

MARILENA ANDRONICO

DONALD DAVIDSON E LA RAGIONEVOLE IRRAZIONALITÀ DELL'AUTOINGANNO

L'autoinganno è considerato un comportamento diffuso nella vita quotidiana, che si manifesta soprattutto nelle situazioni cosiddette "difficili", come quando ci sembra di dover prendere atto di qualcosa che ci addolora profondamente o di dover prendere delle decisioni che non vorremmo affatto prendere. Per la filosofia, invece, l'autoinganno è in primo luogo un comportamento o una condizione irrazionale, la cui semplice descrizione non può essere separata dai principi di razionalità che si intendono far valere. In questo intervento prendo in considerazione la riflessione di Donald Davidson sul tema, sia perché essa rappresenta il termine di confronto obbligato per ogni speculazione sull'argomento, sia perché essa offre spunti di grande interesse alla discussione sulla razionalità e sui suoi limiti. Nella prospettiva di Davidson, infatti, l'irrazionalità dell'autoinganno può essere compresa sullo sfondo di una razionalità ampia e condivisa.

1. *Il problema*

Con l'espressione "autoinganno" si è soliti indicare un insieme di fenomeni che caratterizzano una condizione di scarsa trasparenza, se non addirittura di completa oscurità in cui un individuo può venire a trovarsi nei confronti dei propri pensieri e/o delle ragioni che lo

portano a compiere determinate azioni. Come tale, la condizione dell'autoinganno contravviene ad uno dei motti che la tradizione filosofica ha posto alle origini della propria ragion d'essere: il «conosci te stesso», iscritto sul tempio dell'oracolo di Delfi; e questo ci dà in parte la misura dell'interesse che esso ha da sempre suscitato per la riflessione filosofica¹, alla quale si è aggiunto – soprattutto a partire dal XX secolo – anche quello della riflessione psicologica. Sul versante filosofico, l'analisi dell'autoinganno ha interessato e tuttora interessa l'etica, l'epistemologia, la filosofia della mente e la filosofia dell'azione; sul versante delle prospettive psicologiche, la sua analisi ha sempre coinvolto e ancora coinvolge tanto la psicologia del profondo (o psicoanalisi), quanto la psicologia cognitiva.

Poiché l'argomento è estremamente vasto, in ciò che segue mi limiterò a presentare alcuni problemi generali che concernono l'autoinganno, prendendo in considerazione soprattutto il trattamento che di esso è stato fatto da Donald Davidson. Le sue riflessioni, infatti, restano un punto di riferimento imprescindibile tanto per coloro che ne prendono criticamente le distanze, quanto per chi avanza ulteriori sviluppi simpatetici, in quanto legano in modo profondo le discussioni su questo argomento a quelle sulla razionalità.

Ma che cosa si intende con l'espressione 'autoinganno'? Uno dei problemi che concernono il fenomeno indicato da questo termine è che la risposta a questa domanda dipende in larga misura dalla teoria, o più semplicemente, dalla prospettiva che si adotta per descriverlo; e questo può valere fino al punto che è possibile arrivare a negare che realmente esista qualcosa che sia degno di essere così denominato. Come ha osservato Amélie Oksenberg Rorty, quando ci si accinge ad affrontare questo tema, si presentano subito dei problemi metodologici molto seri, poiché i vari fenomeni di cui l'autoinganno potrebbe consistere rappresentano «una sezione arbitrariamente selezionata di uno spettro di attività tra loro

¹ Poiché chi inganna se stesso è lontano dal conoscere se stesso, già Platone faceva dire a Socrate che «lasciarsi trarre in inganno da solo a solo è la cosa peggiore fra tutte. Poiché quando colui che sta per ingannarti non se ne sta lontano neanche un pochino, ma ti sta sempre al fianco, come potrebbe non essere una sciagura?» (*Cratilo* 428d).

strettamente connesse che riguardano forme ritualizzate di auto-manipolazione; la loro identificazione presuppone teorie relative ai pattern normali della salienza concettuale, percettiva ed emotiva, a norme di razionalità e di trasparenza»². In sostanza, secondo A.O. Rorty (1994), per poter parlare di autoinganno bisogna avere chiare in mente situazioni cosiddette “normali” in cui non ci si autoinganna: cosa che se da un lato appare ovvia, dall'altro non è del tutto scontata, poiché la descrizione di tale normalità raramente è del tutto neutrale, e facilmente dipende dalle norme di razionalità che una data prospettiva filosofica intende far valere.

Stando così le cose, una strategia da molti adottata, per provare a rispondere alla domanda su cosa sia l'autoinganno, consiste nell'assumere come punto di partenza il fatto che nella vita quotidiana si danno casi di autoinganno e nel portare esempi di tali casi. Si tratta di individuare delle situazioni in cui diremmo di qualcuno che inganna se stesso, oppure dei casi in cui qualcuno direbbe di se stesso che si è ingannato da solo. Vediamo alcuni di questi esempi: 1) in una coppia uno dei due coniugi, pur avendo a disposizione tutte le evidenze per capire che l'altro ha una relazione extraconiugale, si convince dell'opposto; 2) una persona ha consapevolezza di non essere sufficientemente preparata per superare il test di guida e tuttavia si convince dell'opposto, presentandosi all'esame³; 3) una persona sa perfettamente di essere in sovrappeso e nonostante ciò, di nuovo, si convince del contrario e mangia una grossa fetta di torta al termine di un lauto pasto⁴; 4) una madre, il cui figlio è in carcere perché ha commesso un orrendo crimine, si ostina ad affermare che il figlio è un bravo ragazzo. Tutti questi sono esempi tipici di autoinganno e prevedono: (a) che un soggetto S creda che p è vero, ma anche, in un senso che richiede per l'appunto di essere spiegato, (b) che il soggetto S creda, o meglio, sappia anche che p è falso (cioè, che creda o sappia che non- p)⁵.

² Rorty A.O. (1994), p. 212.

³ Questo è l'esempio di Carlos, discusso in Davidson (1985).

⁴ Questo esempio è discusso in Evnine (1991).

⁵ Ricordo che quando si usa p si intende dire che p sta per un certo contenuto di credenza espresso da una proposizione, del tipo: «mia moglie mi è fedele», «sono sufficientemente preparato per superare l'esame di guida», «io non sono in sovrappeso», «mio figlio è un bravo ragazzo», ecc.

Una domanda legittima che si presenta a questo punto riguarda il peso che l'inganno ha o può avere negli esempi appena menzionati. Così, per quanto sia un'ovvietà, vale comunque la pena di ricordare quale sia la struttura tipica dell'inganno interpersonale:

- (a) il soggetto S sa che p è vero;
- (b) il soggetto S intende indurre nell'interlocutore I la credenza che p è falso (o che non- p);
- (c) il soggetto S mente a I (dice a I o si comporta con I come se egli S credesse che p è falso (o che non- p));
- (d) I acquisisce la credenza che p è falso (o la credenza che non- p)⁶.

Se assumiamo l'inganno interpersonale come modello per l'autoinganno, abbiamo che il soggetto S occupa anche la posizione di I , egli cioè è al tempo stesso ingannatore e ingannato; o, potremmo anche dire, egli è sia l'agente, sia il paziente, sia colui che perpetra l'inganno, sia colui che lo subisce. Se l'impiego del termine 'inganno' nella parola composta 'autoinganno' ha da essere qualcosa di più di un semplice impiego metaforico, allora proprio come nell'inganno anche nell'autoinganno possiamo dire che vi è un'*intenzione* di indurre una credenza falsa.

Le due descrizioni fin qui fornite mettono in luce che una situazione di autoinganno può contenere due tipi di paradossi:

- (i) uno riferito allo stato di credenza (e per questa ragione nella letteratura sull'argomento viene chiamato «paradosso doxastico»): S crede che p e contemporaneamente crede che non- p
- (ii) l'altro riferito all'intenzione di indurre una credenza (e per questa ragione nella letteratura sull'argomento viene chiamato «paradosso dinamico»): S crede che p e induce se stesso a credere che non- p .

Difficile dire, naturalmente, quale dei due paradossi sia più imbarazzante: quello che prevede un soggetto capace di intrattenere

⁶ Vorrei qui osservare, sulla scia di Demos (1960), che se è possibile ingannare qualcuno senza avere l'intenzione di farlo, non è invece possibile mentire a qualcuno senza averne l'intenzione. Pertanto, il tipo di inganno a cui ci si riferisce nelle discussioni sull'autoinganno è quello che lo avvicina alla menzogna. Per Demos (1960), ad esempio, l'autoinganno esiste quando una persona mente a se stessa, cioè quando persuade se stessa a credere qualcosa che *sa* non essere (vero) (p.588).

contemporaneamente due credenze contraddittorie (doxastico), o quello che prevede un soggetto che pur credendo o sapendo che p , deliberatamente si induce a credere che non- p . Comunque sia, in quanto implica una condizione paradossale, l'autoinganno è considerato fare parte di una più ampia famiglia di comportamenti irrazionali come l'*akrasia* o debolezza della volontà e il *wishful thinking*, espressione questa per la quale non vi è una traduzione italiana canonica, e che indica la situazione per cui si agisce come se le cose stessero nel modo in cui si desidera che stiano⁷.

Di certo sarebbe lecito rifiutare la descrizione dell'autoinganno fin qui data e decidere di trattare gli esempi sopra descritti da una prospettiva differente e con strumenti diversi, ad esempio da una qualche prospettiva psicologica, dotata di diverse griglie concettuali, allo scopo di rappresentarli come situazioni che non implicano né paradossalità, né irrazionalità. Sintetizzando e semplificando, vi sono almeno due vie che tendono a diluire o addirittura a eliminare l'aspetto di irrazionalità dell'autoinganno: una è quella che insiste sulla dicotomia tra conscio e inconscio; l'altra è quella che, senza fare ricorso a tale dicotomia, nega comunque la componente intenzionale del fenomeno. Per la prima via, se la credenza o la conoscenza che non- p è inconscia o pre-conscia, in un qualche senso da definire, allora essa non può essere considerata in contraddizione con la credenza conscia che p ⁸. Per la seconda via, si ipotizzano strutture, istanze o tropismi sub-personali che operano in modo tale che il soggetto giunga, per così dire, quasi naturalmente a formare la credenza (errata) che p , senza che ciò implichi né l'intenzione di formarla, né la contemporanea formazione della credenza contraria. In particolare, questa seconda via non assume l'inganno

⁷ Cfr. Davidson (1982), p. 19; e anche Pedrini (2012) che, nell'Introduzione alla raccolta da lei curata, ricorda come nella letteratura contemporanea sull'argomento questi fenomeni sono etichettati come casi di «motivata irrazionalità».

⁸ Il caso paradigmatico di questo modo di affrontare il problema si trova in Freud, per il quale il fenomeno che qui stiamo prendendo in considerazione rientra tra le strategie di rimozione che l'inconscio mette in atto. Seguono questa strada anche molte prospettive cognitiviste che fanno ampio ricorso all'idea di strutture cognitive inconscie. Per orientarsi sull'argomento si vedano Marraffa (1999), Pedrini (2006) e Pedrini (2012).

interpersonale come modello dell'autoinganno⁹.

Tuttavia, vorrei far notare che, così facendo, se da un lato otterremmo il risultato, per certi aspetti vantaggioso, di dissolvere il problema nella varietà dei suoi possibili componenti psichici o sub-personali, dall'altro lato, toglieremmo a noi stessi l'opportunità di riflettere proprio su uno dei temi più classici della filosofia dai tempi di Socrate: mi riferisco al tema della razionalità, o per meglio dire al tema dell'intreccio di relazioni che, come filosofi, riteniamo possano, o meglio, debbano sussistere tra conoscenza, credenza, ragione e azione¹⁰. Quando Socrate nel *Protagora* afferma che nessuno fa il male o pecca volontariamente, egli non si limita ad esprimere una determinata posizione in etica, ma manifesta l'esigenza, condivisa anche dal senso comune, che sia possibile rendere conto delle azioni degli uomini a partire da ciò che essi credono (sanno) o affermano di credere, e getta così le basi per tutte le future discussioni filosofiche sulla relazione tra credenza (conoscenza, pensiero) e azione, di cui quelle sull'autoinganno rappresentano soltanto una parte.

Per questa ragione, dunque, mi attengo alla rappresentazione dell'autoinganno come di una condizione paradossale e irrazionale e, seguendo Davidson, preferisco cercare di chiarire cosa significhi qui "irrazionale" e come possa determinarsi una simile situazione.

2. Razionalità

Nel saggio *Paradossi dell'irrazionalità*, Davidson afferma che «l'idea di un'azione, di una credenza, di un'intenzione e di un'inferenza o di un'emozione irrazionale è un paradosso. Perché l'irrazionale non è meramente il non-razionale, ciò che sta al di fuori dell'ambito del razionale; l'irrazionale è un fallimento che si verifica all'interno della casa della ragione. Quando Hobbes dice che soltanto l'uomo ha "il privilegio dell'assurdità", egli lascia intendere che soltanto una

⁹ Sono anti-intenzionalisti, ad esempio, Johnston (1988) e Mele (2001). Una posizione critica dell'anti-intenzionalismo si trova in Pedrini (2006).

¹⁰ Il problema, ad avviso di chi scrive, è che tali relazioni non possono non essere vincolanti e quindi comunque presenti nelle descrizioni che muovono da prospettive psicologiche e cognitive, che si servono di risultati sperimentali, poiché tali relazioni concorrono a determinare il tipo di griglia entro cui leggere quei risultati.

creatura razionale può essere irrazionale. L'irrazionalità è un processo o uno stato mentale – un processo o uno stato razionale – che fallisce» (1982, p. 18). Con quello che può sembrare un gioco di parole, Davidson ci dice che la nozione stessa di irrazionalità è paradossale, nel senso che può essere compresa e considerata come tale solo su uno sfondo molto ampio e condiviso di razionalità. L'irrazionale va distinto da ciò che è a-razionale, in quanto non costituisce un'uscita dalla sfera della ragione, ma al contrario, rappresentando – potremmo dire: semplicemente – un fallimento nella casa della ragione, l'irrazionale è sempre riconducibile alla ragione; l'irrazionale è razionalizzabile.

Ma di quale ragione o di quale razionalità stiamo qui parlando? Una risposta un po' sbrigativa, ma con la quale Davidson sarebbe d'accordo, è che stiamo parlando dell'unica ragione o dell'unica razionalità che conosciamo, cioè della nostra. Questo, a tutta prima, può apparire riduttivo, ma non lo è affatto se seguiamo la sua riflessione su ciò che ci permette di interpretare un soggetto umano come un essere che desidera, crede, teme, spera, ecc.¹¹, nonché su ciò che ci permette di rendere conto delle sue azioni e soprattutto del nesso che sussiste o che deve sussistere tra le sue azioni e le sue credenze (o pensieri).

In primo luogo, Davidson precisa che ciò che distingue un'azione da un qualsiasi altro tipo di evento è la sua dimensione intenzionale, che permette (i) sia all'agente di fornire ragioni per le sue azioni, come quando ad esempio, egli dice: «ti saluto, perché esco»; (ii) sia all'interprete di comprendere il comportamento dell'agente, attribuendogli delle intenzioni per le sue azioni, intenzioni tali che l'agente stesso può riconoscere come proprie, come nel caso in cui l'interprete dice o pensa: «mi ha salutato, perché esce»¹². E qui vorrei insistere sul fatto che la possibilità di fornire ragioni per le azioni è ciò che rende queste ultime sempre razionalizzabili, vale a

¹¹ Cioè, come qualcuno a cui attribuiamo stati mentali e/o atteggiamenti proposizionali.

¹² Sempre in Davidson (1982), leggiamo: «È perché ha una ragione per ciò che fa che possiamo dare l'intenzione con cui ha agito. E come ogni azione intenzionale, la sua azione può essere spiegata facendo riferimento alle credenze e ai desideri che l'hanno causata e che le hanno dato origine» (p. 27). Cfr. anche Davidson (1980), trad. it., p. 41.

dire sempre riconducibili a ragioni. Secondo Davidson, è sempre possibile da una qualche prospettiva individuare delle ragioni in base a cui un agente ha agito (o attribuire ad un agente delle ragioni in base alle quali egli ha agito), non importa quanto estreme o bizzarre siano le azioni da interpretare. Data questa riflessione sul rapporto tra azioni e ragioni, bene si comprende in che cosa consista propriamente il paradosso dell'irrazionalità:

Il paradosso soggiacente dell'irrazionalità è questo: se lo spieghiamo troppo bene, lo trasformiamo in una forma occulta di razionalità; mentre se assegniamo l'incoerenza con eccessiva scioltezza, molto semplicemente compromettiamo la nostra capacità di diagnosticare l'irrazionalità col togliere quello sfondo di razionalità che è necessario per giustificare una qualsiasi diagnosi. (Davidson 1982, p. 37).

In secondo luogo, Davidson non si stanca mai di sottolineare che «se non potessimo interpretare gli altri come esseri che condividono una gran parte di ciò che costituisce il nostro senso comune non saremmo in grado di identificare nessuna delle loro credenze, intenzioni e desideri, e nessuno dei loro atteggiamenti propositivi» (ID., p. 36). Tanto per cominciare, se attribuiamo in modo intelligibile ad una persona la credenza che un pezzo di ghiaccio si sta sciogliendo, non possiamo farlo in un contesto di isolamento totale, non possiamo cioè non attribuirgli anche altre credenze (vere) circa la natura del ghiaccio, circa le sue proprietà fisiche connesse all'acqua, al freddo, alla solidità, ecc. (Cfr. ID., p. 37). È questa la concezione olistica del mentale, secondo cui una credenza ne implica molte altre: una credenza è ciò che è, in quanto occupa una certa posizione all'interno di un reticolo più ampio di credenze. Ma la rappresentazione dei vincoli che la nostra razionalità ci impone nel valutare quella altrui si fa ancora più impegnativa quando Davidson prova a rendere conto di ciò che ci autorizza, proprio dalla prospettiva dell'olismo del mentale, a ritenere comunque irrazionali certi comportamenti o credenze, senza che ciò implichi collocarli del tutto al di fuori della ragione, cioè senza considerarli a-razionali. Davidson dichiara, infatti, che il tipo di irrazionalità a cui è interessato non consiste nel fatto che qualcuno non si conformi ai valori e/o alle norme a cui un altro trova ragionevole e razionale conformarsi. Per fare un esempio: egli non vede problema a considerare irrazionale uno che trovi ragionevole scalare l'Everest in condizioni climatiche

sfavorevoli, perché, poniamo, intende mettere alla prova la sua preparazione atletica in condizioni estreme, o perché misura il valore degli uomini in funzione del coraggio che essi mostrano nello sfidare gli elementi naturali¹³. L'irrazionalità che interessa Davidson è quella che si origina da, o che consiste in un'*incoerenza interna* al sistema di credenze e di principi ai quali un agente normalmente si attiene (o dice di volersi attenere). In particolare, Davidson individua due principi, la cui violazione autorizza un interprete ad attribuire incoerenza interna o irrazionalità all'agente: il *principio di continenza* e il *principio dell'evidenza totale*.

Il *principio di continenza* è un principio normativo, secondo cui non si dovrebbe compiere intenzionalmente un'azione quando si giudica, sulla base di ciò che si ritengono essere tutte le considerazioni disponibili, che un corso alternativo e accessibile di azioni sarebbe stato meglio. Questo principio «prescrive un genere fondamentale di coerenza nel pensiero, nell'intenzione, nella valutazione e nell'azione»¹⁴ e una persona che agisce in accordo con questo principio si dice che possiede la virtù della continenza. Tale principio è tipicamente violato nei comportamenti cosiddetti *acratichi*, cioè in quei comportamenti caratterizzati dalla *debolezza della volontà* (*weakness of the will*). Facciamo un esempio banale: so che fumare fa male; ritengo che sia bene non fumare; tuttavia fumo. In questo caso, è lecito anche dire che il mio è un comportamento incontinente, oppure che io sono una persona incontinente.

Il *principio dell'evidenza totale* è un principio normativo che Davidson ci dice riprendere quel che Hempel e Carnap hanno chiamato il *requisito dell'evidenza totale per il ragionamento induttivo*¹⁵. Secondo tale principio, quando dobbiamo decidere tra una serie di ipotesi che si escludono reciprocamente, dobbiamo dare

¹³ Così si esprime Davidson (1982): «Il tipo di irrazionalità che costituisce effettivamente un problema concettuale non è quello posto da chi non crede o non sente o non fa ciò che noi riteniamo ragionevole, ma quello posto dall'incoerenza o dalla contraddittorietà all'interno della struttura di credenze, comportamenti, emozioni, intenzioni, azioni di una singola persona. Gli esempi possono essere il *wishful thinking*, l'agire in maniera contraria al proprio meditato giudizio, l'auto-inganno, il credere a qualcosa che si ritiene invalidato dal peso dell'evidenza» (p. 19).

¹⁴ Cfr. Davidson (1985), p. 139.

¹⁵ Ivi, p. 140.

credito all'ipotesi più fortemente *supportata da tutta l'evidenza rilevante a disposizione*. E qui non è importante il tipo di evidenza a cui ci si attiene; ciò che conta è che qualcosa *sia considerato dall'agente stesso come un'evidenza rilevante per qualcos'altro*. Questo principio è tipicamente violato nei *ragionamenti* caratterizzati dalla *debolezza della giustificazione* (*weakness of the warrant*). Facciamo un esempio molto semplice: ho tutta l'evidenza a disposizione per formarmi la credenza corretta e condivisa che, data la quantità di pioggia caduta, un certa strada è diventata impraticabile, e formo invece l'ipotesi per cui, nonostante la quantità di pioggia caduta, non c'è alcun problema a percorrere quella strada.

Di nuovo, notiamo che questi due principi sono tali innanzitutto per l'interprete; essi cioè sono costitutivi della razionalità per lui e per la sua comunità di appartenenza, cosicché egli non può non aspettarsi che – *almeno in larga misura* – essi vengano rispettati anche da un agente a cui viene attribuita razionalità, cioè da parte di qualcuno di cui egli ritiene spiegabile il comportamento¹⁶. Avrebbe senso, si chiede Davidson, affermare che una persona che non accetta, poniamo, il requisito dell'evidenza totale e il cui pensiero non soddisfa tale requisito sia irrazionale per gli standard di un'altra persona che invece accetta il requisito, ma non per i suoi propri standard? Oppure dovremmo rendere l'incoerenza interna una condizione necessaria dell'irrazionalità? La risposta di Davidson è che non è facile vedere come queste due questioni possano essere separate, visto che la coerenza interna è essa stessa una norma fondamentale del pensiero e aggiunge:

Nel caso delle norme fondamentali le questioni non possono essere chiaramente separate. Infatti quanto più strano (*striking*) un caso di incoerenza interna possa apparire ad un estraneo, tanto meno l'estraneo, nel tentativo di spiegare l'apparente aberrazione, può fare uso di una presunta distinzione tra le sue norme e quelle della persona osservata. (Davidson 1985, p. 141)¹⁷.

¹⁶ Sull'analogia tra le riflessioni di Davidson sulle attribuzioni di razionalità e le sue riflessioni sulle attribuzioni di significato, in una situazione di traduzione radicale, si veda Hällén (2011).

¹⁷ Sempre per la medesima ragione, Davidson precisa: «Differenze relativamente piccole prendono forma e sono spiegate su uno sfondo di norme condivise, ma vere e proprie deviazioni dagli standard fondamentali della

3. L'autoinganno

Una volta chiariti i termini in cui si articola il discorso di Davidson sulla razionalità, vediamo come l'autoinganno propriamente si distingua da altre forme di irrazionalità e anche che cosa lo avvicini ad esse. In primo luogo, l'autoinganno si distingue dall'*acrasia* o debolezza della volontà poiché quest'ultima prevede un fallimento o un'incoerenza nell'azione (o anche, nella relazione tra credenza e azione), laddove l'autoinganno prevede soprattutto un'incoerenza cognitiva, cioè un'incoerenza tra credenze. Connesso a ciò è il fatto che non è richiesto ad una persona con comportamenti *acratichi*, o incontinenti di mentire a se stessa, o ad altri. Basti pensare che si può essere pienamente consapevoli del fatto che fumare fa male alla salute e tuttavia fumare, avendo chiaro in mente che ci si sta comportando male.

D'altro canto, *acrasia* e autoinganno, nella prospettiva di Davidson, sono accomunati dal fatto che la loro spiegazione si avvale di una medesima rappresentazione del mentale (o di una medesima teoria della mente), la quale prevede – tra le altre cose – che si diano parti o aree della mente, per così dire, tra loro distinte, che operano come strutture semiautonome, ciascuna delle quali è rappresentabile come un reticolo di credenze e di stati mentali che si interseca e si sovrappone ad altre strutture¹⁸; è in virtù di tali relazioni di sovrapposizione e di intersezione che possono generarsi dei conflitti, vale a dire, che possono originarsi situazioni di incoerenza interna alla mente di un singolo individuo, vista come un tutto. Per quanto Davidson riconosca a Freud il merito di avere

razionalità, sono più adatte ad essere negli occhi dell'interprete che nella mente dell'interpretato» (Ib. 1985, p. 142).

¹⁸ Davidson (1982), alla n. 5 della trad. it. così si esprime in proposito: «Qui come altrove la mia versione altamente astratta della partizione della mente devia da quella di Freud. In particolare, non ho nulla da dire sul numero o la natura delle divisioni della mente, né sulla loro permanenza o eziologia. In questa sede mi interessa esclusivamente difendere l'idea dei compartimenti mentali, e sostenere che essi sono necessari se intendiamo spiegare una comune forma di irrazionalità. Dovrei forse sottolineare che espressioni quali "partizione della mente", "parte della mente", "segmento", ecc. sono fuorvianti se lasciano intendere che ciò che appartiene a una divisione della mente non può appartenere a un'altra. Il quadro che desidero prevede dei territori che si sovrappongono» (p. 33).

affrontato il tema dell'irrazionalità suddividendo la sfera psichica in ambiti distinti, che variamente interagiscono tra loro, diversamente da Freud, egli non ritiene tuttavia necessario ipotizzare che qualcuno degli ambiti in cui la mente può essere suddivisa sia inconscio e inaccessibile, ed agisca come una forza cieca ed oscura. Anzi, dal punto di vista di Davidson l'inaccessibilità dell'inconscio postulata da Freud ha anch'essa qualcosa di paradossale, visto che non solo lo scopo della terapia, bensì quello dell'intera impresa psicoanalitica è di individuare le ragioni e le cause di ogni singola manifestazione del comportamento umano, soprattutto di quelle più stravaganti. Vi è dunque un ulteriore merito che Davidson riconosce a Freud, ed è quello di avere «aumentato enormemente il numero e la varietà dei fenomeni che possono essere considerati razionali: perché si scopre che ci sono delle ragioni per le nostre dimenticanze, i nostri sbagli linguistici, le nostre paure esagerate» (Davidson 1982, p. 21).

Per spiegare un comportamento acratice, dunque, è sufficiente ipotizzare che vi siano delle parti semi-autonome della mente, una delle quali «tutto considerato, trova un certo tipo di azioni migliori», mentre un'altra spinge (*prompts*)¹⁹ verso un altro tipo di azione. «Da entrambi i lati» – afferma Davidson – «vuoi il lato del giudizio meditato, vuoi il lato dell'interazione e azione incontinente, vi è una struttura portante di ragioni, di credenze interrelate, di aspettative, assunzioni, atteggiamenti e desideri» (p. 34). Come si è detto, ciò che tra le due parti semi-autonome spinge verso un altro tipo di azione non ha bisogno di essere concepito come un che di inconscio²⁰, ma (soltanto) come una causa che non è, o meglio, non è ancora una ragione; e più precisamente come una causa mentale, cioè come uno stato mentale (di desiderio, di timore, ecc.) capace di determinarne o causarne un altro.

Non è certo questa la sede per entrare nel merito del dibattito sulla distinzione tra cause e ragioni, né in particolare su quello scaturito dall'uso peculiare che di essa ha fatto Davidson. In questo contesto, mi sembra sufficiente ricordare che, dato un fenomeno (sia

¹⁹ Cfr. Davidson (1982), p. 181 della vers. ingl.

²⁰ «Il tipo normale di *acrasia* è quello in cui l'agente sa cosa sta facendo e perché lo sta facendo, e sa che non è il meglio e perché non lo è. Egli riconosce la propria irrazionalità. Se tutto ciò è possibile, allora la descrizione non può essere resa insostenibile col supporre che a volte alcuni dei pensieri o desideri coinvolti sono inconsci» (Davidson 1982, p. 39).

esso fisico o di altra natura), è lecito asserire delle sue cause che esse sono sconosciute, mentre non è lecito asserire che le ragioni per cui qualcuno afferma che il fenomeno si è verificato possano essere ignote a chi fa tale affermazione. Introdurre delle cause che non sono ancora ragioni è dunque un modo per fare svolgere ad esse il ruolo che in Freud svolge l'inconscio, senza tuttavia doverne postulare l'impenetrabilità. D'altronde, quando parla di divisione (o partizione) della mente, ciò che Davidson propone «non corrisponde per natura o funzioni all'antica metafora della lotta tra la Virtù e la Tentazione o la Ragione e la Passione» (ibid.); ciò che la sua prospettiva prevede è soltanto che vi siano «degli elementi organizzati, all'interno di ciascuno dei quali vi è un notevole grado di coerenza, e dove un elemento può operare sull'altro secondo le modalità della causalità non-razionale» (ibid.)²¹. Di nuovo, che la causalità non-razionale possa sempre, in linea di principio, essere razionalizzata, cioè ricondotta a ragioni, mina anche la rappresentazione classica di ciò che viene chiamato "passione" e/o "tentazione": a poco serve concepirle come forze oscure ed indomabili, rispetto alle quali gli uomini non possono che essere passivi²², quando ciò che interessa al filosofo è ampliare il raggio della comprensione dell'agire umano, riconducendolo alle sue ragioni.

Dato l'assunto della divisione (o partizione) della mente, Davidson ridecrive l'autoinganno come una situazione in cui: (a) un soggetto S crede che *p*; (b) tale credenza, in quanto è parte integrante di un sistema di altre credenze e stati mentali (la struttura semi-autonoma di cui si è detto) che si interseca con altri sistemi di credenze e stati mentali, causa o determina nel soggetto S uno stato mentale (di timore, o di desiderio) che lo porta a contravvenire ad un qualche principio di razionalità, al quale egli invece *normalmente* si attiene;

²¹ L'esigenza qui espressa da Davidson è stata criticata come una nuova forma di homuncularismo; si veda su questo punto Johnston (1988).

²² Sempre Johnston (1988), su questo punto critica Davidson per il suo iper-razionalismo. A ben vedere l'analisi del rapporto tra pensiero e azione, tra stati mentali e azioni si colloca per Davidson all'interno di un *continuum* che ha alle sue estremità il *Principio di Platone*, cioè l'agire mossi dal proprio miglior giudizio (un principio senza dubbio razionalista) e il *Principio di Medea*, cioè l'agire mossi da forze cieche e oscure, fuori dal nostro controllo; cfr. Davidson (1982, p. 25 s.).

(c) pertanto, il soggetto S è messo nella situazione di credere che non- p . Il principio di razionalità che per lo più viene violato in questa situazione è quello *dell'evidenza totale per il ragionamento induttivo*; il che significa che il soggetto S si ritrova: (i) o a dare maggior peso alle evidenze a favore di non- p (arrivando persino a crearle); (ii) oppure a trascurare il più possibile le evidenze contro non- p . Per fare un esempio molto semplice: la moglie che ha un numero sufficiente di evidenze per credere che il marito la tradisca, a causa dell'incompatibilità di tale credenza con altre credenze (le strutture semi-autonome) – ad esempio con quelle che coinvolgono l'immagine di sé –, può indursi a violare il principio dell'evidenza totale, elaborando la credenza opposta, cioè che il marito non la tradisca. Applicando anche all'autoinganno il «Principio di carità interpretativa», che vieta di attribuire ad un soggetto credenze palesemente contraddittorie, Davidson, lo considera come una situazione in cui qualcuno crede che p e crede che non- p , senza che ciò comporti che egli creda la congiunzione di p e non- p ²³. Per questa ragione, infine, la peculiarità dell'autoinganno, nella descrizione di Davidson, consiste nel fatto che chi si autoinganna crede che non- p , proprio perché (o, a causa del fatto che) crede che p (Davidson 1985, p. 145). La moglie crede che il marito non la tradisca proprio perché (o, a causa del fatto che) essa crede, o meglio, ha le giuste evidenze per sapere che il marito la tradisce; uno stato mentale negativo, di timore, di dolore, o di disprezzo associato all'idea del tradimento rappresenta qui una causa, che però non è ancora una ragione, per il suo violare il principio dell'evidenza totale.

Così descritto l'autoinganno condivide sia aspetti del *wishful thinking*, sia aspetti della *debolezza della giustificazione* e

²³ Davidson (1985) è del tutto esplicito al riguardo: «This step [that “D believes that (he is bald and he is not bald)”] must be resisted: nothing a person could say or do would count as good enough grounds for the attribution of a straightforwardly and obviously contradictory belief, just as nothing could sustain an interpretation of a sincerely and literally asserted sentence as a sentence that was true if and only if D was both bald and not bald, though the words uttered may have been “D is and is not bald”. It is possible to believe each of two statements without believing the conjunction of the two» (p. 138 s.). È lecito considerare queste parole come l'applicazione al caso dell'autoinganno del Principio di Carità interpretativa, che vieta di attribuire credenze palesemente contraddittorie ad un soggetto.

l'intenzionalità in esso presente assume l'aspetto di una motivazione che trae origine da una credenza. Questo lo distingue da una presunta forma di insincerità verso se stessi, che Davidson non ritiene possibile (cfr. Davidson 1985, p. 145)²⁴; quest'ultima, infatti, dal suo punto di vista, contiene un'intenzione che si autodistrugge. Per questa medesima ragione, possiamo infine osservare che l'autoinganno non è rappresentato come una situazione pregiudizievolemente negativa sotto il profilo morale. Per rimanere all'esempio del tradimento coniugale, rispetto all'infedeltà del marito, la moglie può autoingannarsi anche a fin di bene, ad esempio per evitare ai propri figli il trauma di una probabile separazione. D'altro canto, essa può invece farlo per scopi meno nobili, come la convenienza economica e sociale, o per mancanza di rispetto verso se stessa.

Nella rappresentazione che ne dà Davidson, l'autoinganno si profila innanzitutto come una situazione lacunosa sotto il profilo epistemologico, la cui irrazionalità è comunque sempre e in linea di principio leggibile, e quindi dissolvibile, sullo sfondo di una più ampia razionalità condivisa. Che l'autoinganno talvolta possa essere ragionevole, come quando ricorriamo ad esso per allentare una tensione non toglie nulla al fatto che esso *sia* di base una condizione irrazionale. Pertanto, solo di volta in volta, solo caso per caso, saremo autorizzati a descriverlo con sguardo indulgente come una forma irrazionale di ragionevolezza.

ABSTRACT. – Self-deception is both a common phenomenon of everyday life and a topic of great interest for philosophers since the time of Socrates. A self-deceiver seems in fact to behave irrationally if compared to someone striving to know oneself and capable of producing the right reasons for his actions. But how can we properly speak of the "rightness" of reasons and how can we characterize irrationality? I will here present Davidson's reflections on this issue.

²⁴ La posizione di Davidson si distingue dunque da quella di Demos (1960) riportata alla n. 6.

Testi citati

- DAVIDSON D. (1980), *Actions, Reasons, [spazio] and Causes*, in Id., *Essays on Actions and Events*, Oxford, O.U.P.; trad. it. 1992, *Azioni, ragioni, cause*, in *Azioni ed eventi*, Bologna, Il Mulino, pp. 41-61.
- DAVIDSON D. (1982), *Paradoxes of Irrationality*, in Id., *Problems of Rationality*, Oxford, O.U.P., 2004, pp. 169-187; trad. it.: *Paradossi dell'irrazionalità*, in *Studi Freudiani*, a c. di D. Meghnagi, Milano, Guerini e Associati, 1989, pp. 17-41. [cui si riferisce il numero di pagina riportato nel testo].
- DAVIDSON D. (1985), *Deception and Division*, in *Actions and Events-Perspectives on the Philosophy of Donald Davidson*, ed. by E. Lepore, B.P. McLaughlin, Oxford, Blackwell, pp. 139-148.
- DEMOS R. (1960), *Lying to oneself*, «The Journal of Philosophy», 57, 18, 1960, pp. 588-595.
- EVNINE S. (1991), *Donald Davidson*, Cambridge, Blackwell, 1991.
- HÄLLÉN E. (2011), *A Different Kind of Ignorance - Self-Deception as Flight from Self-Knowledge*, Dissertation at Upsala Universitet; <http://uu.diva-portal.org/smash/get/diva2:408360/FULLTEXT01>
- JOHNSTON M. (1988), *Self-Deception and the Nature of Mind*, in B.P. MCLAUGHLIN - A.O. RORTY (eds.) (1988), *Perspectives on Self-Deception*, Berkeley, California University Press, 1988 pp. 63-91.
- MARRAFFA M. (1999), *Il problema dell'autoinganno: una guida per il lettore*, «Sistemi intelligenti», *Il problema dell'autoinganno* (numero monografico), XI, 1999, pp. 373-403.
- B.P. MCLAUGHLIN - A.O. RORTY (1988), (eds.), *Perspectives on Self-Deception*, Berkeley, California University Press, 1988.
- MELE A. (2001), *Self-Deception Unmasked*, Princeton, P.U.P., 2001.
- PEDRINI P. (2006), *Self-Deception: What is it to Blame After All?*, «Annali del Dipartimento di Filosofia 2005», Firenze, University Press, pp. 147-179.
- PEDRINI P. (2012), (ed.), *Philosophy of Self-deception*, in: http://www.humanamente.eu/PDF/complete_issue_20.pdf
- PLATONE, *Cratilo*, in Id., *Opere complete*, Bari, Laterza Ed., 1998.
- RORTY A.O. (1994), *User-Friendly Self-Deception*, «Philosophy», 69, 268, pp. 211-228.