may be said that the doxastic behavior are transparent to factual aspects of the world.

### 1 Transparency and conceptual evidentialism

Evidentialism can be understood as the view that the doxastic behavior of a rational subject – or at least the conscious doxastic behavior of a rational subject – is governed by the principle of believing that  $p$  only if  $p$ . Bernard Williams (1973, p. 136) suggests that a reference to truth is one of the basic features of beliefs without apparently trying to defend evidentialism: "Beliefs aim at truth". Miriam McCormick (2015, p. 15) notes that Williams establishes that the aim at truth which characterizes a person's doxastic behavior is the expression of a conceptual truth about belief. The evidentialist principle can only be accepted once the subject has the concept of belief. Understanding what it means to have a belief involves admitting that it aims at truth.

This form of evidentialism is called "conceptual". The main characteristic of conceptual evidentialism as opposed to "doxastic freedom" evidentialism is that it considers that the norm of believing that  $p$  on the basis of sufficient evidence in favor of the truth of  $p$  is not a moral imperative. Conceptual evidentialism has different versions which can be distinguished according to their intensity. Jonathan Adler (2002), who presents the most extreme evidentialist conception, argues that it is conceptually impossible to harbor a belief while knowing that there is not enough evidence which supports it. Adler argues that certain assertions inspired by Moore's paradox are unsustainable. Consider the case of the following conjunction inspired in Moore's paradox: "The number of stars is even, but I do not have enough evidence that the number of stars is even". Since both are statements which can be true at the same time, the contradiction which prevents the subject from accepting the conjunction should manifest itself in his thought. By asserting that the number of stars is even, we express our belief that the number of stars is even, but if we do not have sufficient evidence in favor of the truth of the proposition that the number of stars is even, we are not entitled to consider it true while judging whether to adopt it as the object of any of our

beliefs.

Adler's evidentialism has been widely contested. McCormick (2015, pp. 19-21) has described cases of gaps or inconsistencies between a subject's best epistemic judgments and the beliefs he adopts. Jens Dam Ziska (2016, pp. 41-44) Ziska has denied that a conjunction of the form "p, but I do not have sufficient evidence that p is true" entails by itself any contradiction in the subject's thought without the additional premise that when we realize that the evidence we possess is not enough to assure that p is true, we cease to believe that p is true.

This premise seems weak if we consider the examples of people who have certain doxastic behavior while knowing that the evidence they count with is not enough to assure that their beliefs are true. Adler could answer that many of these cases are "esoteric [...] like those which can occur in the thought of a mentally disturbed person" (Adler, 2002, p. 35), so that these objections reinforce his theory. But even if that were true, a philosophical doctrine must measure its success according to its ability to respond to the most peripheral and challenging cases.

A softer version of conceptual evidentialism is put forward by Nishi Shah. In an article written in collaboration with David Velleman (see Shah and Velleman 2005), Shah argues that transparency occurs only when a belief is adopted in a deliberative context, that is to say, in a context in which the subject makes use of the concept of belief. According to Shah, transparency implies that in deliberative circumstances the subject only takes into account those reasons which can be seen as relevant to determine the truth or falsehood of the proposition under consideration. The concept of belief imposes a condition which must be accepted by the subject if he wants to conceive his personal deliberation as aimed at adopting a belief. This condition is that deliberating about whether to believe p implies accepting that the belief that p is only correct if p is true: "When one deliberates about whether to have an attitude conceived as the belief that p is true, one deliberates about an attitude to which the standard of being correct only if p is true already applies" (Shah and Velleman 2005, p. 501). This explanation of transparency as exclusively linked to those contexts in which the subject makes use of the concept of belief is compatible with the existence of elements perceived as irrelevant with respect to the truth or falsehood of the proposition under consideration, which can influence on whether it is adopted as doxastic content. When the subject does not make use of the concept of belief, there is not any conceptual imperative which implies a selection based on the truth or falsity of the proposition.

Shah and Velleman's argument is based on transparency understood as an epistemologically neutral phenomenon. In another article which was published a little earlier (Shah 2003, p. 466), Shah refers to the need for an epistemological framework capable of integrating the fact that the subject who reflectively

adopts a belief only takes into account the considerations he sees as relevant to determine whether the proposition in question is true or false. The solution suggests the existence of different contexts of belief formation in relation to whether or not the subject makes use of the concept of belief. As long as the subject conceives his deliberation as directed to the adoption of a belief, he must accept that the resulting propositional attitude would be correct only if it is true, so that he is obliged to base either his belief or his non-belief on the rational considerations he is able to gather about the truth or falsehood of the proposition he is reflecting about. These are the main lines of Shah and Velleman's strategy: this particular version of conceptual evidentialism provides a plausible explanation for transparency.

## 2 Objections against Shah and Velleman's approach

Shah and Velleman represent the use of transparency as case for evidentialism. We will now raise some objections against their approach and examine an attempt to correct some of its fundamental points. Shah and Velleman seem to suppose a conception of doxastic correction which can be described as externalist, since it seems to follow from it that the subject does not have access to the truth or falsehood of his beliefs, but only to the sum of what he perceives as evidence in favor of one or the other. The subject cannot be fully certain that his belief is true. However, the standard of correction for belief affects the doxastic behavior of the subject and allows it to be framed in a scheme of assessment. That is to say, if the belief that *p* is correct when it is true and incorrect when it is false, then there is an imperative which affects the behavior of the subject who has been asked the question whether he believes *p* and is trying to answer it. The subject must base his attitude with respect to whether *p* is the case on the considerations that he sees as relevant to determine the truth value of *p*. The problem with Shah and Velleman's approach here is that it seems to assume that the subject can stop basing his doxastic attitudes on evidence about what is the case. The fact that any given behavior is correct in any sense does not make it mandatory, but seems to involve the possibility of violating the norm.

If there is the possibility of acting both correctly and incorrectly in the reflexive formation of a belief, the subject who reflects on that matter should be able to make the decision of not taking into account the evidence at his disposal. The subject should be able to evade the alleged gravitational force of evidence at the cost of wrongdoing. This seems incompatible with understanding transparency in Shah's manner as conceptually inseparable from the reflective formation of beliefs: it would be the subject who identifies the question of whether to believe that *p* with the question of whether *p*, so that transparency could not be epistemologically neutral. Transparency should be the requirement for compli-

ance with an epistemic obligation analogous to the ordinary moral duty. The introduction a norm of correction specifically linked to belief seems to imply an ethical consideration: the subject must be free not to act in accordance with the norm. In the reflective formation of a belief, the subject may disregard the evidence and accept a proposition on the basis of some kind of spiritual peace.

Until now we have tried to illustrate the apparent inconsistencies which can be found in Shah and Velleman's approach. If one speaks of correct doxastic behavior, it seems that one must admit that the subject can evade transparency even in reflective contexts only by refraining from following the relevant norm. Transparency would no longer have a factual character, which is precisely the starting point of Shah's argument, but would be presented as the consequence of a previous choice of the subject. This would make it possible to introduce multiple criteria on the basis of which to judge beliefs (see Zalabardo 2010). Cannot a belief be correct or incorrect in different ways? Cannot the subject who does not take into account the evidence be wrong according to one criterion, but right from the point of view of other criterion? A belief would not be right or wrong in an absolute sense.

Other objections which can be raised against Shah's approach relate to the nature of the considerations involved in the reflective formation of beliefs. Shah seems to characterize the reflective adoption of beliefs as a deductive argument whose premises are constituted by considerations that the subject understands as relevant to determine the truth value of the proposition in question, that is to say, its conclusion. This is the basic criteria. Starting from this way of understanding the logical structure of the reflexive adoption of beliefs, it is possible to explore two strategies against Shah's approach: the first takes into account that the conclusion of personal deliberation about what to believe as Shah understands it is nothing more than a factual proposition: when the subject raises the question of whether to believe that  $p$ , his logical activities lead, according to Shah's model, to the utterance that something is the case. However, it is not clear that a factual statement arrived at as the conclusion of an argument will be enough to make it an object of belief. If its logical structure is that of a deductive argument which seeks to determine whether the proposition being examined is true or false, the deliberation on what to believe will end just with a fact, that is, the conclusion of the argument is not a belief, but the statement that something is the case. In any case, an additional premise would be needed to link what is known as true with the corresponding belief. But the problem with taking that step is that it seems to introduce a certain discordance which goes against Shah's requirement of immediacy. It would seem that in this way the subject of personal deliberation about what to believe does not do anything more than a logical calculation based on premises with which he maintains a purely intellectual relationship. With that additional premise it seems that Shah would incur in

circularity: considerations unrelated to the truth or falsity of the proposition do not intervene in the reflective formation of beliefs, since there is a premise in the personal deliberation about what to believe which prevents it. But it is the relevance of introducing such a premise that is under discussion. The evidentialist characterization is based on the deliberation about what to believe as part of the concept of belief, and at the same time the concept of belief characterized in the evidentialist way is used to explain a truth about the reflective formation of beliefs. It would be said that there is something problematic in approaching the logic of the reflective formation of beliefs from the understanding of the concept of belief that is obtained from the analysis of deliberation about what to believe.

The second argumentative strategy that we will explore against Shah's approach has to do with the kind of reasons on the basis of which the subject adopts a belief. Is it not possible to find examples of beliefs which are conscious and contradict the evidence? It seems that a person may sincerely declare a belief while knowing that the evidence does not support the proposition. The believer can do without the evidence. However, this does not occur under whichever condition, but when the belief fulfils certain characteristics. Let us think of the case of a religious person who believes in the metempsychosis of souls or the transubstantiation. The religious believer perseveres in his doxastic behavior without seeming to care that the evidence does not support it. Is it enough to accuse the religious believer of epistemological negligence? The same should be said of much of humankind. Would not our position be arrogant? In relation to Shah's approach, religion offers examples of conscious doxastic behavior which cannot be adopted at the end of a process of personal deliberation about what to believe which concludes with the affirmation of a factual proposition as true. Why does the religious believer maintain, even in reflective conditions, doxastic behavior whose content he cannot consider true? Here, there is something similar to an asymmetry between purely factual and critical-religious contexts. Shah's conception may be valid when it is about proposition corresponding to truth conditions which can be specified. But the reasons which the religious believer may invoke in favour of his doxastic conduct will not be such that they can be seen as relevant to determine whether processes such as the metempsychosis or the transubstantiation are taking place, that is, they cannot function as premises in the personal deliberation about what to believe in order to determine whether the proposition is true or false. The evidence does not seem to be taken into account in the reflective examination of one's doxastic behavior when the content of the belief is within the bounds of religion.

### 3 **Steglich-Petersen and his attempt to set out a new model of transparency**

We have raised until now three objections against Shah's evidentialism: we have seen that the imposition of a standard of correction over conscious doxastic behavior seems to imply the possibility of believing something against the evidentialist requirement; we have examined the problems which arise from personal deliberation understood as a deductive argument; and we have considered the difficulties of Shah's approach to the doxastic behavior of certain agents who adopt beliefs in reflective contexts while knowing that they cannot invoke reasons to determine the truth or falsity of the given content. But if we accept these objections, we will not only have to abandon Shah's approach. What is the state of transparency once objections have been raised to the point of rejecting any preference for this way of understanding the reflective examination of one's doxastic behavior? In so far as it is understood in the manner of Shah, transparency does not seem to occur in the reflective formation of beliefs aimed at the question of whether to adopt a religious content as object of doxastic endorsement. But according to the other objections, it would also be said that it does not fulfil some of the characteristics which Shah attributes to it while exposing the phenomenon. If the subject may not fulfil evidentialist requirements and it is possible to speak of doxastic inadequacy as a possibility against doxastic adequacy, should we not consider that transparency is not a genuine phenomenon, that is, that it occurs as a consequence of a choice of the subject? And if the considerations not relevant to determine the truth or falsity of the given proposition are excluded from the deliberation about what to believe understood according to the logical form of a deductive argument whose conclusion is the affirmation of a fact, is it not also because the subject has preferred an evidentialist standard rather than other?

Absjorn Steglich-Petersen (2013) has tried to offer a formulation of transparency designed to avoid having to consider it as a feature of the subject's deliberation about what to believe by making use of the concept of belief. Steglich-Petersen's strategy is based on an interpretation of what we often mean when we think of the question of whether to believe that *p*. Starting from the description of a phenomenon analogous to transparency, transparency\*, Steglich-Petersen seeks a common explanation for both. Transparency\* consists in the fact that, when we consider the question of whether *p* is the case, we base our response on those considerations which we take as relevant to determine the truth or falsity of *p*. In contrast to transparency, transparency\* invokes a type of deliberation in which the concept of belief does not appear. Apart from that, transparency and transparency\* are identical phenomena, including the possibility that both the question of whether to believe that *p* and the question of whether *p* have an

unconscious or thoughtless answer without need for the subject to engage in a deliberative process.

Steglich-Petersen argues that, given similarities between transparency and transparency\*, they should be cases of the same general principle. This implies that transparency is not to be explained by using the concept of belief, i.e. as a feature of deliberation in which the subject makes use of the concept of belief, but in a different and more general way which also explains transparency\*. In Steglich-Petersen's view, this common explanation comes from considering that both transparency and transparency\* are instances of a principle which states that it is not possible to do something for a certain purpose while being aware that what is being done is not conducive to the achievement of that purpose. According to this, deliberation about whether p is the case, in which transparency\* is given, is an activity aimed at correctly answering the question of whether p is the case. Considerations which are not relevant to determining the truth or falsity of p will also fail to answer correctly the question of whether p is the case. Thus a person cannot attempt to answer the question of whether p on the basis of considerations which are known to be irrelevant to determine the truth or falsity of p.

How could transparency be considered as an instance of the principle that it is not possible to do something aiming at a particular purpose while knowing that the purpose in question will not be achieved by that means? Transparency is characterised by the fact that it links two apparently different questions: transparency consists in that the question of whether to believe that p is treated from the point of view of the subject in the same way as the question of whether p. With respect to transparency\*, however, there is only one question involved: the subject asks whether p is the case and bases his answer on those considerations which can be selected as relevant to determine whether p is the case. The solution proposed by Steglich-Petersen is to assume that the question of whether to believe that p can be addressed in different contexts. Transparency only occurs when what the subject means in posing the question whether to believe that p is just whether p. Then the subject expresses with the verb "believe" a question about a fact. The differences between transparency and transparency\* are merely linguistic. Transparency and transparency\* are identical in fundamental aspects, so that the explanation for both must be the same. Transparency must be explained in the same way as transparency\*: it must receive an explanation in terms of the principle that it is not possible to do something for a particular purpose while being aware that what is being done does not fulfil that purpose. Both phenomena are instances of this more general principle.

#### 4 Steglich Petersen's criticism of Shah's approach

It seems that Steglich-Petersen's conception is free of the apparent weaknesses on the basis of which we have formulated some of our objections against Shah's approach. In turning personal deliberation about what to believe into personal deliberation about what is the case, Steglich-Petersen does not need the concept of belief. That is to say, it is no longer possible to speak of a standard of correction which is given by the concept of belief and from which the subject can deviate by denying that transparency is conceptually inseparable from the reflective formation of beliefs, so that the objection formulated above is deactivated here. Likewise, Steglich-Petersen renounces the evidentialist model of doxastic legitimacy extracted from the concept of belief, so his approach eludes the logical accusation of circularity. If deliberation about what to believe is dispensed with when the concept of belief is used, the considerations as to whether *p* is the case in the relevant deliberation cannot be a consequence of the application of the concept of belief.

But Steglich-Petersen's model does not seem so effective when integrating the case of religious beliefs. We have seen that issues such as the metempsychosis or the transubstantiation seem immune to the evidence. Steglich-Petersen seeks to solve the problem by distinguishing a special deliberative context in which the question of whether to believe *p* is not identified with the question of whether *p* is the case. Steglich-Petersen uses the example of the belief in life beyond death. Aware of the dizzy transience of life and paralysed by the perspective of an absolute annihilation, a person may wonder whether, although the evidence does not allow to reach any conclusion on the truth or falsity of the matter, the belief in life beyond death should be adopted. A person in such a situation is not merely concerned with life beyond death, but rather with whether the question about a belief should be approached on the basis of nothing more than the truth value of a proposition. In cases like that, deliberation is genuinely about whether to believe something and should not be identified with the determination of the truth value of a proposition. Steglich-Petersen suggests that transparency does not occur in deliberative contexts where the concept of belief is important by itself rather than merely expressing the subject's interest in determining whether something is the case. As a result, religious beliefs are excluded from transparency.

Should we resign ourselves to this conclusion? Transparency would then be a property of certain beliefs. Deliberation about whether to believe in life beyond death is not a mere consideration of the truth or falsity of the proposition. The problem is not that the question of whether or not there is life beyond death cannot be asked, but that it would be asked as a consequence of the subject's willingness to approach the question from a certain point of view. The fact that

Steglich-Petersen is forced to deny that transparency manifests itself in relation to certain beliefs can be seen as a sign of the weakness of his approach. If the Steglich-Petersen's conception is taken for granted, transparency cannot be a conceptual property of beliefs. The foreseeable recourse to deny that religious beliefs are beliefs would contradict the non-problematic everyday practice of considering them as beliefs.

We encounter a dilemma. We could take a view such as Steglich-Petersen's and admit that transparency has exceptions in certain beliefs whose truth or falsity cannot be determined on the basis of the evidence, so that transparency would not be a genuine phenomenon. The other option is to develop a new formulation of the concept of transparency which is capable of explaining those cases which are problematic. Success in this purpose would ensure that transparency is a genuine phenomenon. If we look at the second term of the dilemma, we can try to find out whether transparency can in any sense be attributed to religious beliefs. This means acting on the concept of transparency both by purging it of Shah's evidentialist servitudes and by avoiding converting it to the trivial consequence of the subject's interest in facts, masked by the use of the verb of propositional attitude.

Remember the feature we have remark on religious beliefs. They are immune to the evidence. But being immune to the evidence in the case of religious beliefs does not mean that they are immune to the truth of their content. The religious believer can behave independently of evidence without avoiding truth, that is to say, without abandoning the aspiration that his belief is true. According to these considerations, we could perhaps develop the second term of the dilemma. Could not transparency be defined as the phenomenon consisting of that the subject would answer the same to the question whether to believe that  $p$  than to the question whether  $p$ ? That would involve not taking into account which considerations are used as the basis for the answer. The approach we are suggesting here probably implies not to imposing any qualification on the reasons on the basis of which the subject adheres to a belief. Thus, for transparency to be attributed to beliefs such as the metempsychosis or the transubstantiation, it might be enough for the person who has them to answer affirmatively to the question about the truth of those doxastic contents.

These considerations, the inclusion of which is intended to establish a contrast with our previous destructive work, suggest what might be the essential lines of a concept of transparency purged of evidentialist servitudes. Emancipating the transparency from evidentialism is the same as denying a basic assumption on which some of the most widespread approaches to the subject's relationship with his own doxastic behavior are based — that there is an epistemic rule of law which the concept of belief imposes on deliberative contexts— and implies opting for a weak interpretation of the principle that beliefs aim

at truth. Steglich-Petersen adopts a similar starting point, but his transparency without doxastic deliberation suffers from explanatory deficiencies with regard to religious beliefs which are also found in the evidentialist version. Developing a concept of transparency free of the failures due to an evidentialist affiliation is a task that we hope to tackle at another time.



## References

- Adler, Jonathan (2002). *Belief's Own Ethics*. Cambridge, Massachusetts: The MIT Press.
- McCormick, Miriam (2015). *Believing Against the Evidence: Agency and the Ethics of Belief*. New York: Routledge.
- Shah, Nishi (2003). "How Truth Governs Belief". In: *The Philosophical Review* 112.4, pp. 447–482.
- Shah, Nishi and David Velleman (2005). "Doxastic Deliberation". In: *The Philosophical Review* 114.4, pp. 497–534.
- Steglich-Petersen, Absjorn (2013). "Transparency, Doxastic Norms and the Aim of Belief". In: *Teorema* XXXII.3, pp. 59–74.
- Williams, Bernard (1973). *Problems of the Self*. New York: Cambridge University Press.
- Zalabardo, José (2010). "Why Believe the Truth? Shah and Velleman on the Aim of Belief". In: *Philosophical Explorations* 13.1, pp. 1–21.
- Ziska, Jens Dam (2016). "Deliberation and the First-Person Perspective". In: *Teorema* XXXV.1, pp. 35–57.





## Wilfrid Sellars e la normatività del significato

Andrea Raimondi

**Abstract.** Una tesi che è stata ultimamente discussa dai filosofi analitici del linguaggio è la tesi della *normatività del significato linguistico* (TNS). Secondo TNS, i significati delle espressioni delle lingue naturali hanno una componente normativa: essi fissano uno standard di correttezza per l'applicazione o l'uso (da parte dei parlanti) delle espressioni stesse e, in conseguenza di ciò, impegnano i parlanti a determinati obblighi semantici nell'applicazione o nell'uso delle espressioni. In questo articolo, ricostruisco e difendo alcune riflessioni di Wilfrid Sellars a favore di questa tesi. In particolare, mi concentro sulle riflessioni contenute in "Empiricism and the Philosophy of Mind" (1956) e sviluppate in "Language as Thought and as Communication" (1969) e "Meaning as Functional Classification. A Perspective on the Relation of Syntax to Semantics" (1974). La struttura dell'articolo è la seguente. Nella prima sezione, presento TNS nella sua versione più generale. Nella seconda sezione, mostro alcuni casi di enunciati descrittivi che sono in grado di trasmettere, in determinati contesti di proferimento, una forza normativa: sostengo che un fenomeno simile sia caratteristico anche degli *enunciati semantici* – enunciati della forma "L'espressione linguistica *e* significa *M*". Questo è il cuore della versione di TNS che vado a discutere. Nella terza, quarta e quinta sezione, discuto l'analisi di Sellars degli enunciati semantici. La conclusione della sua analisi è che un enunciato semantico predica l'identità dei ruoli funzionali di due parole. Mostro come Sellars arrivi a questa analisi, dedicando ampio spazio alla critica ad un'analisi alternativa, secondo la quale gli enunciati semantici sono enunciati descrittivi che predicano una relazione di associazione tra un'espressione linguistica e un oggetto o proprietà. Mi soffermo sulle ragioni teoriche che giustificano il rifiuto di Sellars

di questa analisi, concentrandomi su alcune tematiche classiche del progetto epistemologico sellarsiano. Nella sesta sezione, dopo aver argomentato che la nozione di *pragmema* proposta da Capone (2005, 2018) può aiutare a comprendere e ampliare la nozione sellarsiana di ruolo funzionale, mostro come tale analisi possa essere sfruttata per sostenere che gli enunciati semantici sono enunciati metalinguistici che hanno delle conseguenze normative sui comportamenti linguistici dei parlanti. Accenno infine al dibattito contemporaneo e alla formulazione di un argomento oggi discusso a favore di TNS, a partire da una rivisitazione della nozione di ruolo funzionale.

**Keywords.** Normatività, Significato, Sellars, Pragmatica del linguaggio.

## Introduzione

Una tesi che è stata ultimamente discussa dai filosofi analitici del linguaggio è la tesi della *normatività del significato linguistico* (TNS). Alcuni tra i più importanti sostenitori contemporanei di una qualche variante di TNS sono Boghossian (1989), Gibbard (2012), Glock (1995), Peregrin (2012, 2014), Whiting (2007, 2009, 2013) e Zalabardo (2012). Secondo TNS, i significati delle espressioni delle lingue naturali hanno una componente normativa: essi fissano uno standard di correttezza per l'applicazione o l'uso (da parte dei parlanti) delle espressioni stesse e, in conseguenza di ciò, impegnano i parlanti a determinati obblighi semantici nell'applicazione o nell'uso delle espressioni. In questo articolo, ricostruisco alcune riflessioni di Wilfrid Sellars a favore di questa tesi, nella convinzione che un'adeguata rielaborazione di queste riflessioni possa giovare all'attuale dibattito sulla normatività del significato linguistico. La struttura dell'articolo è la seguente.

Nella prima sezione, presento TNS nella sua versione più generale e meno impegnativa, accennando al problema della priorità metafisica delle norme. Nella seconda sezione, mostro alcuni casi di enunciati descrittivi che sono in grado di trasmettere, in determinati contesti di proferimento, una forza normativa: sostengo che un fenomeno simile sia caratteristico anche degli *enunciati semantici*<sup>1</sup> – enunciati della forma “*e* significa *M*” (dove *e* è un'espressione linguistica, e *M* il suo significato). Questo è il cuore della versione di TNS che le riflessioni sellarsiane sembrano difendere. Nella terza, quarta e quinta sezione, discuto l'analisi di Sellars degli enunciati semantici. La conclusione della sua analisi è che un enunciato semantico predica l'identità dei ruoli funzionali di due parole (una nella lingua oggetto e una nella lingua in cui è formulato l'enunciato, la lingua che funge da metalinguaggio). Mostro come Sellars arrivi a questa analisi, dedicando ampio spazio alla critica ad un'analisi alternativa. Nella sesta sezione, argomento che la nozione di *pragmema* proposta da Capone (2005, 2018) chiarisce e amplia la nozione sellarsiana di ruolo funzionale; in seguito, mostro come l'analisi sellarsiana possa essere sfruttata per sostenere che gli enunciati semantici sono enunciati metalinguistici che hanno delle conseguenze normative sui comportamenti linguistici dei parlanti. Infine, accenno ad un argomento contemporaneo a favore di TNS, ricollegandolo ad alcune nozioni discusse da Sellars.

<sup>1</sup>deVries e Triplett (2000).

## 1 La normatività del significato e il problema della priorità metafisica delle norme

Nella sua versione più generale, TNS dice che il significato linguistico è essenzialmente normativo. Un altro modo di formulare TNS consiste nel sostenere che è metafisicamente necessario che non ci sia significato linguistico senza norme. Come metterò in luce più avanti, tali norme sono norme per l'applicazione delle espressioni linguistiche: il significato di una qualsiasi espressione linguistica sembra essere intimamente connesso alla possibilità che quell'espressione venga applicata o usata *correttamente* o *scorrettamente* dai parlanti (laddove la correttezza è l'adeguamento alla norma e la scorrettezza è la sua violazione).

Di qui, è possibile porsi un problema di *priorità metafisica*: le espressioni linguistiche hanno un significato in virtù del fatto che certe norme per la loro applicazione sono in vigore *oppure* certe norme per l'applicazione di espressioni linguistiche sono in vigore in virtù del fatto che le espressioni linguistiche hanno un significato? In altri termini: le norme associate ai significati linguistici sono norme *costitutive* o *non-costitutive* dei significati linguistici stessi? Pur essendo una domanda interessante, non credo che la risposta che si è portati a dare sia centrale per apprezzare il contenuto di TNS. TNS, infatti, si limita a dire che le norme sono essenziali per i significati linguistici. Non specifica – e non è tenuta a specificare – in che “ordine” dobbiamo intendere le norme rispetto ai significati. TNS è compatibile sia con la tesi secondo la quale il significato delle espressioni è costituito dalle norme per la loro applicazione, sia con la tesi secondo la quale il significato è costituito da altri fattori (battesimi e catene causali, ruoli inferenziali, convenzioni e disposizioni, ecc.) e, nelle pratiche linguistiche, viene associato a certe norme per l'applicazione. Sostenitori autorevoli di TNS, come Blackburn (1984), Boghossian (1989), Kripke (1982) e Whiting (2007) rimangono neutrali sulla questione della priorità metafisica.

Che si impegni o meno alla priorità metafisica delle norme, il normativista accetta la tesi che ho formulato sopra e che ora riscrivo più rigorosamente:

**TNS** Un'espressione linguistica *e* significa *M* per un parlante *S* al tempo *t* solo se una norma per l'applicazione di *e* è in vigore per *S* a *t*.

Nel dibattito contemporaneo, la letteratura sul contenuto prescrittivo delle norme semantiche è molto ricca: alcuni sostengono che le norme semantiche sono norme *ipotetiche*<sup>2</sup> (cioè che sono in vigore per i parlanti solo se questi intendono metter capo ad atti comunicativi di successo); altri ritengono che esse siano norme *categoriche*<sup>3</sup> (cioè che valgano indipendentemente dalle intenzio-

<sup>2</sup>Si veda, ad esempio, Glüer e Wikforss (2015).

<sup>3</sup>Si vedano, ad esempio, Peregrin (2014) e Whiting (2013).

ni comunicative dei parlanti). Credo ci siano buone ragioni per sostenere che la seconda tesi sia quella più adeguata: è parte centrale del significato, ad esempio, della parola “gatto” il suo essere associata a delle regole d’uso, regole che stabiliscono che essa può essere applicata a tutti e soli quegli oggetti che hanno certe caratteristiche rilevanti, ovvero quelle dei gatti (*quali* siano poi queste caratteristiche è una questione distinta, che in parte dipende dalle nostre classificazioni del mondo animale e in parte dipende dalle scoperte delle scienze empiriche). Se i parlanti decidono di usare la parola “gatto” per riferirsi ai cani, allora potremmo scegliere se valutare questa situazione come un caso in cui i parlanti violano sistematicamente la norma semantica rilevante per “gatto”, o come un caso in cui vi è un cambiamento sostanziale nel vocabolario della comunità linguistica (“gatto” diventa un’espressione che significa *cane*).

Non è però della natura delle norme semantiche ciò di cui desidero occuparmi qui. Piuttosto, voglio soffermarmi su TNS e ricostruire quelle che ritengo essere osservazioni interessanti in suo favore; queste osservazioni consistono sostanzialmente nell’analisi condotta da Wilfrid Sellars degli enunciati semantici, enunciati della forma “*e* significa *M*”. Con *analisi* intendo uno studio della forma logica degli enunciati semantici, forma che dovrebbe rivelare ciò che questi enunciati *dicono* o, come ci si può anche esprimere, l’informazione trasmessa da questi enunciati. La domanda di partenza può dunque essere formulata nei modi seguenti: cosa stiamo dicendo quando diciamo che una certa parola dell’italiano significa questo-e-questo? Che tipo di informazione stiamo trasmettendo quando asseriamo che un’espressione di una lingua straniera significa una certa cosa ed è traducibile quindi con una certa espressione della nostra lingua?

Nel celebre saggio del 1956, “Empiricism and the Philosophy of Mind” (EPM), Sellars dedica qualche pagina ad una critica di un’analisi classica (di matrice tarskiana e carnapiana) degli enunciati semantici e avanza una proposta alternativa, già abbozzata in testi precedenti, come “Inference and Meaning” (1953) e “Some Reflections on Language Games” (1954), e sviluppata successivamente in “Language as Thought and as Communication” (1969) e “Meaning as Functional Classification. A Perspective on the Relation of Syntax to Semantics” (1974). È a questa analisi che dedicherò la mia attenzione. In via preliminare, occorre però chiarire perché l’analisi degli enunciati semantici dovrebbe essere rilevante ai fini di una difesa di TNS.

## 2 Casi comuni di enunciati descrittivi con usi normativi

Sostengo che l’analisi degli enunciati semantici possa svolgere un ruolo rilevante ai fini di una difesa di TNS poiché abbracciare TNS (o una sua versione) coin-

cide con il sostenere che gli enunciati semantici non sono enunciati descrittivi, ma sono, piuttosto, in qualche modo analoghi ad enunciati normativi; un'analisi che mostrasse la natura normativa degli enunciati semantici rappresenterebbe quindi un argomento a sostegno di TNS. Credo che l'esito dell'analisi di Sellars possa essere interpretato in modo tale da sostenere questa posizione. Occorre tuttavia notare che in EPM (e negli altri testi), Sellars non propone la sua analisi come un argomento direttamente a sostegno di TNS né la sua analisi si conclude con una esplicita riformulazione degli enunciati semantici nei termini di enunciati che fanno uso di un vocabolario irriducibilmente normativo; come metterò in luce, credo che Sellars sfrutti la sua analisi con il duplice intento di attaccare l'incrocio di due versioni del Mito del Dato e di rimanere coerente con la sua posizione epistemologica. Prima di entrare nel vivo del discorso, propongo alcuni esempi di enunciati semantici:

- (1) 'Giallo' (in italiano) significa *giallo*.
- (2) 'Yellow' (in inglese) significa *giallo*.
- (3) 'E' (in italiano) significa *e*.
- (4) 'Fenice' (in italiano) significa *fenice*.

(1)-(4) non contengono termini deontici e dunque non sembrano avere una forza normativa, a differenza dei seguenti enunciati:

- (5) Devi mangiare la minestra.
- (6) Tutti dovrebbero avere accesso ai beni primari.
- (7) Non si deve mangiare a bocca aperta.
- (8) Puoi stare a torso nudo.

Il fatto che (1)-(4) siano superficialmente diversi da (5)-(8) non implica che non sia possibile proporre un'analisi di (1)-(4) che sveli la natura normativa del verbo "significare". Di fatto, interpreterò l'analisi di Sellars esattamente come una strategia per compiere questa mossa: svelare la natura normativa di enunciati (quelli semantici) che superficialmente appaiono essere descrittivi. Non credo che questa sia una mossa bizzarra. A tal proposito, val la pena spendere qualche parola per presentare un caso non controverso di enunciati comuni che non contengono termini deontici pur potendo essere usati come enunciati normativi. Sellars non sfrutta questa linea argomentativa "per analogia"; ritengo, tuttavia, che la discussione di questi casi rappresenti un'introduzione al trattamento dell'analisi di Sellars che farò nell'ultima sezione, quando mostrerò in che senso si tratta di un argomento indiretto per TNS.

Si considerino gli enunciati (9)-(11):

- (9) I maschi non piangono.  
 (10) Gli amici non abbandonano gli amici.  
 (11) Le bambine non dicono le parolacce.

Gli enunciati (9)-(11) contengono termini generici che fungono da soggetti grammaticali (rispettivamente: “i maschi”, “gli amici” e “le bambine”) seguiti da predicati che non sono predicati modali deontici (rispettivamente: “non piangono”, “non abbandonano gli amici” e “non dicono le parolacce”). Ecco ora altri enunciati che contengono termini generici nella posizione di soggetti grammaticali cui seguono predicati che non sono predicati modali deontici:

- (12) Le tigri non volano.  
 (13) Le porte hanno le maniglie.  
 (14) I politici guadagnano molti soldi.

Gli enunciati (12)-(14) vengono usati per comunicare che, tipicamente o normalmente, le tigri non volano, le porte hanno le maniglie e i politici guadagnano parecchio. Sono cioè enunciati descrittivi o fattuali. Invece, (9)-(11), pur essendo simili a (12)-(14), possono venire usati per due scopi: (i) in certi contesti d’uso, per comunicare che, tipicamente o normalmente, i maschi non piangono, gli amici non abbandonano gli amici e le bambine non dicono le parolacce (e, in questi casi, (9)-(11) funzionano come enunciati descrittivi analoghi a (12)-(14)); (ii) in altri contesti d’uso, per comunicare come le cose devono o non devono essere, cioè per comunicare che i maschi non devono piangere, gli amici non devono abbandonare gli amici, e le bambine non devono dire le parolacce (e, in questi casi, (9)-(11) funzionano come enunciati che hanno, o comunque pretendono di avere, delle conseguenze normative sui nostri comportamenti, e in ciò sono analoghi ad enunciati normativi come (5)-(8)).

Si noti che spesso questi usi normativi di enunciati descrittivi sono enfatizzati dall’aggiunta di alcuni aggettivi applicati ai termini generici che fungono da soggetti grammaticali. Ad esempio, l’uso normativo di (9) viene enfatizzato dicendo “I veri maschi non piangono”, e qualcosa di simile accade anche nel caso di (10) (“I veri amici non abbandonano gli amici”) e di (11) (“Le brave bambine non dicono le parolacce”). Se ci accontentassimo di dire che questi enunciati con l’aggiunta degli aggettivi sono enunciati descrittivi esattamente come (9)-(11), non renderemmo conto né del perché sia sensato e non superfluo o ridondante aggiungere gli aggettivi né di che tipo di modifica questa aggiunta comporti. Si confronti uno qualsiasi di questi enunciati con l’enunciato (13) modificato, “Le vere porte hanno le maniglie”: in questo caso, abbiamo l’intuizione che l’aggiunta dell’aggettivo “vere” sia meno adeguata.

Questo breve excursus è servito solo a mostrare che vi sono alcuni casi intuitivi di enunciati ordinari che non contengono termini deontici e, ciononostante, possono essere utilizzati come enunciati dotati di forza normativa. Ritengo che un fenomeno analogo sia ascrivibile anche agli enunciati semantici, sebbene in questo caso il contesto d'uso non determini la funzione descrittiva o normativa dell'enunciato: enunciati come (1)-(4) sono sempre normativi – o, come metterò in luce più avanti, hanno sempre conseguenze normative sui nostri comportamenti linguistici. Mi rivolgo ora all'analisi di Sellars.

### 3 L'analisi ingenua degli enunciati semantici

Si considerino nuovamente gli enunciati semantici (1) - (4). Questi hanno la seguente forma: “*e* significa *M*”. Sellars denuncia un errore in una certa analisi di enunciati come (1)-(4). Per capire di che errore si tratti, mi concentro su (2), un enunciato semantico in un contesto di traduzione:

(2) ‘Yellow’ (in inglese) significa *giallo*.

L'analisi che Sellars critica (che chiamo *analisi ingenua*) muove dalla somiglianza superficiale che enunciati come (2) presentano con enunciati come (15):

(15) Milano confina con Rho.

Occorre avanzare due osservazioni su (15). La prima è che (15) è un enunciato descrittivo: dice come stanno le cose (in particolare, dice che Milano confina con Rho) ed esprime una proposizione che ha un valore di verità. La seconda osservazione è che (15) è un enunciato relazionale, cioè un enunciato che ha la seguente forma: “*xRy*” (così scrive Sellars) oppure, nella notazione standard della logica predicativa, “*R(xy)*” (io utilizzerò questa notazione). Altri esempi di enunciati relazionali sono i seguenti:

(16') Tizio insegue Caio.

(17') Qualcuno insegue Caio.

(18') Tutti inseguono qualcuno.

Gli enunciati (16)-(18) possono essere formalizzati nel modo seguente:

(16)  $I(tc)$

(17)  $\exists xI(xc)$

(18)  $\forall x\exists yI(xy)$

Dove  $I$  è una costante predicativa a due posti che sta per il predicato “inseguire” e  $t$  e  $c$  sono costanti individuali che stanno per gli individui denotati rispettivamente dai nomi propri “Tizio” e “Caio”.

Si consideri ora (15): si tratta di un enunciato relazionale che non è né esistenziale (a differenza di (17)) né universale (a differenza di (18)); è quindi analogo a (16) e può essere formalizzato così:

(15')  $C(mr)$

dove  $C$  è una costante predicativa a due posti che sta per il predicato “confinare con” e  $m$  e  $r$  sono costanti individuali che stanno per le città denotate rispettivamente dalle espressioni “Milano” e “Rho”.

Ora, l'analisi ingenua propone di trattare (2) alla stregua di un enunciato descrittivo relazionale come (15), dove il predicato non è più “confinare con” ma “significare”. L'analisi ingenua formalizza (2) così:

(2')  $S(yg)$

dove  $S$  è una costante predicativa a due posti che sta per il predicato “significare” e  $y$  e  $g$  sono costanti individuali che stanno per gli oggetti denotati rispettivamente da “Yellow” e “giallo”. Ciò che vi è di interessante e, per Sellars, di inadeguato in questa analisi è che “significare” viene inteso come una relazione di associazione tra un oggetto linguistico (una parola dell'inglese: la parola ‘Yellow’<sup>4</sup>) e un oggetto extralinguistico (il giallo o la qualità cromatica “giallo” o la Giallezza). Secondo l'analisi ingenua, (2) è dunque un enunciato descrittivo relazionale-associativo.

## 4 I problemi dell'analisi ingenua

Che cosa c'è di scorretto in ciò? In EPM Sellars non è esplicito sui motivi che giustificano il rifiuto dell'analisi ingenua, ma credo che la *posizione* occupata dalla sua discussione all'interno di EPM dia qualche indizio per capire che cosa Sellars reputi scorretto in essa.

La discussione dell'analisi ingenua in EPM si trova al paragrafo 31, nella sezione intitolata *The Logic of 'Means'*, la settima del saggio. Questa sezione è immediatamente successiva alla sezione intitolata *Impressions and Ideas: An Historical Point*. Una presentazione analitica delle sezioni che ho nominato non è utile ai fini del problema di cui sto discorrendo: dedico dunque qualche parola esclusivamente a quei punti che sembrano essere rilevanti per capire perché Sellars voglia occuparsi degli enunciati semantici proprio là dove se ne occupa.

<sup>4</sup>Si noti che “Yellow” in (2) compare tra “...” perché l'uso che, in (2), viene fatto del segno “Yellow” è quello di riferirsi a se stesso in quanto segno linguistico, cioè in quanto parola.

Nella sezione *Impressions and Ideas: An Historical Point*, Sellars mette a confronto le posizioni di Locke, Berkeley e Hume sul problema dell'acquisizione della consapevolezza di ripetibili sensoriali (determinabili e determinati)<sup>5</sup>. Secondo Sellars, tutti gli empiristi britannici danno per scontate due idee: (i) l'idea che la mente umana sia naturalmente dotata della capacità non-acquisita (innata) di avere consapevolezza di certi ripetibili sensoriali determinati – “consapevolezza” significa qui “consapevolezza classificatoria”, cioè la capacità di conoscere fatti della forma “ $y^n(x_1, \dots, x_n)$ ” (detto altrimenti: la capacità di sussumere il particolare, ovvero  $x, \dots, x_n$  sotto il generale, ovvero  $y^n$ )<sup>6</sup>; (ii) l'idea che tale capacità non-acquisita coincida con la capacità naturale di avere sensazioni.

Occorre accennare al fatto che questi postulati dell'empirismo classico sono per Sellars tra i responsabili storicamente più noti di un progetto epistemologico fondazionalista, secondo il quale il sistema della conoscenza empirica ha un fondamento costituito da una serie di credenze basilari dotate delle seguenti caratteristiche: (a) ciascuna di esse è conseguita non-inferenzialmente; (b) ciascuna di esse non presuppone altre credenze, fondazionali o inferenziali che siano; (c) ciascuna di esse, insieme alle altre, si pone a fondamento di tutte le credenze empiriche. Si noti che la caratteristica (b) non coincide con la caratteristica (a); di fatto Sellars adotta una posizione che accetta (a) ma rifiuta (b): vi sono credenze conseguite non-inferenzialmente ma queste credenze *presuppongono* necessariamente altre credenze – quelle che servono per maneggiare i concetti rilevanti messi in gioco dalle credenze conseguite non-inferenzialmente. In secondo luogo, si osservi che (a) e (b) conferiscono assoluta *indipendenza epistemica* alle credenze in questione (e, come accennavo, Sellars sostiene che tale indipendenza non sia possibile in linea di principio), mentre (c) garantisce loro *efficacia epistemica*, ossia la capacità di porsi come premesse indubitabili per inferenze le cui conclusioni coincidono con le credenze fattuali che costituiscono

<sup>5</sup>Val la pena chiarire la terminologia di Sellars. La dicotomia ripetibile-irripetibile è un altro modo di esprimere la dicotomia *type-token* (o, in italiano, tipo-occorrenza). Una mela verde ha il suo colore (una sua parte non-indipendente): il verde della mela è un irripetibile collocato spazio-temporalmente e, per questo motivo, non può essere condiviso con altri oggetti distinti da quella mela. Un maglione verde (dello stesso verde della mela) ha il suo colore (una sua parte non-indipendente): il verde del maglione è un irripetibile. Ma vi è un senso secondo cui è corretto dire che la mela e il maglione hanno lo stesso colore, cioè condividono un ripetibile, che non è collocato spazio-temporalmente e che, per questo motivo, può essere condiviso da più oggetti. Ci si può esprimere anche così: il verde della mela è un'occorrenza del tipo “verde”, il verde del maglione è un'altra occorrenza dello stesso tipo “verde”. Alla dicotomia ripetibile-irripetibile si accompagna la dicotomia determinabile- determinato. Questa dicotomia si applica ai ripetibili: un ripetibile A è un determinabile se e solo se esiste un ripetibile B ad esso subordinato; un ripetibile A è un determinato se e solo se non esiste un ripetibile B ad esso subordinato. Un ripetibile B è subordinato ad un ripetibile A se e solo se è possibile che qualcosa sia A senza essere B ed è impossibile che qualcosa sia B senza essere A.

<sup>6</sup>Questa formulazione è tratta da Guardo (2007).

il sistema della conoscenza empirica<sup>7</sup>.

Cosa pensa Sellars dei due postulati dell'empirismo classico e delle caratteristiche (a), (b) e (c) attribuite alle credenze basilari è noto: Sellars riconosce in questo atteggiamento filosofico un episodio diffuso del Mito del Dato, l'episodio epistemologico, in base al quale vi è, nell'esperienza, qualcosa di *dato* (ad esempio, per gli empiristi, le sensazioni - o, come si esprime Sellars, la consapevolezza di ripetibili sensoriali determinati) che, pur non essendo concettualmente strutturato, è in grado di fondare il sistema della conoscenza empirica, che evidentemente è concettualmente strutturato. Per questioni di spazio, non posso riprendere qui l'argomento sellarsiano contro il Mito del Dato; ma il cuore dell'argomento è questo: l'errore dell'empirismo e di qualsiasi epistemologia fondazionalista è la postulazione di un *dato originario*, come la sensazione, a partire dal quale l'edificio della conoscenza può e deve fondarsi; si tratta di un errore perché un dato originario così inteso non è dotato di struttura proposizionale, e dunque non è possibile che esso sia in grado di fondare un sistema di conoscenza che invece è proposizionalmente strutturato. L'idea portante dell'argomento di Sellars è che le esperienze sensoriali sono semplici *eventi* (plausibilmente, eventi mentali di qualche tipo) e non sono dunque strutturate proposizionalmente; in quanto tali, le esperienze sensoriali non possono avere o trasmettere uno status epistemico positivo ad una qualche credenza empirica a fondamento della nostra conoscenza (Sellars assume che soltanto qualcosa dotato di forma proposizionale possa avere o trasmettere un tale status).

Infine, è utile menzionare il fatto che Sellars abbozza, nell'ultimo paragrafo della sezione, una posizione che chiama "nominalismo psicologico" (che giocherà un ruolo centrale nella sua analisi degli enunciati semantici). Secondo questa posizione, la consapevolezza dei ripetibili è una faccenda linguistica; come tale, essa non è data, ma presuppone un linguaggio e un significativo apparato concettuale. In EPM Sellars non argomenta approfonditamente a favore di questa posizione (e può permettersi di non farlo perché, in questa sezione - e in gran parte del saggio - il suo avversario principale è il sostenitore dell'empirismo classico, con il quale condivide le "tendenze prevalentemente nominalistiche della tradizione empiristica" (Sellars 1956, §6). Ad ogni modo, l'idea generale è la seguente: aderendo alla prospettiva *lato sensu* wittgensteiniana e sviluppandola originalmente, Sellars sottoscrive la tesi secondo la quale possedere un concetto equivale a conoscere il significato che una parola assume nelle quotidiane pratiche linguistiche; a sua volta, conoscere il significato di una parola implica la capacità di padroneggiare le basilari regole d'uso di un (segmento di) linguaggio, regole che Sellars invita ad analizzare nei termini di processi inferenziali (Sellars 1969, 1974).

Quanto detto circa la sezione *Impressions and Ideas: An Historical Point* può

<sup>7</sup>Questa è la ricostruzione dell'argomento formulata in deVries e Triplett (2000).

bastare. La sezione successiva è *The Logic of 'Means'*. Questa è costituita di due paragrafi, il 30 e il 31. Il secondo è quello in cui viene condotta l'analisi degli enunciati semantici. Nel primo paragrafo, invece, viene presentata la versione linguistica del Mito del Dato. Riporto il brano per intero, per mettere in luce che ciò di cui Sellars sta parlando è un presupposto teorico che fa da sfondo alla versione epistemologica del Mito che è stata criticata nella sezione precedente:

[...] when we picture a child [...] learning his first language, we, of course, locate the language learner in a structured logical space in which we are at home. Thus, we conceive him as a person (or, at least, a potential person) in a world of physical objects, coloured, producing sounds, existing in Space and Time. But though it is we who are familiar with this logical space, we run in danger, if we are not careful, of picturing the language learner as having ab initio some degree of awareness - "pre-analytic", limited and fragmentary though it may be - of this same logical space.

(Sellars 1956, §30)

Sellars vuole problematizzare quella che Wittgenstein, nei primi paragrafi delle *Philosophical Investigations*, chiama *immagine agostiniana* del linguaggio, secondo la quale (i) il linguaggio consta anzitutto di nomi, e (ii) questi sono assimilabili ad etichette per i particolari (individui, oggetti, eventi, ecc.). Conseguentemente, l'apprendimento linguistico coincide con l'apprendimento della relazione tra certe etichette linguistiche e certi particolari, e avviene, tipicamente, per via ostensiva.

Sellars (sulla scia di Wittgenstein) è un severo critico di questo genere di resoconti. Sellars volge la propria attenzione al fatto che, secondo questa concezione, il soggetto che impara un linguaggio si trova già (ed è consapevole di trovarsi) in uno spazio logico strutturato in maniera del tutto familiare, uno spazio logico costituito anzitutto di particolari, ma anche di universali, fatti o, più in generale, elementi già chiaramente discriminati e classificati. Pertanto, l'apprendimento linguistico consiste, secondo TRS, in una prassi di etichettatura (via via più complessa e precisa) attraverso simboli verbali di ciò che si dà già come discriminato e classificato. Si tratta, in definitiva, della versione linguistica del Mito del Dato, secondo la quale il discente, nel momento in cui si appresta ad apprendere un linguaggio, è già dotato di una struttura concettuale (per quanto rozza e primitiva) che gli permette di avere quel tipo di consapevolezza classificatoria che gli empiristi britannici consideravano una capacità, appunto, naturale e innata della mente umana.

Queste poche nozioni su due sezioni di "Empiricism and the Philosophy of Mind" sono sufficienti per rispondere alla seguente domanda: perché, al paragrafo 31, Sellars rifiuta l'analisi ingenua degli asserti semantici? Credo che ci siano due risposte che possono essere date.

La prima risposta è la seguente: l'analisi ingenua propone di trattare (e dunque di formalizzare) gli enunciati semantici come enunciati descrittivi relazionali-associativi proprio perché essa è uno degli esiti del Mito del Dato nelle sue versioni epistemologica e linguistica. Chi casca nella trappola del Mito è portato a sostenere che l'apprendimento del significato di una parola debba essere identificato con l'associare ad un certo simbolo qualcosa di cui si ha già consapevolezza (e così le due versioni del Mito si fondono in un unico errore). Ed è questo di fatto il modo ingenuo di intendere enunciati semantici come (2): come enunciati esprimenti una relazione di associazione tra un oggetto linguistico (una parola in una qualche lingua) e un oggetto extralinguistico che è dato.

La seconda risposta dipende da una conseguenza del nominalismo psicologico di Sellars. Si assuma per ipotesi che l'analisi ingenua di (2) sia corretta: 'Yellow' è una parola che è associata alla qualità cromatica "giallo" o alla Giallezza (non è importante qui scegliere quale sia precisamente l'entità cui l'etichetta si applica) e la formalizzazione di (2) è, come spiegato sopra, (2'). Se così stanno le cose per (2) e se si desidera mantenere una certa uniformità nell'analisi degli enunciati semantici, allora si deve fornire la medesima analisi per tutti gli enunciati semantici, dunque anche per (1), (3) e (4). La formalizzazione ingenua di questi enunciati ricalca la formalizzazione ingenua di (2):

$$(1') S(g_1g)$$

$$(3') S(e_1e)$$

$$(4') S(f_1f)$$

dove  $g_1$ ,  $g$ ,  $e_1$ ,  $e$ ,  $f_1$  e  $f$  sono costanti individuali che stanno per gli oggetti denotati rispettivamente da "Giallo", "giallo", "E", "e", "Fenice", "fenice". (1') dice che 'Giallo' è una parola che è associata alla qualità cromatica "giallo" o alla Giallezza; (3') e (4') dicono, rispettivamente, che 'E' e 'Fenice' sono due parole dell'italiano che stanno per due oggetti distinti. Ma che tipi di oggetti sono?

Per poter rispondere, occorre riprendere rapidamente il nominalismo psicologico sellarsiano e chiarirne due punti. Anzitutto, si chiama *nominalismo* perché è una posizione che nega che i fenomeni cognitivi coinvolgono una qualche relazione diretta con oggetti astratti (a prescindere dal loro essere intesi come oggetti dipendenti o indipendenti dalle menti); cerca quindi di sostituire questa relazione diretta con oggetti astratti con una relazione diretta con oggetti *linguistici*. In secondo luogo, è un nominalismo *psicologico* cioè non è un nominalismo ontologico; un nominalismo ontologico afferma che gli oggetti astratti non esistono; quello di Sellars è un nominalismo secondo cui gli oggetti astratti non sono direttamente coinvolti nei fenomeni psicologici, in particolare in quelli cognitivi. In questo senso, il nominalismo psicologico implica la falsità di tutte le posizioni epistemologiche che, in un modo o nell'altro, sono vittime del Mito del Dato.

Se questi due punti sono chiari, è semplice capire che non è possibile abbracciare congiuntamente il nominalismo psicologico e l'analisi ingenua degli enunciati semantici. Discutendo di (3') e (4'), avevo formulato la domanda "che tipi di oggetti sono gli oggetti cui si applicano le etichette 'E' e 'Fenice' secondo l'analisi ingenua?". La risposta a questa domanda dipende strettamente dal rifiuto o dall'accettazione del nominalismo psicologico. Se si rifiuta il nominalismo psicologico, allora la risposta è: gli oggetti cui si applicano 'E' e 'Fenice' sono oggetti diversi dagli oggetti concreti ma sono pur sempre oggetti (ad esempio: oggetti astratti) che sono direttamente coinvolti in certi fenomeni cognitivi (ad esempio sono direttamente coinvolti quando penso alla congiunzione come connettivo logico vero-funzionale e alla sua tavola di verità - è il caso di 'E' - oppure quando racconto una storia sul celebre uccello immaginario e mi trovo quindi in un certo stato mentale che verte su questo oggetto immaginario - è il caso di 'Fenice'); il problema di questa soluzione è che, secondo Sellars, rifiutare il nominalismo psicologico implica accettare il Mito del Dato; ma accettare il Mito del Dato ha un costo troppo alto, perché impegna ad un'epistemologia fondazionalista che esce sconfitta dall'argomentazione di Sellars in EPM. Se, al contrario, si accetta il nominalismo psicologico, allora la risposta è: gli oggetti cui si applicano 'E' e 'Fenice' sono oggetti linguistici, non oggetti astratti o immaginari con i quali si possa avere una relazione diretta di consapevolezza; com'è ovvio, questa soluzione è incompatibile con l'analisi ingenua, che, cedendo alle lusinghe del Mito del Dato, invita invece a riconoscere che gli oggetti cui si applicano le etichette linguistiche non sono a loro volta etichette linguistiche, ma oggetti extralinguistici di qualche tipo dei quali possiamo naturalmente avere consapevolezza diretta e pre-linguistica.

In definitiva, o si accetta il nominalismo psicologico o si adotta l'analisi ingenua degli enunciati semantici. Sellars non è disposto a rinunciare al nominalismo psicologico (poiché rinunciarvi implicherebbe tornare al Mito del Dato) e quindi rifiuta l'analisi ingenua. È ovvio che, così come il rifiuto dell'analisi ingenua dipende dal rifiuto del Mito del Dato, allo stesso modo l'analisi proposta da Sellars dipende dall'assunzione del nominalismo psicologico.

## 5 La proposta di Sellars

Rimane ora da presentare la proposta di Sellars per l'analisi degli enunciati semantici. La sua proposta è quella di considerare "significa" come una "forma specializzata di copula" (Sellars 1974) ("è") in grado di esprimere una relazione d'identità. Considero a titolo esemplificativo ancora (2). Secondo Sellars, (2) è analizzabile nei termini di un enunciato d'identità:

(2") 'Yellow' = giallo.

Si noti che, da un punto di vista logico, (2'') è un enunciato analogo ai seguenti enunciati d'identità (o alla loro formalizzazione):

(19) Espero = Fosforo

(20) Espero = Marte

(21) Acqua =  $H_2O$

Il valore di verità della proposizione espressa da un enunciato d'identità si stabilisce verificando che l'oggetto denotato dall'espressione linguistica a destra del simbolo d'identità sia o non sia lo stesso oggetto denotato dall'espressione linguistica a sinistra del simbolo d'identità: se lo è, la proposizione è vera, mentre se non lo è, essa è falsa. (20) è un esempio di enunciato d'identità esprimente una proposizione falsa (gli altri enunciati della lista esprimono proposizioni vere).

Quella che ho schizzato è un'immagine molto povera delle condizioni di verità degli enunciati d'identità (un'immagine che, ad esempio, non tiene conto del variare del valore di verità al variare del mondo possibile dal quale gli enunciati vengono valutati), ma credo sia sufficiente per procedere verso l'analisi degli enunciati semantici. Se (2'') è un enunciato d'identità, allora le sue condizioni di verità sono del tutto analoghe alle condizioni di verità di (19)-(21). Vale a dire: se (2'') è un enunciato d'identità, allora esso è vero se e solo se l'oggetto denotato dall'espressione linguistica a destra del simbolo d'identità è lo stesso oggetto denotato dall'espressione linguistica a sinistra del simbolo d'identità. Che tipi di oggetti sono quelli denotati dalle espressioni in (2'')? Si potrebbe essere tentati di rispondere così: l'espressione a sinistra del simbolo d'identità ("Yellow") denota un oggetto linguistico (la parola 'Yellow') e l'espressione a destra del simbolo d'identità ("giallo") denota un oggetto extralinguistico (la qualità cromatica "giallo" o la Giallezza). In EPM questa opzione non viene presa in considerazione per ovvi motivi: essa rappresenta una nuova ricaduta nel Mito del Dato (con l'unica differenza che "significa" non è più inteso come una relazione associativa bensì come una relazione d'identità). Aggiungerei che l'opzione di trattare gli oggetti di cui si predica l'identità in (2'') come un oggetto linguistico ed uno extralinguistico è poco attraente al di là dell'eventuale ricaduta nel Mito del Dato: infatti, dire che la parola 'Yellow' è identica alla qualità cromatica "giallo" (o alla Giallezza) è semplicemente falso. Per un motivo o per l'altro, non è questa la via da seguire.

La proposta di Sellars è la seguente: (2'') è un enunciato che esprime l'identità tra *ruoli funzionali* delle espressioni linguistiche. Quindi, (2'') è vero se e solo se il ruolo funzionale dell'espressione linguistica a destra del simbolo d'identità è lo stesso ruolo funzionale dell'espressione linguistica a sinistra del simbolo d'identità. Che cosa sia un ruolo funzionale è presto detto: il ruolo funzionale dell'espressione linguistica  $e$  è il ruolo svolto da  $e$  in un sistema-linguaggio  $L$

(l'italiano, l'inglese, ecc.), ovvero l'uso (o gli usi) che di *e* è consentito fare in *L*. Questa definizione non rende giustizia alla complessità del concetto coniato da Sellars in "Meaning as Functional Classification. A Perspective on the Relation of Syntax to Semantics" (1974) (MFC). Tuttavia, per questioni di spazio, non posso che dedicarvi solo qualche battuta.

Introducendo il nominalismo psicologico, ho fatto un rapido riferimento al trattamento sellarsiano dei concetti dicendo che, per Sellars, possedere un concetto equivale a conoscere il significato di una parola; a sua volta, conoscere il significato di una parola implica la capacità di padroneggiare le regole d'uso di un linguaggio, regole che Sellars invita ad analizzare nei termini di processi inferenziali. In questo senso, conoscere il contenuto del concetto di "rosso" comporta sapere in quale rete di concetti esso si colloca e quale posizione ha in essa. In MFC, Sellars adotta una terminologia "funzionalistica" ma la sua posizione rimane pressoché invariata: la capacità di sviluppare inferenze (o di percorrere la rete inferenziale che connette i concetti) è l'abilità di classificare il ruolo funzionale di un'espressione linguistica entro un certo sistema-linguaggio (per esempio quello dell'italiano). Ancora una volta, può essere utile accennare alla questione dell'apprendimento linguistico: quando il discente entra nel sistema-linguaggio, progressivamente impara a conoscere le diverse funzioni dei giochi linguistici e, contemporaneamente, apprende la facoltà di selezionare, di volta in volta, la funzione che più risulta idonea nel contesto linguistico(-conversazionale) nel quale si trova a muoversi:

To say what a person says, or, more generally, to say what a kind of utterance says, is to give a functional classification of the utterance.

(Sellars 1974, p. 419)

Padroneggiare i significati delle espressioni linguistiche significa saper cogliere le funzioni che quelle espressioni svolgono in un certo sistema-linguaggio: questa è, in fondo, la ragione teorica che sottende il tipo di analisi degli enunciati semantici proposta da Sellars.

Per riassumere: l'analisi sellarsiana consiste nel sostenere che (2) ha la forma logica di un enunciato d'identità, l'enunciato (2"), che predica l'identità tra i ruoli funzionali delle due espressioni linguistiche che compaiono a sinistra e a destra del simbolo d'identità. In altri termini: (2) dice che le funzioni, gli usi che l'espressione linguistica 'Yellow' svolge nella lingua inglese sono assolti, in italiano, dall'espressione linguistica 'Giallo'. Con ciò, non è ancora stato asserito nulla circa tali usi: (2) non necessariamente dice che l'espressione inglese 'Yellow' sta in una relazione di associazione con la qualità cromatica "giallo" o la Giallezza. Il che significa sostenere che solo dalla verità della proposizione espressa da (2) non è possibile inferire alcunché circa la complessità del ruolo svolto dall'espressione 'Yellow' in inglese o circa il modo preciso in cui essa è

correlata alla qualità cromatica “giallo” o alla Giallezza. Nella prossima sezione, abbozzerò una spiegazione della relazione tra il significato di (2) (e (2’)) e il fatto che ‘Yellow’ si riferisce alla qualità cromatica “giallo” o alla Giallezza.

Sellars adotta due dispositivi tecnici per formalizzare enunciati come (2): le “*asterisk quotes*” e le “*dot quotes*” (Sellars 1974). Ecco come viene formalizzato (2):

(2+) \*yellow\* significa •giallo•

Oppure, mantenendo il simbolo d’identità:

(2+’) \*yellow\* = •giallo•

(2+) e (2+’) esprimono il fatto che i ruoli funzionali equiparati sono relativi l’uno – il primo – alla lingua che, in (2), svolge la funzione di lingua oggetto (l’inglese), e l’altro - il secondo - alla lingua che, in (2), svolge la funzione di metalinguaggio (la lingua cioè in cui (2) è formulato, l’italiano).

## 6 Dalla proposta di Sellars alla tesi della normatività

Propongo ora alcune osservazioni. La prima è volta a risolvere un problema che potrebbe sorgere dall’analisi di Sellars. Nella seconda osservazione, considero la nozione di *pragmema* (introdotta da Mey (2001) e sviluppata da Capone (2005, 2017, 2018)) e la utilizzo per argomentare che il significato dei pragmemi esemplifica una nozione ampliata di ruolo funzionale. La terza osservazione, che nasce come una riflessione sulla nozione di metalinguaggio, vuole mostrare in che senso l’analisi di Sellars sia interpretabile come un’analisi che parla a favore della tesi della normatività del significato (TNS). Concludo con una rapida discussione del modo in cui, dalle osservazioni di Sellars, sia possibile passare alla formulazione dell’idea secondo la quale le espressioni linguistiche hanno condizioni di applicazione corretta. Mostro quindi che questa idea è al centro del dibattito attuale su TNS.

Ecco la prima osservazione. Riconosco che c’è un senso in cui sembra corretto dire che (2) significa che la parola ‘Yellow’ è usata in inglese per riferirsi alla qualità cromatica “giallo” o alla Giallezza. L’analisi di Sellars nega questo. Come si può sanare questa discrepanza tra ciò che suggerisce l’intuizione e ciò che insegna l’analisi? Credo così: nel caso di (2), essere parlanti competenti della lingua in cui l’enunciato è formulato (l’italiano) è condizione necessaria per poter dedurre inferenzialmente quale sia l’effettivo ruolo funzionale svolto dal termine su cui verte l’enunciato semantico (il termine inglese ‘Yellow’). In sostanza, il fatto che (2) dica che la parola ‘Yellow’ sta per la qualità cromatica “Giallo” o per la Giallezza è la conclusione di un ragionamento inferenziale, non il significato di (2). Il ragionamento inferenziale potrebbe essere qualcosa di questo genere:

(2) predica l'identità di due ruoli funzionali, quello di 'Yellow' in inglese, e quello di 'Giallo' in italiano; io, parlante competente dell'italiano, so che 'Giallo' è un termine osservativo che significa questo e questo, cioè conosco il ruolo funzionale svolto da 'Giallo' in italiano (o, il che è lo stesso per Sellars, comprendo il termine 'Giallo' perché so tracciare un certo numero di percorsi inferenziali "autorizzati" a partire dal termine 'Giallo' – ad esempio: so che si tratta di un termine di un colore, e che quindi non può essere correttamente utilizzato per predicare una proprietà di un suono o di un odore; so anche che se qualcosa è giallo, allora non può assumere certe gradazioni cromatiche; conosco molte cose appartenenti a quella che Wittgenstein avrebbe chiamato la *grammatica* del termine 'Giallo'); da tutto ciò riesco ad inferire i casi in cui un parlante competente dell'inglese è autorizzato a utilizzare il termine 'Yellow', e quindi associo il termine 'Yellow' alla qualità cromatica "giallo" o alla Giallezza. Tale associazione non è parte del significato di (2) ma è implicata dal significato di (2).

Ovviamente, un punto cruciale in questo ragionamento è svolto dal fatto che essere parlanti competenti della lingua in cui l'enunciato semantico è formulato sia condizione necessaria per dedurre inferenzialmente il ruolo funzionale dell'espressione in questione. Questo punto può essere ulteriormente precisato. Sostengo che una regola implicita nei contesti tipici nei quali vengono proferiti enunciati semantici è che sia il parlante che l'ascoltatore devono conoscere il significato del termine oggetto dell'enunciato semantico proferito (in (2) il significato di 'giallo'). In sostanza, fa parte del *common ground*<sup>8</sup> dei partecipanti a una conversazione il fatto che i partecipanti conoscano (almeno sommariamente) il significato dell'espressione linguistica che viene utilizzata per spiegare il significato di un'altra espressione linguistica.

La seconda osservazione verte sulla nozione di *pragmema*: intendo sostenere che il significato dei pragmemi rappresenta un esempio paradigmatico di ruolo funzionale *ampliato*. Per spiegare perché, occorre anzitutto introdurre la nozione di *pragmema*.

Secondo Capone (2005, 2018), un *pragmema* è un'*unità linguistica situata* nella quale le regole del linguaggio, della società e della cultura si combinano per determinarne il significato<sup>9</sup>. Di seguito riporto una definizione efficace della nozione di *pragmema*, che rende conto della sua peculiarità rispetto a nozioni più tradizionali come quelle di atto linguistico e di gioco linguistico:

A *pragmeme* is an utterance requiring various degrees of enrichment. First of all, we need completions and saturation processes in order to fix the reference of pronominals or other referential expressions [...] Then pragmatic increments are needed which are determined/affected by cultural norms [...]

<sup>8</sup>Stalnaker (2002).

<sup>9</sup>Per un approfondimento della tematica del rapporto tra *pragmema* e cultura si vedano in particolare Wong (2010) e Capone (2018).

Finally, we also need to take into account increments due to the influence of the local co-text. [...] Further information is needed about the peculiarity of pragmemes with respect to traditional notions such as the speech act in context or the language game. Two further forms of embedding need to be added: embedding in a cultural context [...] and, finally, embedding in expectations about the rationality of the speakers [...]

(Capone 2018, pp. 91-92)

A ciò va aggiunta una precisazione fondamentale: una caratteristica importante del pragmemma è che non è necessario che esso coincida con un singolo proferimento linguistico; esso può coincidere (e di fatto spesso coincide) con unità linguistiche complesse e strutturate, come un intero (frammento di) discorso<sup>10</sup>.

Uno degli esempi di pragmemma più persuasivi (discusso da Capone (2005, 2017, 2018)) è quello rappresentato dagli scambi comunicativi nel contesto dell'interrogazione in una classe italiana. Tipicamente, l'insegnante seleziona un'alunna per l'interrogazione proferendo il comando "Vieni". Un'interpretazione letterale del comando non sembra catturare il suo reale contenuto prescrittivo: per eseguire il comando, non è sufficiente che l'alunna si avvicini alla cattedra; di fatto, non è questo ciò che è primariamente inteso dall'insegnante (anche se ciò potrebbe essere parte di ciò che è inteso). Ciò che l'insegnante primariamente intende comunicare è che vuole interrogare l'alunna. Ella può rispondere affermativamente, dicendo "Ok, vengo", o negativamente, dicendo "No, non vengo"; nel secondo caso, l'insegnante potrà chiedere "Non hai studiato?" e l'alunna potrà scegliere tra diverse opzioni di risposta, tra le quali val la pena di menzionare "Ho letto"; a proposito di questa risposta, Capone scrive:

[i]t is not quite clear what to make of this utterance, as this may be interpreted as "I have not studied it properly" or as "I have tried my best to understand". Inferential pragmatics does not really come to our aid here. [...] One could infer from the usage of the weaker item ('read') that the person in question has not really studied the passage in question, let alone learned anything [...] But this is not what we should infer in this context. What we are supposed to understand is that the student has made an effort to study the lesson, but she is not completely sure that she has grasped it.

(Capone 2005, pp. 1361-1362)

Il punto su cui Capone vuole attirare l'attenzione è che questi proferimenti vanno intesi anzitutto come frammenti di una certa cultura (quella scolastica italiana contemporanea), nella quale il linguaggio è situato, e la quale a sua volta trasmette il linguaggio e i suoi usi. L'idea fondamentale è che questi atti linguistici non possono essere studiati in maniera isolata dal contesto di proferimento, dove questo non coincide semplicemente con un certo luogo fisico (l'aula

<sup>10</sup>Si veda Capone (2017).

scolastica), ma comprende una situazione più ampia in cui rientrano convenzioni sociali e culturali, regole implicite condivise, nonché abitudini di comportamento consolidate (quelle abitudini che contribuiscono a costituire la pratica dell'interrogazione scolastica).

Vorrei ora argomentare che la nozione di pragmemma può essere utile nell'analisi degli enunciati semantici: essa può rappresentare un modo per chiarire e ampliare la nozione sellarsiana di ruolo funzionale.

Come spiegato sopra, il pragmemma è un'unità linguistica (semplice o complessa) il cui significato è determinato in maniera sostanziale dal contesto arricchito entro il quale l'unità linguistica compare. Si consideri l'esempio proposto da Capone, il proferimento di "Vieni" nel contesto dell'aula scolastica italiana. In questo caso, sembra evidente che il significato di quel proferimento coincide (almeno in parte) con il *ruolo* che quel proferimento gioca nella conversazione tra insegnante e alunna in quel contesto particolare che viene a costituirsi nella pratica "istituzionalizzata" dell'interrogazione. Sostengo che i casi di pragmemma considerati da Capone esemplificano con molta chiarezza unità linguistiche caratterizzate da una tipologia *arricchita* di ruolo funzionale. A prova di ciò, se consideriamo i pragmemmi, il ruolo funzionale non coincide semplicemente con gli usi ammessi di un'unità linguistica entro un certo sistema linguistico, ma include anche le modalità d'uso peculiari di quell'unità che sono sancite non tanto dal suo significato letterale, quanto piuttosto dal contesto sociale e culturale entro cui tale unità linguistica viene di volta in volta utilizzata. Ad esempio, se consideriamo il proferimento di "Vieni" nell'esempio di Capone, sembra adeguato specificarne il ruolo funzionale sottolineandone l'uso concreto che ne viene fatto entro il contesto sociale della lezione, e specificando le implicazioni che ha nella conversazione e, più in generale, nella strutturazione della pratica stessa dell'interrogazione scolastica.

In sintesi, il ruolo funzionale può essere inteso sia in un senso ristretto sia in uno allargato. Nel senso ristretto, esso si identifica con il ruolo inferenziale rivestito da una certa espressione o unità linguistica entro un certo sistema linguistico, come spiegato sopra. Nel senso allargato, esso va ad includere anche il significato che quell'espressione o unità linguistica assume entro un contesto che è caratterizzato da canoni, convenzioni sociali e culturali, regole implicite condivise, e abitudini di comportamento (linguistico e non-linguistico) consolidate. Indubbiamente, quando vogliamo specificare il significato di un pragmemma dobbiamo far riferimento alla nozione allargata di ruolo funzionale, una nozione molto feconda per rendere conto del significato di molteplici parti del discorso. Pertanto, la nozione di pragmemma amplia il modo in cui possiamo intendere la nozione sellarsiana di ruolo funzionale.

La terza osservazione verte sul concetto di metalinguaggio. Credo che l'analisi di Sellars di (2), e soprattutto la sua formalizzazione attraverso le *asterisk quo-*

*tes* e le *dot quotes*, metta in luce un aspetto interessante degli enunciati semantici: essi non sono enunciati equiparabili agli ordinari enunciati nel linguaggio (enunciati come (12)-(14), che sono i casi paradigmatici di enunciati descrittivi), perché sono enunciati che mettono in relazione due diversi livelli linguistici, cioè il livello del linguaggio oggetto (in (2), l'inglese) e il livello del metalinguaggio (in (2), l'italiano). Ovviamente, non è necessario che il linguaggio oggetto e il metalinguaggio siano diversi; in enunciati semantici come (1), essi coincidono, e infatti non risultano particolarmente informativi. Tuttavia, nella loro forma logica, essi sono analoghi a enunciati semantici dove il linguaggio oggetto e il metalinguaggio sono distinti. Ad ogni modo, credo che gli enunciati semantici, per queste ragioni, siano enunciati che non hanno un loro posto *nel* linguaggio, perché sono enunciati *sul* linguaggio: sono cioè enunciati *metalinguistici* (ancora, non è rilevante *quale* sia il metalinguaggio in cui li esprimiamo). Il cuore teorico di questa tesi è che gli enunciati semantici che vertono su espressioni di una certa lingua naturale non sono enunciati ordinari di quella lingua naturale, ma sono enunciati su quella lingua naturale; la tesi è dunque banale, ma per essere sottoscritta richiede l'abbandono dell'analisi ingenua, che, di primo acchito, può sembrare più convincente, poiché associa a espressioni linguistiche oggetti, proprietà, entità o altri pezzi di mondo. In una lettera a Roderick Chisholm<sup>11</sup>, Sellars scrive che gli enunciati semantici rappresentano una modalità del discorso ordinario che va tenuta distinta dalla mera descrizione e spiegazione di fatti empirici, poiché fa implicitamente appello ad un vocabolario prescrittivo. L'idea di Sellars è esattamente quella espressa dalla sua analisi degli enunciati semantici: essi non vanno intesi come enunciati che stabiliscono relazioni tra espressioni linguistiche e oggetti o proprietà, ma come classificazioni delle funzioni delle espressioni stesse in uno o più sistemi linguistici; ovviamente, tali funzioni possono dipendere dall'esistenza di associazioni del tipo *world-word*, ma tali enunciati non esprimono direttamente tali associazioni.

Un buon modo per chiarire questa tesi è riflettere sul ruolo logico dell'operatore "significa". Questo operatore in (2) non va inteso né come un qualsiasi predicato in un ordinario enunciato nel linguaggio (ad esempio, "confina con") né come un "operatore agostiniano" che mette in relazione etichette a oggetti, ma come un operatore *metalinguistico decitazionale* che ha il compito di esplicitare i ruoli funzionali delle parole, gli usi delle parole che sono permessi. E se ciò è vero, allora è chiaro in che senso questa analisi degli enunciati semantici rappresenti un'analisi a favore della tesi della normatività del significato, dalla quale ho preso le mosse. Lo ricordo: TNS dice che i significati sono essenzialmente normativi, cioè che per essi sono essenzialmente in vigore delle norme che stabiliscono l'uso corretto delle corrispondenti espressioni linguistiche. Nella terminologia di Sellars, il ruolo funzionale di un'espressione linguistica è

<sup>11</sup>Chisholm e Sellars (1958, p. 527).

esattamente l'uso che di un'espressione linguistica è *consentito* fare in un certo sistema-linguaggio, l'uso *autorizzato* o *corretto*; la nozione di ruolo funzionale è dunque una nozione eminentemente normativa. Tuttavia, sarebbe forse una mossa forzata sostenere che, alla luce di ciò, la conclusione che si dovrebbe trarre dall'analisi di Sellars è che (2) è un enunciato normativo *tout court*. Preferisco proporre un'interpretazione meno impegnativa: la conclusione che si deve trarre dall'analisi di Sellars è che (2) è un enunciato metalinguistico che ha delle conseguenze normative sui nostri comportamenti linguistici, ma in sé non è un enunciato normativo analogo a (5):

(5) Devi mangiare la minestra.

Credo che la differenza tra (5) e (2) stia nel seguente fatto. Per quanto riguarda (5), non è in linea di principio possibile dire se l'enunciato è vero o falso; anzi, è del tutto insensato domandarsi se sia vero o falso: esso veicola un contenuto prescrittivo che può essere soddisfatto o meno dalle azioni dell'interlocutore. Invece, nel caso di (2), è sensato interrogarsi circa il suo valore di verità ed è possibile attribuirgliene uno. A prova di ciò, posso formulare intenzionalmente enunciati semantici *falsi* per ingannare qualcuno, dicendo ad esempio che 'Yellow' in inglese significa *nero*; formulerei così un enunciato metalinguistico che esprimerebbe una falsità sul linguaggio, cioè predicherebbe un'identità inesistente tra due ruoli funzionali di termini appartenenti a lingue diverse, un'identità che potrebbe convincere qualcuno dell'esistenza di una certa regola nella lingua inglese secondo la quale la parola 'Yellow' si applica correttamente al colore delle lettere di questo testo.

In questo senso credo ci sia una certa analogia tra gli enunciati semantici e gli enunciati comuni che ho analizzato inizialmente, enunciati come "I (veri) maschi non piangono": (i) i primi non hanno la forma degli enunciati normativi e non sono nemmeno enunciati descrittivi ordinari, bensì sono enunciati metalinguistici che hanno conseguenze normative; (ii) i secondi non hanno la forma degli enunciati normativi e, in determinati contesti d'uso, non sono equiparabili agli ordinari enunciati descrittivi, e, in tali contesti, hanno delle conseguenze normative. Com'è ovvio, si tratta solo di un'analogia: gli enunciati del secondo tipo non sono infatti enunciati metalinguistici, e in questo senso si distinguono dagli enunciati semantici; gli enunciati semantici, inoltre, hanno *sempre* delle conseguenze normative per i parlanti, mentre non può essere detto lo stesso degli enunciati del secondo tipo, che, come mostrato sopra, talvolta hanno un uso meramente descrittivo.

Usando l'analisi di Sellars a sostegno della posizione secondo cui un enunciato semantico è un enunciato metalinguistico che ha delle conseguenze normative sui comportamenti linguistici dei parlanti, intendo sottolineare l'idea che le espressioni linguistiche hanno delle condizioni di applicazione corretta.

Si consideri TNS (una formulazione standard nel dibattito contemporaneo sulla normatività). Che cosa sia la norma cui TNS fa riferimento può essere oggetto di dibattito. La risposta più intuitiva è esattamente quella cui l'analisi di Sellars conduce: una norma semantica in vigore per l'uso di  $e$  dipende strettamente dalle condizioni di applicazione corretta associate ad  $e$ . Si noti che dire che  $e$  ha un ruolo funzionale è equivalente a dire che  $e$  ha delle specifiche condizioni di applicazione corretta, ha un ruolo nel linguaggio che consente certi usi e certe applicazioni e ne vieta altri ( $e$ , in questo senso, l'analisi sellarsiana di (2) rende esplicita l'identità delle condizioni di applicazione corretta di 'Yellow' in inglese e di 'Giallo' in italiano). Si può esprimere questa idea nel seguente modo:

**CAP** Un'espressione linguistica  $e$  significa  $M$  per un parlante  $S$  al tempo  $t$  solo se per ogni  $x$ ,  $e$  è applicata correttamente da  $S$  a  $x$  se e solo se si dà il caso che  $x$  sia  $f$ .

Dove CAP sta per condizioni di applicazione corretta e  $f$  è la caratteristica o l'aspetto in virtù del quale  $e$  si applica a  $x$ . Ad esempio, l'espressione "cat" si applica correttamente ad un oggetto  $x$  se e solo se quell'oggetto è un gatto (ovvero, se e solo se ha certe caratteristiche fisiche tipiche dei gatti); la parola "acqua" si applica correttamente ad un oggetto  $x$  se e solo se quell'oggetto è  $H_2O$  – e non se, ad esempio, quell'oggetto conserva le caratteristiche fenomeniche dell'acqua, ma ha una struttura chimica diversa da  $H_2O$ . Come sottolinea Whiting (2007, 2009), quella che ho chiamato CAP esprime la seguente idea: per un'espressione linguistica, l'aver un determinato significato implica l'aver delle determinate condizioni di applicazione corretta. Di qui, è facile metter capo a TNS: TNS appare essere un'ovvia conseguenza di CAP, se si accetta che la nozione di "applicazione corretta" che compare in CAP sia una nozione normativa. Questo è, in sostanza, un argomento semplice e diretto a favore di TNS che, in modi diversi, viene avanzato da molti attuali sostenitori di una qualche versione di TNS (come Boghossian (1989), Kripke (1982) e Whiting (2007, 2009)). Questo argomento invita a compiere una mossa intuitivamente accettabile: far discendere dalla verità di CAP la verità di TNS in virtù della normatività della nozione di applicazione corretta o, più semplicemente, della nozione di correttezza.

Occorre menzionare che l'argomento non è esente da critiche. Nonostante sembri ragionevole sostenere che una qualche nozione di correttezza sia parte essenziale del concetto stesso di significato, vi sono tentativi di falsificare questa posizione (si vedano Glüer e Wikforss (2015) e Hattiangadi (2008)). Simili tentativi cercano di argomentare a favore dell'idea secondo la quale vi è un senso normativo ed uno non-normativo della nozione di correttezza; di qui, l'anti-normativista intende mostrare che la nozione di correttezza semantica che compare in CAP è una nozione non-normativa. Se questo è vero, allora non è più legittimo far discendere dalla verità di CAP la verità di TNS perché viene negata la

condizione che dovrebbe legittimare questa mossa, ovvero la normatività della nozione di correttezza. Il dibattito contemporaneo verte su queste obiezioni e sui tentativi, da parte dei normativisti, di rispondere ad esse. Il mio lavoro sull'analisi sellarsiana degli enunciati semantici vuole rappresentare un contributo ad una riflessione sull'ampia tematica della normatività del significato linguistico, a partire da un punto di vista ignorato nel dibattito contemporaneo.

## Riferimenti bibliografici

- Blackburn, Simon (1984). "The Individual Strikes Back". In: *Synthese* 58.3, pp. 281–301.
- Boghossian, Paul (1989). "The Rule-Following Considerations". In: *Mind* 98.392, pp. 507–549.
- Capone, Alessandro (2005). "Pragmemes (A Study with Reference to English and Italian)". In: *Journal of Pragmatics* 37.9, pp. 1355–1371.
- (2017). "Introducing the Notion of the Pragmeme". In: *Pragmemes and Theories of Language Use*. A cura di Keith Allan, Alessandro Capone e Istvan Kecskes. Berlino: Springer.
- (2018). "Pragmemes (Again)". In: *Lingua* 209, pp. 89–104.
- Chisholm, Roderick e Wilfrid Sellars (1958). "Intentionality and the Mental: Chisholm-Sellars Correspondence on Intentionality". In: *Minnesota Studies in the Philosophy of Science (Vol. I)*. A cura di Herbert Feigl e Michael Scriven. Minneapolis: University of Minnesota Press.
- deVries, Willem e Timm Triplett (2000). *Knowledge, Mind and the Given. Reading Wilfrid Sellars's "Empiricism and the Philosophy of Mind"*. Indianapolis: Hackett Publishing Company.
- Gibbard, Allan (2012). *Meaning and Normativity*. Oxford: Oxford University Press.
- Glüer, Kathrin e Åsa Wikforss (2015). "Meaning Normativism: Against the Simple Argument". In: *Organon (Supplementary Issue)* F.22, 63–73.
- Guardo, Andrea (2007). *Empirismo senza fondamenti – Cinque lezioni su Empirismo e filosofia della mente*. Milano: Cuem.
- Hattiangadi, Anandi (2008). "Some More Thoughts on Semantic Oughts". In: *Analysis* 69.1, pp. 54–63.
- Kripke, Saul (1982). *Wittgenstein on Rules and Private Language*. Oxford: Blackwell.
- Mey, Jacob L. (2001). *Pragmatics*. Oxford: Blackwell.
- Peregrin, Jaroslav (2012). "Inferentialism and the Normativity of Meaning". In: *Philosophia* 40.1, pp. 75–97.
- (2014). *Inferentialism: Why Rules Matter*. Basingstoke: Palgrave Macmillan.
- Sellars, Wilfrid (1954). "Some Reflections on Language Games". In: *Philosophy of Science* 21.3, pp. 204–228.

- Sellars, Wilfrid (1956). "Empiricism and the Philosophy of Mind". In: *Minnesota Studies in the Philosophy of Science (Vol. I)*. A cura di Herbert Feigl e Michael Scriven. Minneapolis: University of Minnesota Press.
- (1969). "Language as Thought and as Communication". In: *Philosophy and Phenomenological Research* 29.4, pp. 506–527.
- (1974). "Meaning as Functional Classification. A Perspective on the Relation of Syntax to Semantics". In: *Synthese* 27, pp. 417–437.
- Stalnaker, Robert (2002). "Common Ground". In: *Linguistics and Philosophy* 25, pp. 701–721.
- Whiting, Daniel (2007). "The Normativity of Meaning Defended". In: *Analysis* 67.2, 133–140.
- (2009). "Is Meaning Fraught with Ought?" In: *Pacific Philosophical Quarterly* 90.4, 535–555.
- (2013). "What is the Normativity of Meaning?" In: *Inquiry* 56.2, pp. 219–238.
- Wong, Jock (2010). "The Triple Articulation of Language". In: *Journal of Pragmatics* 42.11, pp. 2932–2944.
- Zalabardo, José (2012). "Semantic Normativism and Naturalism". In: *Continuum Companion to Philosophy of Language*. A cura di M. Kölbel e M. García Carpintero. London: Continuum Publishing Corporation.



## Alternative Possibilities in Frankfurt-style Cases

*Marco Giovanni Stucchi*

**Abstract.** In this article I will criticize the so called Frankfurt-style cases. These cases have been built with the purpose of denying that in order to hold a person responsible, it's necessary this person could have done otherwise, namely there were alternative possibilities. I present here two arguments according to which Frankfurt's cases did not succeed in this attempt. In the first argument I will focus, from and ethical and a logical point of view, on the Principle of Alternative Possibilities, and I will show that the Frankfurt-style cases are not truly able to falsify it. The second argument is semantical and concerns the formulation of Frankfurt example.

**Keywords.** Free Will, Harry Frankfurt, Liberty, Responsibility.

## 1 Introduction to the problem

It seems obvious that in order to hold a person responsible for something, we need to assume that she could have done otherwise. In other words we naturally take for granted the so called Principle of Alternative Possibilities (PAP): a person is morally responsible for what she has done only if she could have done otherwise (see Frankfurt (1969, p. 829)). This principle was accepted not only by the common sense, but also from the majority of philosophers, at least until Frankfurt challenged it in his very influential “Alternative possibilities and Moral Responsibility”. That paper, according to many, has proved that PAP is false. In this paper I would like to criticize Frankfurt’s argument by showing, first of all, that PAP must be clarified and, thereafter, that there are alternative possibilities in Frankfurt-style cases.

I will build here a Frankfurt-style case that should undermine PAP. In the literature there have been many complications and reformulations of Frankfurt’s original case in order to defend Frankfurt-style cases from some criticisms (see Mele and Robb (1998, p. 101) and see Fischer (2011, p. 249)). I believe however that the following example can be representative of the debate in general, and can work for the purpose of this paper. Suppose Mr. Black knows that Jones is a political dissenter and he desires an armed revolution. Black knows that tomorrow morning Jones will show up in front of the parliament with a gun in his coat in order to kill the Prime Minister<sup>1</sup>. Black, however, is aware that Jones is kind of a coward man, and it would not be the first time that he show up armed in front of the parliament without doing anything. This time Black, who wants the death of the Prime Minister, has placed in Jones’ brain, without Jones knowing, a sophisticated mechanism, able to screen off Jones’ thought, and able to force him to kill the Prime Minister if Black pushes a certain button. Black does not want to interfere unnecessarily: he will push the button only if he sees that Jones shows the intention (and Black can check this “mental process” thanks to the mechanism) to not kill the Prime Minister. However that morning Jones is particularly resolute to kill the Prime Minister and, on his own, he shoots her with the gun. If, as Frankfurt persuasively suggests, Jones is morally responsible it seems that PAP is false, since he cannot do otherwise. In any case, in fact, he would have killed the Prime Minister.

Most philosophers have found Frankfurt-style cases compelling (Ginet 1996, p. 404), and this kind of argument has been used by the supporters of compatibilism about free will. Although many philosophers have taken for granted that Frankfurt has shown PAP to be false, there are some dissenters (Widerker 1995, p. 247 and see O’Connor 2005). There are different ways to attack Frankfurt-style

<sup>1</sup>In the example at stake the agent is blameworthy, because I assume that the action is morally wrong. In this paper I will not consider whether Frankfurt-style cases can work both for blameworthy and praiseworthy actions (Nelkin 2011, p. 1), and I will assume a symmetry between these two.

cases; in this paper I will refer to some elements of Widerker's criticism, one of the most influential voices conflicting with Frankfurt.

## 2 Alternative possibilities in Frankfurt-style cases

I believe that in order to criticize Frankfurt-style cases we should look carefully at PAP, trying to understand what such a principle is supposed to mean. PAP as stated above, narrows down the responsibilities just for actions. But since, as van Inwagen points out (Fischer 2011, p. 245), we can hold individuals morally responsible also for others items such as omissions, consequences and intentions, I would like to generalize PAP. In this new formulation PAP claims that a person is morally responsible for S, only if the person could have avoided S. I will take PAP as divided in two different parts. The first part (*i*) is: "a person is morally responsible for S" and the second part (*ii*) is: "the person could have avoided S". These two parts are connected by a classic material conditional:  $(i) \rightarrow (ii)$ . If a person could have avoided S it means that is possible that S does not occur. Moreover, if it possible that S does not occur the person could have avoided S<sup>2</sup>. If we formalize this we have  $(ii) \leftrightarrow \Diamond \neg S$ . Putting these elements together we have that if  $(i) \rightarrow (ii)$  and  $(ii) \leftrightarrow \Diamond \neg S$ , it will follow that  $(i) \rightarrow \Diamond \neg S$ .

Consider now the example of the Frankfurt-style case in the first section. In that case we can substitute S with killing the Prime Minister. Frankfurt claims to have found a case in which  $\neg \Diamond \neg S$ , namely in which it is not possible that Jones does not kill the Prime Minister, but in which (*i*) is true, because we consider Jones responsible for killing the Prime Minister. In other words Frankfurt has shown that  $(i) \rightarrow \Diamond \neg S$ , which follows directly from PAP, is wrong.

I would like to take a moment to analyse what S means in this case. Just above we have taken S for 'killing the Prime Minister'. So at the question about what Jones is responsible for, it seems *prima facie* plain that the correct answer would be, simply, an action. If we say that S, the item for which Jones is hold responsible, corresponds with the action A 'killing the Prime Minister', it seems that Frankfurt is right. He actually gave us a case in which  $(i) \rightarrow \Diamond \neg S$  is false. In fact, if we substitute A in this conditional we have that Jones is responsible for A, that is for his killing the Prime Minister, and that  $\neg \Diamond \neg A$ , since in any case, whether on his own or for the interference of Black, he would have killed the Prime Minister.

If it is just the action at stake then Frankfurt is right. But are we sure that Jones is responsible just for his action? It seems that in the example there is another item that plays a relevant role in Jones' responsibility. In my made-up case Jones does not just kill the Prime Minister, but he kills her with the intention to do it.

<sup>2</sup>I believe that I can claim this without claiming that the person must be responsible for her avoiding S.

Jones decides<sup>3</sup> to perform a certain action, and he acts on the basis of this decision. With a closer look to my example we could say that S is the action A based on the intention  $\Upsilon$  to perform such an action. Can Frankfurt's argument work if  $S = A + \Upsilon$ ? In other words is it true that Jones is responsible for the action based on the intention to perform that action, despite the fact that it was impossible that his action based on the intention to perform that action does not occur? To find out let's substitute S with  $A + \Upsilon$ . I am assuming, and it seems reasonable to do it, that Jones is responsible for  $A + \Upsilon$ , so the antecedent of  $(i) \rightarrow \Diamond \neg S$  is true. As we already know, in order to succeed in her aim a Frankfurt's supporter should prove that the consequent of our conditional is false, namely that  $\neg \Diamond \neg S$ . Since I am examining the case in which S is  $A + \Upsilon$ , it is clear that Frankfurt's supporter needs to prove that  $\neg \Diamond \neg (A + \Upsilon)$ . Frankfurt's supporter has an easy time in doing so because she can argue in the same way we have seen above. Since there is no possibility that the action does not occur, because of the presence of Black's mechanism, there is no possibility that the action based on the intention to perform that action would not occur. Frankfurt's supporters can claim in a convincing way that  $\neg \Diamond \neg (A + \Upsilon)$ , and therefore that PAP is wrong.

We have seen two cases so far: one in which S is identical to A and one in which S is identical to  $A + \Upsilon$ . In the first Jones is morally responsible for killing the Prime Minister, in the second Jones is morally responsible for killing the Prime Minister based on his intention for acting in this way. In these two cases we have seen that Frankfurt's argument is successful. I suggest however that now we should test carefully our intuitions about what Jones is responsible for. It seems sensible to affirm that Jones is responsible for his acting in such a way with the intention to do it (our second case). I suggest now to divide the two items at stake in this case: the action A and the intention  $\Upsilon$ . It could be, in fact, that it is just one of these two items that plays a relevant role in holding Jones responsible. I suggest the following device in order to identify the item we are looking for: the item essential for Jones' responsibility (Mele 2008, p. 277). In this situation we have three elements: (i) Jones' responsibility, A Jones' killing the Prime Minister,  $\Upsilon$  Jones having the intention to kill the Prime Minister. I have already established that if both A and  $\Upsilon$  occur we have (i) as an outcome. Now I am going to see if (i) depends just on one of these two items. If I verify that removing one of these two items, and leaving the other as it is, there is a change in the truth value of (i), I can claim that (i), Jones' responsibility, depends on the former item; therefore I will conclude that such an item plays a fundamental role in Jones' responsibility.

<sup>3</sup>Although it could be useful, at a deeper level, to distinguish between 'have the intention to perform an action' and 'decide to perform an action', I believe that, in regard to this paper, this distinction is not necessary.

<sup>4</sup>I will use the non-logical sign + without any precise meaning.  $A + I$  means just the action based on the intention to perform such an action. If I have used the logical connective  $\wedge$ , I would have run into some complications, as the reader can certify.

Take the case in which Jones, that morning, was overwhelmed by his same old cowardice, and he abandoned his intention to kill the Prime Minister. What would happen in this case? Black, who can read Jones' mind as a book thanks to his mechanism, would interfere and make Jones kill the Prime Minister. What should we say about Jones' responsibility in a such a situation? It seems obvious that, if we know the circumstances, namely that in Jones' brain there is a mechanism that will force him in the murder if he does not show the intention to kill the Prime Minister, nobody would hold Jones morally responsible for his killing the Prime Minister. If we remove Jones' intention to kill the Prime Minister ( $i$ ) is false. On the other hand, according to our example, given that Jones has the intention to kill the Prime Minister, we would hold, as Frankfurt clearly states, Jones responsible for his killing the Prime Minister. The point seems to be clear: if there is Jones' intention to kill the Prime Minister then Jones is responsible ( $\Upsilon \rightarrow (i)$ ); if there is no intention to kill the Prime Minister then Jones is not responsible ( $\neg\Upsilon \rightarrow (i)$ ). From this follows that  $\Upsilon \leftrightarrow (i)$ . What should we say, in the Frankfurt-style case at issue,<sup>5</sup> about the role of the action A in the explanation of Jones' responsibility? I have shown two cases in which ( $i$ )'s truth value changes: the case in which there is Jones' intention to kill the Prime Minister, and we have that ( $i$ ) is true, and the case in which Jones has no intention to kill the Prime Minister, and we have that ( $i$ ) is false. In both cases A occurs. It seems evident therefore, that the action does not play an essential role in Jones' responsibility, while the intention does, since, as I have shown,  $\Upsilon \leftrightarrow (i)$ .

I believe I have just proved that S, the item for which Jones' is morally responsible, corresponds with  $\Upsilon$ , namely Jones' intention to kill the Prime Minister. Therefore PAP, in my case, takes this shape:  $(i) \rightarrow \diamond\neg\Upsilon$ . To reiterate, in order to prove the validity of Frankfurt's criticism to PAP, it must be proved that the consequent is false. Hence, in this last case, Frankfurt's supporter should prove that  $\neg\diamond\neg\Upsilon$ . But this, as we have just seen, is not the case. About  $\Upsilon$  there are alternative possibilities: both  $\diamond\neg\Upsilon$  and  $\diamond\Upsilon$  are true. Jones has the possibility to have the intention to kill the Prime Minister and the possibility not to have such an intention. The interference system prevents Jones only from the possibility of  $\neg A$ , because in any case Jones will act A, namely  $\neg\diamond\neg A$ . Frankfurt's supporter is not able to show that  $(i) \rightarrow \diamond\neg\Upsilon$  is false. I have proved that, as regarding the item that truly plays a role in my case, there are alternative possibilities. These alternative possibilities are relevant for responsibility because, if my argument works, they concern the essential item that explains responsibility itself.

<sup>5</sup>Which I take as model for the most common Frankfurt-style cases.

### 3 A criticism to my argument and another proposal

The main argument of this paper has been already presented in the second section. In this section I would like to show a possible criticism to my argument and to address to it, and finally, I will sketch a new argument against Frankfurt-style cases that should prove what I have already shown in the second section.

A Frankfurt supporter can raise an objection to my argument. She can build a different example in which Black's interference can happen in a different moment. In my previous example Black pushes the button after Jones shows an intention about his action. If he shows an intention to kill the Prime Minister Black will not interfere; on the other hand if he shows an intention to not kill the Prime Minister Black will interfere. The moment in which Black can interfere, removing alternative possibilities, is after the intention. Frankfurt's supporter could make up a case in which Black can interfere before the intention. This move should allow Frankfurt's supporter to remove the alternative possibility that I discovered in section 2. Let's make explicit the example I am talking about. Jones is the same person I described above, and he shows up in front of the parliament with a gun. The night before Black had installed a special mechanism in Jones' brain. This mechanism is able to show if Jones is *forming* the intention to kill the Prime Minister or not, and, in the case that Jones is forming the intention not to kill the Prime Minister, Black can push a button that triggers a process in Jones' brain, that ends with Jones' intention to kill the Prime Minister. That morning in front of the parliament Jones forms on his own the intention to kill the Prime Minister and he acts according to his intention. In this case it seems that we hold Jones responsible for his killing the Prime Minister, and there was not an alternative possibility about his intention to do it. Black's mechanism prevents Jones from the intention not to kill the Prime Minister. If we come back to the logical formalization I proposed in section 2, it means that, given that  $S$  is  $\Upsilon$ , ( $i$ ) is true and  $\neg\Diamond\neg\Upsilon$ . In this case  $(i) \rightarrow \Diamond\neg\Upsilon$  is false, therefore Frankfurt is right about PAP.

I think that this criticism to my argument can be rejected (see Widerker (1995, pp. 249-251)). I suggest thinking about what 'forming an intention' really means, since forming an intention plays an important role in the last example. I believe that 'Jones is forming an intention' can have two different meanings relevant for my inquiry. Jones is forming an intention  $\Upsilon$  can mean either 1) Jones will have the intention  $\Upsilon$ <sup>6</sup> or 2) Jones will likely have the intention  $\Upsilon$  but he can change his mind. It seems to me that these two cases are jointly exhaustive. Take the case 1). It seems that in this case there is no difference with the case I have analysed in section 2. If forming the intention will necessarily bring about the in-

<sup>6</sup>It this case there is a necessary relation between forming an intention and having such an intention.

tention, Frankfurt's supporter has just pushed the debate one step back without any change regarding the outcome. In fact I can just move the responsibility from the intention to the forming the intention,<sup>7</sup> and I can show that in this last example there are indeed alternative possibilities about forming the intention. Let's move on to 2). In this case it seems obvious, by definition, that Jones has alternative possibilities: he can form his intention to kill the Prime Minister, but still he could change his mind and give up that intention. If what I have just said is coherent I see no reason why my argument against Frankfurt-style cases should be undermined by this last more sophisticated example.

I would like now to mention another possible way to attack Frankfurt-style cases. I will not go into detail with this part, because I believe I have already proved what was of concern, namely the fact that in Frankfurt-style cases there are alternative possibilities about the item relevant for the moral responsibility.

PAP states that an agent is morally responsible for what she has done only if the agent could have done otherwise. I believe that Frankfurt-style cases do not prove that PAP is wrong since they change the meaning of the term 'agent' in PAP. More precisely I think that when Frankfurt's supporter claims that "an agent is morally responsible for what she has done, even if she could not have done otherwise", 'agent' and 'she' do not have the same meaning since 'she' is not an agent at all. Frankfurt's supporter needs to prove that the individual that is morally responsible for something is the same individual that could not have done otherwise. I believe that as regarding (*i*), namely holding responsible Jones, we are holding responsible an agent, in the genuine sense of agent. But what should we say about (*ii*)? By saying that Jones does not have alternative possibilities, are we talking about an agent in the genuine sense of agent? If the answer is negative, as I would like to suggest, it seems that Frankfurt's supporter, in her attack to PAP, is playing with words since the subject of (*i*) is not the same subject as in (*ii*). Saying that Jones has not alternative possibilities, in other words that he could not do otherwise, means, in the case at stake, that Jones cannot not kill the Prime Minister. Jones cannot not kill the Prime Minister because Black's mechanism, by brainwashing Jones' mind, would force him to kill the Prime Minister if he shows the intention not to kill the Prime Minister. If we say that Jones has no alternative possibilities, we are saying that in the possible world in which Jones does not show the intention to kill the Prime Minister, Black would push the button, causing Jones' to kill the Prime Minister. I believe that in this possible world Jones, since he would be brainwashed, would lack his property to be an agent because he would lack any kind of control that we expect from an agent.

<sup>7</sup>If we assume that not only forming an intention  $\Upsilon$  necessarily implies having an intention  $\Upsilon$ , but also that, as it should be plain, having an intention  $\Upsilon$  necessarily implies that one had forming an intention  $\Upsilon$ ; there is, as a matter of fact, no difference between forming an intention  $\Upsilon$  and having an intention  $\Upsilon$ .

If the reader shows some resistances regarding what I have just argued, I think that this it is due to the particular story of the example. In my example Jones, that morning, goes out with a gun in his coat and with hate for the Prime Minister in his heart. If he will not shoot, it is just because of his cowardice. Moreover we know that his deciding to kill the Prime Minister is up to him, and it is a matter of subtle feelings and thoughts in his mind. Maybe we could be inclined to see the interference mechanism as a little push in a certain direction. I would like to change the story in order to give force to my argument.

That morning, in front of the parliament, there will be a parade in which the Prime Minister will walk among the people and she will shake some hands. Also in this example Black hates the Prime Minister, but not so much as in the previous one. Black decides to put (on the sly) a mechanism in the brain of a random citizen, who will participate at the parade. This mechanism shows to Black if the citizen will have the intention the punch the Prime Minister in the face, and it will enable Black, if the citizen does not show such an intention, to press a button that will cause the citizen to punch the Prime Minister. We do not know anything about this citizen, but as it normally happens, we would not think that she is going to punch the Prime Minister. As expected, the unlucky citizen does not show the intention to punch the Prime Minister, therefore Black interferes. Would we say that the citizen, in this case, preserves her property to be an agent? I do not think so. She is totally forced to do something nobody would ever expect from her. If I am right Frankfurt's supporter is not able to prove that the agent, properly speaking, has no alternative possibilities. The subject of *(ii)* for whom there are no alternative possibilities, is not an agent as the subject of *(i)* that we hold responsible. If I am right Frankfurt's supporter has not proved that PAP is false.

## 4 Conclusion

I think more should be said about this last criticism against Frankfurt-style cases. First of all the notion of agent could be discussed in details. However I believe that I have already carried out the task of this paper. In conclusion I would like to note that even if Frankfurt-style cases do not represent a valid argument for compatibilism, there can be still some good reasons for compatibilism. For example one could conceive liberty in the Kantian negative sense, namely as the freedom from specific bounds and interferences. The reader can easily note that, according to such conception, liberty can survive the loss of PAP.

## References

- Fischer, John Martin (2011). "Frankfurt-Type Examples and Semicompatibilism: New Work". In: *The Oxford Handbook of Free Will (second edition)*. Ed. by Robert Kane. Oxford: Oxford University Press, pp. 243–265.
- Frankfurt, Harry (1969). "Alternate Possibilities and Moral Responsibilities". In: *Journal of Philosophy* 66.1, pp. 829–839.
- Ginet, Carl (1996). "In Defense of the Principle of Alternative Possibilities: Why I Don't Fine Frankfurt's Argument Convincing". In: *Philosophical Perspectives* 10, pp. 403–417.
- Mele, Alfred (2008). "Manipulation, Compatibilism and Moral Responsibility". In: *The Journal of Ethics* 12.3, pp. 263–286.
- Mele, Alfred and David Robb (1998). "Rescuing Frankfurt-Style Cases". In: *The Philosophical Review* 107.1, pp. 97–112.
- Nelkin, Dana Kay (2011). "Abilities". In: *Making Sense of Freedom and Responsibility*. Ed. by Dana Kay Nelkin. Oxford: Oxford University Press.
- O'Connor, Timothy (2005). "Freedom with a Human Face". In: *Midwest Studies in Philosophy* XXIX, pp. 207–227.
- Widerker, David (1995). "Libertarianism and Frankfurt's Attack on the Principle of Alternative Possibilities". In: *The Philosophical Review* 104.2, pp. 247–261.