

T E M I

# AUTOINGANNO

di Patrizia Pedrini

*ABSTRACT - In questo tema viene esplorato il fenomeno dell'autoinganno, che appartiene alla famiglia della cosiddetta "irrazionalità motivata". Ripercorreremo i nodi problematici principali su cui la discussione filosofica si concentra, tra i quali il problema di stabilire se l'autoinganno sia o non sia intenzionale; affronteremo il problema del contenuto motivazionale dell'autoinganno, riassunto dalla domanda "Che cosa vuole chi si autoinganna?"; cercheremo di aggiudicare la questione se l'autoingannato sia davvero convinto della verità di quel che va raccontandosi. Vedremo anche che cos'è che l'autoingannato non sa di sé mentre si autoinganna e analizzeremo come dobbiamo giudicare l'autoinganno in qualità di osservatori esterni.*

1. INTRODUZIONE

2. L'AUTOINGANNO È INTENZIONALE?

3. CHE COSA VUOLE CHI SI AUTOINGANNA?

4. L'AUTOINGANNATO È DAVVERO CONVINTO DI QUEL CHE CREDE?

5. CHE COS'È CHE L'AUTOINGANNATO NON SA DI SÈ?

6. COME DOBBIAMO GIUDICARE L'AUTOINGANNO?

BIBLIOGRAFIA

## 1. INTRODUZIONE

Ogni uomo ha dei ricordi che racconterebbe solo agli amici. Ha anche cose nella mente che non rivelerebbe nemmeno agli amici, ma solo a se stesso, in segreto. Ma ci sono anche altre cose che un uomo ha paura di rivelare anche a se stesso, e ogni uomo perbene ha un certo numero di cose del genere accantonate nella mente». Così scriveva Fëdor Dostoevskij in una memorabile pagina di Memorie del sottosuolo. Con il consueto acume psicologico e con una magistrale capacità di sintesi, Dostoevskij sta rilevando che spesso gli esseri umani non si raccontano la verità e, altrettanto spesso, al suo posto, si raccontano invece una storia dai risvolti più desiderabili, nonostante le evidenze che le cose non stiano così siano a disposizione, o siano perlomeno facilmente ottenibili. Ecco qua, *in nuce*, il fenomeno dell'autoinganno. Una varietà di evidenze empirico-sperimentali e cliniche, nonché una gran quantità di contesti ordinari, mostrano che si tratta di un fenomeno relativamente pervasivo, dal quale nessuno è completamente immune, e dal quale la nostra vita quotidiana è spesso pesantemente condizionata. Vi sono persone platealmente tradite dal partner che non credono che il compagno sia fedifrago; storie d'amore ormai logore di cui non si riesce a credere che siano finite, segnali di malattia preoccupanti che trascuriamo imputando gli inequivocabili disturbi alla stanchezza da lavoro; saggi legittimamente rifiutati che ci paiono ottimi lavori ingiustamente respinti; dipendenze da sostanze tossiche che crediamo siano il frutto di un comportamento scelto autonomamente. In situazioni come queste, raccontiamo a noi stessi una storia che non corrisponde alla verità e a un interprete attento non sfugge che, se fossimo imparziali o non coinvolti emotivamente e cognitivamente nella situazione nella quale ci troviamo, se non avessimo una motivazione o un interesse a credere qualcosa di più piacevole, non sarebbe difficile vedere e dirsi come stanno le cose. [Pedrini 2013, pp. VII-VIII]

Insomma, ci autoinganniamo in molte occasioni e lo facciamo per un “motivo”, ovvero per il desiderio che le cose stiano un certo modo. Non ci autoinganniamo, insomma, a causa di limiti cognitivi strutturali congeniti alla mente della specie umana o per distrazione e inaccuratezza epistemica, come può accadere per altre forme di irrazionalità. Rilevando che l'autoinganno è probabilmente causato da questa forte componente motivazionale, che mette in moto l'intero processo cognitivo che lo produce e lo devia nella direzione della credenza desiderata, i più accreditati teorici di questo fenomeno lo classificano come appartenente alla più ampia famiglia di fenomeni noti appunto come casi di “irrazionalità motivata”. Si dice “motivata” ogni forma di

irrazionalità in cui non incorreremmo se non fossimo appunto condizionati da uno stato motivazionale, ad esempio un desiderio.<sup>1</sup>

Al fine di offrire al lettore una panoramica coerente del dibattito in corso e dei principali problemi che vengono attualmente discussi nella comunità anglofona, organizzo questo tema nel modo seguente. Nel Paragrafo 1 ripercorrerò i nodi problematici principali su cui la discussione filosofica degli ultimi 30 anni si è concentrata, tra i quali figura centralmente il problema di stabilire se l'autoinganno sia o non sia intenzionale. Due teorie principali, l'intenzionalismo e anti-intenzionalismo, si sono divisi il campo polemico, fino alla consacrazione dell'anti-intenzionalismo come posizione prevalentemente accettata dagli studiosi e di cui sono disponibili diverse varianti. Al termine del paragrafo accennerò anche ai lineamenti di una terza via, il cosiddetto "intenzionalismo debole", che assieme ad altri anche io difendo. Nel Paragrafo 2 affronterò il problema del contenuto motivazionale specifico dell'autoinganno, riassunto dalla domanda "Che cosa vuole chi si autoinganna?". Alcuni autorevoli studiosi sostengono che l'autoingannato vuole non già che uno stato del mondo sia quello che essi desiderano, bensì vuole trovarsi in uno stato di credenza che quello stato sia quello che desidera che sia. Avrebbe cioè un desiderio rivolto verso se stesso, e non (o non necessariamente anche) rivolto verso il mondo. Cercherò di articolare brevemente le ragioni del mio dissenso, in virtù delle quali a parer mio l'autoingannato ha un desiderio rivolto verso il mondo e vuole che un certo stato di cose sia vero. Nel Paragrafo 3, illustrerò un problema che, accanto agli altri, costituisce una delle maggiori ragioni di

---

<sup>1</sup> All'interno di questa classe di fenomeni non figura soltanto l'autoinganno, ma anche l'*akrasia* (o «debolezza della volontà»), la formazione di intenzioni illusorie o irrazionali, il cosiddetto *wishful thinking* (letteralmente «pensiero che è espressione di un desiderio», e che potremmo caratterizzare come un'attività di pensiero al limite della fantasticheria in cui il soggetto indulge in maniera anomala e che, in modi da chiarire, condiziona il suo comportamento e una parte della sua ulteriore attività di pensiero).

fascino che l'autoinganno suscita, e cioè se l'autoingannato sia davvero convinto della verità di quel che va raccontandosi. Ripercorrerò alcuni argomenti secondo cui l'autoingannato in fondo sa come stanno le cose e cercherò di mostrare che l'*appeal* di questi argomenti risiede nel carattere non dinamico della teoria in seno alla quale sono concepiti. Nel Paragrafo 4 vedremo che cos'è che l'autoingannato non sa di sé mentre si autoinganna. Occorre infatti anche una spiegazione del perché non si accorge che si sta autoingannando. Infine, nel Paragrafo 5 affronteremo il problema di come dobbiamo giudicare l'autoinganno come osservatori esterni. Secondo alcuni dovremmo guardare a questo fenomeno con simpatia. Cercherò di sostenere che la simpatia non è un atteggiamento che abbiamo invariabilmente ragione di adottare verso l'autoingannato.

## 2. L'AUTOINGANNO È INTENZIONALE?

Due gruppi principali di teorie si contendono la descrizione corretta dell'autoinganno: le teorie «intenzionaliste» (o «tradizionaliste») e le teorie «anti-intenzionaliste» (o «deflazionistiche»). Gli intenzionalisti sostengono che l'autoinganno è un processo psicologico che avviene per il tramite di una precipua attività “intenzionale”, nel senso di “volontaria”, condotta dal soggetto: l'assunzione minima dell'intenzionalismo è che l'autoingannato gioca un ruolo decisivo nel produrre il suo stato di autoinganno, «inducendosi intenzionalmente a credere che  $p$ » a dispetto dell'evidenza che  $\text{non-}p$ , un'evidenza che è o sarebbe facilmente accessibile. Talora gli intenzionalisti fanno anche assunzioni più forti, ad esempio che l'autoingannato creda quel che in origine «sapeva» essere falso, o che continui a credere che la proposizione autoingannevole sia vera mentre crede anche la sua falsità, cioè «sa» che è falsa; o ancora, che

l'autoingannato crede che  $p$  sia vera proprio perché sa che è falsa. In quest'ultimo caso, si dice che l'autoingannato «deve volere il suo autoinganno» [Davidson 1985a, p. 87].

Ora, il problema è che qualora fosse vera la teoria degli intenzionalisti, la razionalità umana sarebbe gravemente minacciata: in che senso, infatti, sarebbe possibile indurci intenzionalmente a credere ciò che dobbiamo perlomeno sospettare essere una falsità o una verità estremamente improbabile?<sup>2</sup> Data la forza di questo «sospetto», lo stato doxastico complessivo del soggetto che si autoinganna finirebbe per essere simile o persino identico a quello di una persona che intrattiene una coppia di credenze contraddittorie (il cosiddetto “paradosso statico”). Inoltre, non è chiaro come sia possibile che un'intenzione deliberata di credere quello che non si crede, o che è immensamente difficile credere, possa avere successo (il cosiddetto “paradosso dinamico”). Il problema a cui sto alludendo ha a che fare con il fatto che la credenza, per sua natura, sembra mirare ad essere vera e a rappresentare un fatto del mondo. In altre parole, è lo stato di cose che una credenza rappresenta ciò che determina la risposta doxastica di un agente epistemico competente. L'evidenza con cui un soggetto entra in contatto contiene, cioè, un elemento che si potrebbe definire di «necessità» che è costitutivo di ciò che significa avere quella credenza. Il soggetto non è «libero», per così dire, quando crede ciò che l'evidenza gli mostra o gli addita, bensì è «costretto» a credere ciò che l'evidenza gli indica.

Gli anti-intenzionalisti hanno una risposta diretta a questi due problemi. Secondo gli anti-intenzionalisti, infatti, il processo di autoinganno non consiste affatto in una transizione dall'intenzione di credere che  $p$  alla credenza che  $p$ , bensì è avviato da un

---

<sup>2</sup> Per un esame accurato della «credenza a dispetto dell'evidenza», noto anche come «credenza incontinente» vedi Mele 1986, pp. 212-229. Il concetto di «credenza incontinente» è discusso anche in Heil 1984, pp. 56-70.

*desiderio* che  $p$ , il quale conduce, attraverso meccanismi che devono essere specificati, alla credenza autoingannevole che  $p$ . Avere un desiderio che  $p$  può insomma essere sufficiente a condizionare le nostre operazioni cognitive e a portarci a formare delle credenze autoingannevoli.

Alfred Mele, il più influente tra gli anti-intenzionalisti, ha offerto quattro condizioni che egli indica come *congiuntamente sufficienti* per entrare in uno stato di autoinganno [Cfr. Mele 1997, 1998 (pp. 37-58) e 2001]:

- a) la credenza che  $p$  che  $S$  acquisisce è falsa;
- b)  $S$  tratta i dati pertinenti, o apparentemente pertinenti, per determinare il valore di verità di  $p$  in maniera motivazionalmente pregiudiziale;
- c) Il trattamento pregiudiziale è una causa non deviante dell'acquisizione da parte di  $S$  della credenza che  $p$ ;
- d) Il corpo di dati posseduti in quel momento da  $S$  forniscono una giustificazione più solida per non- $p$ , piuttosto che per  $p$  [Cfr. Mele 2001, cap. 3].

Mele [1997, 2001, cap. 3] adotta il modello del test delle ipotesi ordinarie elaborato da Trope e Liberman [Cfr. Trope e Liberman 1996, pp. 120-137] come dimostrazione empirica del modo in cui la motivazione può pregiudicare la cognizione. Secondo Trope e Liberman, le persone hanno differenti soglie di accettazione o rifiuto delle ipotesi, soglie che dipendono dal costo soggettivo atteso che per ciascun individuo comporta l'errata accettazione o l'errato rifiuto. Il costo è una funzione delle risorse necessarie per acquisire e trattare l'informazione. Più alto è il costo dell'errata accettazione dell'ipotesi che  $p$ , più alta è la soglia di accettazione, e la stessa cosa accade per l'errato rifiuto. Le

ipotesi che hanno una soglia di accettazione alta saranno cioè più rigorosamente testate e valutate, mentre quelle che hanno una soglia bassa di accettazione saranno abbracciate più facilmente. Ora, nell'autoinganno il costo soggettivo atteso associato all'acquisizione della falsa credenza che  $p$  è basso. Per esempio, il marito che crede falsamente che sua moglie sia fedele, quando così desidera, ha delle soglie di accettazione differenti, rispettivamente, per l'ipotesi che sua moglie sia fedele e per l'ipotesi che sua moglie sia infedele. In altre parole, quest'uomo è più felice di credere falsamente che sua moglie sia fedele che non di credere falsamente che sua moglie sia infedele, perché desidera che sua moglie non abbia una relazione extraconiugale. Mele quindi pensa che il desiderio che  $p$  sia vero possa condurre a un trattamento motivazionalmente pregiudiziale delle informazioni, così da produrre un abbassamento della soglia di accettazione dell'ipotesi falsa che  $p$  e un innalzamento della soglia di rifiuto dell'ipotesi vera che  $\text{non-}p$ .<sup>3</sup>

### 3. CHE COSA VUOLE CHI SI AUTOINGANNA?

Un'altra questione centrale che interessa coloro che studiano l'autoinganno è quella di specificare quale sia il contenuto motivazionale specifico di colui che abbraccia una credenza autoingannevole. Insomma, vogliamo sapere che cosa vogliono gli autoingannati, di cosa tutti quanti fondamentalmente vanno in cerca. Eric Funkhouser ha recentemente formulato in maniera lucida questa domanda [Funkhouser 2005] e ha

---

<sup>3</sup> Vi sono altre forme di anti-intenzionalismo che ho discusso in Pedrini 2013, a cui rinvio per un approfondimento che non è qui consentito per ragioni di spazio. Nello stesso contesto elaboro 4 obiezioni all'anti-intenzionalismo, cercando di mostrare come, sebbene questo modello sia di gran lunga migliore, dal punto di vista esplicativo, rispetto all'intenzionalismo, tuttavia deve essere emendato da alcuni difetti importanti, tra cui figurano l'annoverare tra i casi di autoinganno in linea di principio anche certi fenomeni di credenza falsa motivata in animali non umani e il non distinguere a sufficienza tra autoinganno e altri fenomeni di irrazionalità motivata che si verificano nella psicologia umana.

tentato di darle una risposta coerente. Secondo Funkhouser chi si autoinganna sta cercando di trovarsi in un certo stato mentale che desidera (o in ogni caso ha un *interesse* ad avere), cioè nello stato di credenza che  $p$ , e il fatto che questo sia il suo desiderio (o interesse) è il motore motivazionale primario e trainante dell'autoinganno, rispetto al quale il desiderio (o l'interesse) che il mondo sia in un certo modo, ad esempio che  $p$  sia vera, è invece *contingente*. In altre parole, l'autoingannato vuole uno stato mentale e le sue caratteristiche qualitative, che possono variare, a seconda dei tipi di autoinganno in cui incorre, dalla piacevolezza ad altre forme di «soddisfazione» o «importanza» (non sempre necessariamente accompagnate da piacevolezza, come nel caso dell'autoinganno «invertito», cosiddetto, in cui un soggetto  $S$  finisce per credere autoingannevolmente una proposizione  $p$  che non desidera, che lo turba o che gli crea disagio [cfr. Pedrini 2013, cap. 2 e Pedrini 2012]. Volendo principalmente trovarsi in uno stato mentale, l'autoingannato ha quindi *un desiderio, o interesse, rivolto principalmente verso se stesso, e solo contingentemente avrebbe anche (ma non sempre, come vedremo, e com'è ovvio, visto il suo carattere di contingenza) un desiderio, o interesse, rivolto verso il mondo*. Si noti che la questione sollevata da Funkhouser circa il contenuto dello stato motivazionale operativo nell'autoinganno è una disputa interna all'anti-intenzionalismo (o «motivazionalismo», come Funkhouser lo denomina). Funkhouser, cioè, riconosce l'anti-intenzionalismo come la teoria dell'autoinganno sostanzialmente corretta.

Funkhouser si rifà a Dana K. Nelkin [2002], la quale ha difeso quella che Funkhouser ritiene essere la risposta corretta: ogni volta che una persona si autoinganna rispetto a  $p$ , l'autoinganno della persona è motivato dal *desiderio di credere* che  $p$ . Questa sarebbe la cosiddetta teoria del «desiderio di credere» [cfr. Nelkin, 2002], e secondo Nelkin



avrebbe anche la virtù di offrire una spiegazione unificata dell'autoinganno diretto e di quello invertito.

La tesi per cui l'autoingannato è mosso da una motivazione rivolta verso se stesso e i propri stati mentali, e non verso l'accertamento di uno stato di cose del mondo che per lui sia eventualmente anche desiderabile, sarebbe confermata dal "comportamento di evitamento" degli autoingannati, i quali non di rado mostrano comportamenti tesi proprio a evitare le fonti di possibile evidenza di come le cose in realtà stanno.

Vi sono, tuttavia, ragioni di insoddisfazione nei confronti di questa teoria. Ad esempio, affinché la teoria di Funkhouser reggesse, bisognerebbe che il soggetto che si autoinganna, oltre ad evitare le fonti di evidenza che gli creano disagio, non facesse nulla di epistemicamente rilevante per rafforzare la sua credenza autoingannevole; mentre invece è tipico dell'autoinganno un'intensa opera di giustificazione agli occhi di se stessi, e spesso anche degli altri, della proposizione che si vuole credere [cfr. Pedrini 2013, cap. 2 e Pedrini 2012]. Funkhouser sembra insomma operare con una distinzione troppo *tranchant* tra «desideri (o interessi) rivolti verso il mondo» e «desideri (interessi) rivolti verso se stessi», mentre invece le due motivazioni paiono concettualmente co-implicate nella psicologia reale dei soggetti, per ragioni che hanno a che fare con la natura stessa dello stato che ci piacerebbe acquisire, cioè una credenza. Questa risposta deve ovviamente essere corredata da una convincente spiegazione del comportamento di evitamento, che mostri come sia possibile evitare il contatto con la realtà e allo stesso tempo essere interessati a che la realtà sia in un certo modo.<sup>4</sup>

#### 4. L'AUTOINGANNATO È DAVVERO CONVINTO DI QUEL CHE CREDE?

---

<sup>4</sup> Per un approfondimento di questo tema e per i lineamenti di questa spiegazione, cfr. Pedrini 2012, cap. 2, e Pedrini 2013, cap. 2.

Di fronte a molti casi di autoinganno, si ha spesso il sospetto che in fondo questi soggetti non possano essere davvero arrivati a convincersi che le cose stiano come loro desiderano che stiano. Si ha la tentazione di non accordare all'autoingannato la credenza che sostiene di avere probabilmente, in parte, perché l'evidenza a sfavore di quella credenza è talmente ovvia che sembra impossibile che sia davvero potuto sfuggirgli che le cose stiano in un certo modo; e in parte perché gli autoingannati trascorrono molta parte del loro tempo cercando di dare a loro stessi e agli altri delle giustificazioni a sostegno di ciò che credono autoingannevolmente e tipicamente producono una gran copia di storie di copertura. Questo oneroso sforzo epistemico sembrerebbe poter trovare un'agile spiegazione nella possibilità che gli autoingannati in fondo si rendono conto che quel che tanto accanitamente sostengono di credere in realtà non lo credono per davvero.

La caratterizzazione degli stati di credenza degli autoingannati è un compito non facile. Mele [2001] e Nelkin [2002], nelle loro analisi di tipo motivazionalista, assumono come dato incontrovertibile che gli autoingannati riescano effettivamente a credere quello che desiderano credere. Alcuni autori hanno tuttavia obiettato a Mele che c'è invece nell'autoinganno una tensione tra quel che dicono di credere e altri loro comportamenti che Mele non coglie e non spiega. Robert Audi [1988], per esempio, ha caratterizzato la tensione caratteristica dell'autoinganno nei termini della tendenza dell'autoingannato a dire una cosa mentre ne crede un'altra. Anche Kent Bach [1997] ha sostenuto che, anche se l'autoingannato non crede quella che è la verità, ha perlomeno dei sospetti su quale possa essere. Nell'autoinganno, insomma, a differenza che nella cecità o nella negazione, la verità sarebbe pericolosamente a portata di mano. Funkhouser [2005, p.

300] è d'accordo con Audi nel dire che gli autoingannati tendono a dire una cosa mentre ne credono un'altra. Anche per questa ragione Funkhouser ha posto l'accento sul comportamento di evitamento. Secondo Funkhouser, infatti, il comportamento di evitamento non solo mostrerebbe che l'autoingannato non è interessato alla realtà, come si è visto nel Paragrafo 2, ma anche che in fondo l'autoingannato non è convinto di quel che dichiara di credere [Cfr. Pedrini 2013, cap. 3]. Quando non ci fosse più il comportamento di evitamento e, più in generale, questa precipua tensione tra comportamenti manifesti e proferimenti, si avrebbe secondo Funkhouser non più un caso di autoinganno, ma di autoillusione, in cui il soggetto si è davvero definitivamente convinto della verità di quel che desidera credere.

Ma come dobbiamo caratterizzare quindi lo stato doxastico di colui che si autoinganna, se non è un caso di convinzione piena? La proposta specifica di Funkhouser è che colui che è autoingannato, a differenza di colui che è compiutamente autoilluso, crede (falsamente) di stare credendo quello che gli piacerebbe credere. In altre parole, l'autoingannato crede illusoriamente di avere formato la credenza che desidera, ma in realtà non è così. Ha insomma una credenza di second'ordine a cui non corrisponde la credenza di prim'ordine che ne dovrebbe essere l'oggetto.

La domanda che mi sono posta<sup>5</sup> è se la distinzione tra autoinganno e autoillusione sia davvero legittima ed esplicativamente necessaria per catturare la natura dell'autoinganno. Sono convinta che nell'autoinganno vi sia davvero quella tensione psicologica che sottolinea Funkhouser e che invece non è presente nell'autoillusione riuscita. Sono quindi anche d'accordo che non si debba confondere l'autoinganno con

---

<sup>5</sup> Per una versione estesa di questo problema, cfr. Pedrini 2013, cap. 3.

autoillusione, pena la perdita di una delle caratteristiche principali dell'autoinganno, ovvero la tensione psicologica in cui si trova l'autoingannato e il contrasto che gli osservatori colgono tra quel che l'autoingannato dice e ciò che fa. Nonostante questa iniziale convergenza di vedute, tuttavia, non credo che la tensione psicologica dell'autoinganno non possa temporaneamente risolversi, nel corso di un processo di autoinganno. Non è necessario, cioè, che un soggetto che si autoinganna non entri mai in una fase di autoillusione, in cui placa temporaneamente la sua tensione convincendosi che una certa proposizione è vera. L'autoillusione, infatti, non è un processo irreversibile: si può autoilludersi per un po', e poi tornare a dubitare di quello di cui ci si era temporaneamente convinti.

Quel che conta, insomma, è che l'autoinganno esibisca un *excursus* in cui la tensione rimane evidente nel lungo periodo, nel senso che chi ne è afflitto può sperimentare fasi in cui si guadagna la pace di una certa convinzione, che crede definitiva, per poi ripiombare nella fatica epistemica del negoziato con l'evidenza contraria, che può sempre ripresentarsi [Cfr. Pedrini 2013, cap. 3]. Data la possibilità di tali altalene doxastiche e data la nostra intuizione che queste altalene sono senz'altro possibili nei processi di autoinganno, c'è meno ragione di cercare di caratterizzare l'autoinganno come un processo in cui non si raggiunge mai il convincimento totale che *p*. Alla luce di questa considerazione, viene anche da pensare che il fatto che Funkhouser non accetti questa possibilità esplicativa sia ciò che lo induce a elaborare la sua teoria della credenza falsa di second'ordine. Ma se l'autoinganno può toccare anche fasi di credenza di prim'ordine piena, allora la spiegazione di Funkhouser diventa più speciosa e sembra dipendere dall'aver tracciato una distinzione molto forte tra autoillusione e autoinganno

che lo costringe quindi a cercare caratterizzazioni dell'autoinganno in cui non si raggiunga mai la credenza piena che  $p$ .

##### 5. CHE COS'È CHE L'AUTOINGANNATO NON SA DI SÈ?

Un'intuizione generalmente diffusa quando si discute di autoinganno è quella che ci fa pensare che quando qualcuno si autoinganna non c'è soltanto una verità del mondo che costui nega a se stesso e non vede, bensì c'è anche, più in generale, qualcosa di sé che non coglie. Il problema è quello di dare una caratterizzazione coerente e convincente di cosa sia ciò che l'autoingannato non vede di sé e del perché non lo vede. Funkhouser, come abbiamo visto, direbbe che quel che l'autoingannato non sa di sé è che non ha la credenza di prim'ordine che pensa, sbagliandosi, di avere. Altri autori, simpatizzanti nonostante tutto per la condizione doxastica espressa dal vecchio paradosso statico che si è visto nel Paragrafo 1, propendono per l'idea che l'autoingannato non sa di avere due credenze contraddittorie, di cui una non trasparente e su cui quindi il soggetto non ha autorità [Bilgrami 2006, pp. 140 e segg.]. L'autoingannato, cioè, crede genuinamente che  $p$  (la credenza autoingannevole) e crede anche genuinamente che  $\text{non-}p$  (la credenza vera su come le cose effettivamente stanno); solo che, mentre sulla credenza che  $p$  ha autorità e può quindi riferire correttamente di averla (ha cioè una credenza di second'ordine vera sulla presenza nella sua psicologia di tale credenza di prim'ordine), sulla credenza che  $\text{non-}p$  la sua autorità fallisce, dato che questa credenza non è trasparente, non sa di averla e riporterebbe di non averla (avrebbe cioè, una credenza di second'ordine falsa circa l'assenza di tale credenza di prim'ordine nella sua psicologia).

Ho già detto come la penso rispetto alla teoria di Funkhouser. Da quell'argomento si può facilmente evincere che non trovo necessaria a cogliere un caso di autoinganno neppure l'analisi di Bilgrami. Se adottiamo infatti la prospettiva diacronica dell'autoinganno che ho raccomandato nel precedente paragrafo, possiamo attribuire a un soggetto la credenza che  $p$  al tempo  $t$ , poi la credenza che non- $p$  al tempo  $t'$  (cioè il soggetto ha cambiato idea), e ancora di nuovo  $p$  al tempo  $t''$  (il soggetto ha nuovamente cambiato idea, data l'instabilità della sua convinzione), e così via. Oppure possiamo anche aver ragione di sostenere che ad un certo tempo  $t^n$  il soggetto non sta davvero credendo né  $p$  né non- $p$ , bensì sta solo illusoriamente credendo di crederlo, come vuole Funkhouser, per poi magari ritornare a credere non- $p$ , oppure magari invece convincersi temporaneamente proprio che  $p$ ; ma se non avessimo alcuna ragione per sostenere che siamo in presenza di un caso di fallimento dell'autorità, non perderemmo con questo il fenomeno dell'autoinganno. L'altalena doxastica di cui soffre l'autoingannato sembra sufficiente a restituirci un caso paradigmatico di autoinganno senza per forza dover ricorrere a nessuna versione del paradosso statico, che ritorna *in auge* se si caratterizza l'autoinganno come un caso di fallimento dell'autorità del soggetto rispetto a una credenza che sta in contraddizione con un'altra parimenti posseduta.

Se il fallimento dell'autorità e la condizione psicologica di doppia credenza non sono necessari per avere autoinganno, c'è da pensare che allora il fallimento dell'autoconoscenza, che tuttavia continuiamo a considerare tipico dell'autoinganno, debba risiedere da qualche altra parte. A questo proposito, un'interessante teoria che è stata elaborata è quella avanzata da Dion Scott-Kakures [2002]. Secondo Scott-Kakures quello che principalmente non ci è noto nell'autoinganno e che di noi ci sfugge è che

siamo preda di meccanismi motivazionali distorsivi che operano sulla nostra cognizione. Il mancato accesso a quello che ci sta complessivamente succedendo nell'autoinganno, cioè che siamo in balia dei *bias* cognitivi innescati dai desideri che le cose stiano in un certo modo, costituisce il fallimento dell'autoconoscenza caratteristico dell'autoinganno.

In Pedrini 2013 [Capitolo 4] ho anche affrontato estesamente una domanda collegata che considero interessante, e cioè se l'aver accesso a quel che ci sta succedendo mentre ci autoinganniamo sia sufficiente a farci fuoriuscire dall'autoinganno. Sostengo lì infatti che potrebbe non essere sufficiente.

#### 6. COME DOBBIAMO GIUDICARE L'AUTOINGANNO?

C'è un altro tema che considero centrale per la riflessione filosofica sul fenomeno dell'autoinganno. Questo tema non è mai stato propriamente al centro del dibattito analitico contemporaneo, il quale si è concentrato soprattutto sulle questioni affrontate finora (come è possibile un fenomeno come l'autoinganno, qual è il meccanismo che lo genera, come deve essere descritto, e così via). Ma non c'è solo il fenomeno dell'autoinganno per come si genera nel soggetto che lo vive che deve interessarci come filosofi. Deve interessarci anche il modo in cui l'autoinganno viene giudicato, ma soprattutto il modo in cui *deve* essere giudicato, da un osservatore che assiste a questo fenomeno quando si verifica nella psicologia di un altro con cui si desidera entrare in una relazione improntata al rispetto. Vi sono diverse proposte normative attualmente disponibili nel dibattito filosofico circa l'atteggiamento che dovremmo avere verso un autoingannato, in qualità di osservatori interessati a una relazione rispettosa.

Secondo un autorevole studioso di menzogna, inganno e autoinganno, Clancy Martin, curatore di un'interessante raccolta di saggi pubblicata nel 2009 da Oxford University Press [*The Philosophy of Deception*], è raccomandabile «riconoscere l'importanza delle false credenze, e persino delle bugie, per lo sviluppo umano», rendersi conto che «inganno e autoinganno sono importanti per l'automantenimento» dell'individuo, che «le bugie e gli autoinganni sono cruciali per le nostre interazioni sociali», che c'è ragione di sospettare che «i filosofi fissino un prezzo troppo alto per la verità» e che, alla luce di queste considerazioni, sia discutibile che si debba riservare all'autoinganno una condanna di tipo morale [Martin 2009, p. 5]. Martin si appresta anche ad aggiungere che sono «i motivi che stanno dietro all'inganno, e non l'inganno in sé ciò che troviamo moralmente biasimevole» [*ibid.*] e in qualche modo lascia intendere che la stessa conclusione dovrebbe essere estesa anche al caso dell'inganno verso se stessi. Ci ricorda nello stesso contesto il monito di Robert C. Solomon, a sua volta ispirato a Nietzsche, che probabilmente gli inganni sono un male necessario. Il progetto più generale di Martin è quello di analizzare le relazioni tra inganno e autoinganno e a questa analisi si dedicano molti degli autori che contribuiscono nella raccolta sopra citata. Una cordata di loro, tuttavia, si occupa anche esplicitamente del problema di come dobbiamo valutare e reagire all'autoinganno, e qualcuno si spinge persino a dire che dovremmo incoraggiarlo, perché socialmente utile [ad esempio, A.O. Rorty, 2009], perché evolutivamente vantaggioso [cfr. Oliver, 2009], oppure perché «la verità è al servizio dei valori, e non viceversa» [cfr. Solomon, 2009]. Il quadro che ne emerge è quello di una corrente di studiosi che non solo pensa che l'autoinganno sia necessario, nel senso che esiste e non ce ne possiamo liberare, ma che sia anche qualcosa che ha le



sue ragioni (evolutive), la sua utilità (sociale e personale) e il suo valore (strumentale) e che, una volta che si sia compreso tutto questo, non solo non è giustificato giudicarlo negativamente in senso morale, ma non è neppure giustificato tentare di combatterlo o reagire a esso, da osservatori interessati a una relazione rispettosa con l'autoingannato, supponendo che qualcosa di essenziale sia stato sottratto allo sviluppo umano dell'individuo che lo vive. Anzi, il suo sviluppo umano, se seguiamo la lezione di Martin, si accresce, con l'autoinganno, invece che diminuire o soffrirne. Altri filosofi, in altri contesti, si sono chiesti che cosa c'è di speciale nell'aderire alla realtà, dopo tutto [per esempio Kirsch 2005, le cui conclusioni sono moderatamente scettiche sul primato dell'adesione alla realtà]; e c'è però stato anche chi ha sostenuto che la «politica di autoinganno» non può essere superiore a quella di adesione alla realtà per ragioni che hanno a che fare con l'autonomia del soggetto, che l'autoinganno farebbe diminuire, e che quindi «l'autoinganno non ci renderà felici» [Van Leeuwen 2009], nonostante l'opinione di una fiorente corrente di psicologi secondo cui l'autoinganno è essenziale alla salute mentale del soggetto [cfr. Taylor 1989].

Ritengo che, per dirimere la questione su come dovremmo comportarci di fronte a qualcuno che si autoinganna, non solo è importante chiarire se e perché la verità è superiore alla falsità, nella vita quotidiana, e alla luce di cosa lo sarebbe, ma anche dedicarsi a un'indagine sui «motivi» che scatenano l'autoinganno. Si è visto finora che molti sostengono che il desiderio che *p* sia il motivo scatenante dell'autoinganno, ma c'è ragione di pensare che questo desiderio è forse soltanto il «motivo prossimale» dell'autoinganno, dato che il desiderio che *p* può e deve essere a sua volta spiegato alla luce del progetto di vita e del carattere valutativo del soggetto che si autoinganna e che

lo intrattiene. Il regresso esplicativo che ci porta a guardare alle origini stesse del desiderio che *p*, oltre all'indagine sui valori che rendono possibile il suo dispiegarsi e il suo condizionare la cognizione, possono essere di aiuto per farci un'idea circa il modo in cui si dovrebbe reagire di fronte a qualcuno che si autoinganna.

Possono secondo me essere distinti tre complessi di motivi a monte della personalità che si autoinganna, che determinano altrettante varietà di casi e che dimostrano quale sia il rapporto del soggetto con la verità:

- 1) il dolore o il disagio che una certa ipotesi procura alla luce dei certi progetti di vita e dei valori del soggetto (è il caso di qualcuno che non voglia credere che la moglie lo tradisce);
- 2) il dolore oggettivo e cocente, e non relativo a specifici progetti di vita, ma legato alla natura umana in quanto tale, che procura l'ipotesi che le cose stiano in un certo modo (può essere il caso di una madre che non riesce a credere che suo figlio sia morto in un incidente aereo sei mesi prima);
- 3) certe forme di «fede» nel futuro, nelle proprie possibilità o nella forza dei propri sentimenti (anche se risulta meno chiaro che si tratti di casi di autoinganno vero e proprio).

Mentre nel primo caso abbiamo ragione di credere che per reagire con rispetto dobbiamo senz'altro invitare il soggetto a crescere sia psicologicamente sia moralmente, migliorando i suoi stessi progetti di vita, alla luce del valore dell'adesione alla realtà, nel secondo possiamo invece ragionevolmente reagire con pietà, e accettare che per evitare l'autoinganno sarebbe necessaria una crescita e una forza che non è legittimo, almeno entro certi tempi e limiti, chiedere al soggetto che si trova in quella condizione. Nel terzo caso, invece, è probabile che proveremo persino simpatia per il soggetto, sempre

che lo scopo per cui si autoinganna (se di autoinganno di tratta) non dipenda, ancora una volta, da moventi moralmente repressibili o da progetti di vita criticabili. In questo caso soltanto abbiamo ragione di considerare l'utilità sociale e personale dell'autoinganno, e la superiorità in termini utilitaristici e di valore evolutivo di una credenza falsa rispetto a una credenza vera.<sup>6</sup> Non c'è una reazione unica da raccomandare nei confronti dell'autoinganno, come ad esempio la sola simpatia o il solo biasimo morale, che le questioni di utilità vanno valutate caso per caso, alla luce di modelli di vita e dei valori dei soggetti coinvolti, e che i presunti vantaggi evolutivi dell'autoinganno sembrano pressoché limitati ai casi i cui un soggetto riesce a sopravvivere, invece che soccombere al dolore, ma che non sono altrettanto chiaramente connessi con gli altri casi.

#### BIBLIOGRAFIA

- Audi R. (1988), "Self-Deception, Rationalization, and Reasons for Acting",  
McLaughlin B.P., Rorty A.O. (a cura di), *Perspectives on Self-Deception*,  
University of California Press, Berkeley, pp. 92-120.
- Bach K. (1997), "Thinking and Believing in Self-Deception", *Behavioral and Brain Sciences*, 20, 1, p. 105.
- Bilgrami A. (2006), *Self-Knowledge and Resentment*, Harvard University Press,  
Cambridge, MA.
- Davidson D. (1985a), "Deception and Division", LePore E. e McLaughlin B.P. (a cura di), *Actions and Events*, Basil Blackwell, Oxford, pp. 138-148.
- Heil J. (1984) "Doxastic Incontinence", *Mind*, 93, 369, pp. 56-70.

---

<sup>6</sup> Ho offerto un'analisi di questi tre casi, corredandola con articolati esempi, in Pedrini 2013, cap. 5, a cui rimando per approfondimenti.

- Kirsch J. (2005), “What's So Great about Reality?”, *Canadian Journal of Philosophy*, 35, 3, pp. 407-428.
- Martin C. (2009), (a cura di), *The Philosophy of Deception*, Oxford University Press, Oxford.
- Mele A. (1986), “Incontinent Believing”, *Philosophical Quarterly*, 36, 143, pp. 212-229.
- Mele A. (1997), “Real Self-Deception”, *Behavioural and Brain Sciences*, 20, 1, pp. 91-102.
- Mele A. (1998), “Two Paradoxes of Self-Deception”, Dupuy J. (a cura di), *Self-Deception and Paradoxes of Rationality*, CSLI Publications, Stanford, CA, pp. 37-58.
- Mele A. (2001), *Self-Deception Unmasked*, Princeton University Press, Princeton.
- Nelkin D.K. (2002), “Self-Deception, Motivation, and the Desire to Believe”, *Pacific Philosophical Quarterly*, 83, 4, pp. 384-406.
- Oliver K. (2009), “Duplicity Makes the Man”, in Martin C. (a cura di), *The Philosophy of Deception*, Oxford, Oxford University Press, pp.104-17.
- Pedrini P. (2012), “What Does the Self-Deceiver Want?”, *Humana.Mente, Journal of Philosophical Studies*, 20 (a cura di Pedrini P.), pp. 141-157
- Pedrini P. (2013), *L'autoinganno. Che cos'è e come funziona*, Laterza, Roma-Bari
- Rorty A. (2009), “User-Friendly Self-Deception. A Traveller's Manual”, in Martin C. (2009), a cura di, *The Philosophy of Deception*, Oxford, Oxford University Press, pp. 244-259.

- Scott-Kakures D. (2002), “At ‘Permanent Risk’: Reasoning and Self-Knowledge in Self-Deception”, *Philosophy and Phenomenological Research*, LXV, 3, pp. 576-603.
- Solomon R.C. (2009), “Self-Deception, and Self-Deception in Philosophy”, in Martin C. (a cura di), *The Philosophy of Deception*, Oxford, Oxford University Press, pp. 15-36.
- Taylor S.E. (1989), *Positive Illusions: Creative Self-Deception and the Healthy Mind*, Basic Books, New York.
- Trope Y., Liberman A. (1996), “Social Hypothesis Testing: Cognitive and Motivational Mechanisms”, in Higgins E., Kruglanski A. (a cura di), *Social Psychology: Handbook of Basic Principles*, Guilford Press, New York, pp. 239-270.
- Van Leeuwen D.S.N. (2009), “Self-Deception Won’t Make You Happy”, *Social Theory and Practice*, 35, 1, pp. 107-132.

---

**AphEx.it è un periodico elettronico, registrazione n° ISSN 2036-9972. Il copyright degli articoli è libero. Chiunque può riprodurli. Unica condizione: mettere in evidenza che il testo riprodotto è tratto da [www.aphex.it](http://www.aphex.it)**

Condizioni per riprodurre i materiali --> Tutti i materiali, i dati e le informazioni pubblicati all'interno di questo sito web sono "no copyright", nel senso che possono essere riprodotti, modificati, distribuiti, trasmessi, ripubblicati o in altro modo utilizzati, in tutto o in parte, senza il preventivo consenso di AphEx.it, a condizione che tali utilizzazioni avvengano per finalità di uso personale, studio, ricerca o comunque non commerciali e che sia citata la fonte attraverso la seguente dicitura, impressa in caratteri ben visibili: "www.aphex.it". Ove i materiali, dati o informazioni siano utilizzati in forma digitale, la citazione della fonte dovrà essere effettuata in modo da consentire un collegamento ipertestuale (link) alla home page www.aphex.it o alla pagina dalla quale i materiali, dati o informazioni sono tratti. In ogni caso, dell'avvenuta riproduzione, in forma analogica o digitale, dei materiali tratti da [www.aphex.it](http://www.aphex.it) dovrà essere data tempestiva comunicazione al seguente indirizzo ([redazione@aphex.it](mailto:redazione@aphex.it)), allegando, laddove possibile, copia elettronica dell'articolo in cui i materiali sono stati riprodotti.

In caso di citazione su materiale cartaceo è possibile citare il materiale pubblicato su AphEx.it come una rivista cartacea, indicando il numero in cui è stato pubblicato l'articolo e l'anno di pubblicazione riportato anche nell'intestazione del pdf. Esempio: Autore, *Titolo*, <<[www.aphex.it](http://www.aphex.it)>>, 1 (2010).

---