

Roberta De Monticelli

Coscienza e conoscenza tacita

Abstract - This paper is a defence of the so-called argument of consciousness against the significance of the Turing test as a criterion for ascribing thought or mental activity to an artifact (and hence against such parts of a functionalist philosophy of mind as might depend on the significance of such a criterion). This argument, attributed to a Professor Jefferson, is rejected by Turing (1950). It is rejected on the basis of the following remark: according to this argument, the only way to be sure that a machine thinks is *to be* that machine, and to feel oneself thinking. But this assumption obviously leads to solipsism.

My defense of Jefferson's argument, which is indeed slightly reformulated and presented as Socrates' argument, is based on the denial that the above mentioned assumption leads to solipsism. The reason why this consequence seems to (obviously) follow from such an assumption is called *the Cartesian Mistake*.

The reformulated version of Jefferson's thesis reads: for some predicates P (including "mental" predicates, but probably not only them), knowing whether x is P requires *being a P* (and not being x itself, i.e. *that* particular thing which is P). This at least if knowing whether x is P *presupposes* knowing the meaning of "P". I call this knowledge *essential knowlege*. Essential knowledge is knowledge "from inside", or reflexive knowledge. I argue that, in spite of its being *essentially tacit knowledge*, it is a constitutive part of our *knowledge of the meaning of "P"*; and as such it is by no means solipsistic, being on the contrary exactly that tacit component of propositional knowledge which is responsible for the non-purely behavioristic *meaning of "mental" predicates* ("feeling pain" *does not* mean "behaving in a pain-like manner", whatever its criterion of application might be).

The Cartesian Mistake would consist in an arbitrary characterisation of *consciousness* (i.e., in our terms, tacit essential knowledge) as the way in which one's own existence and one's own present acts are given to one. What is arbitrary is the implied particularisation of the concerned knowledge. On the contrary, it is argued, any act of reflection on the exercise of some function gives the subject universal information on the essential nature of that function. This is why by such reflection, "eidetic" and not "psychological", philosophers were able to learn much on the nature of, say, logical thought ("what do we *mean* by 'valid inference'?" Exemples can be multiplied according to the meaning of a given "P").

Why was the Cartesian Mistake so influential even on philosophers of quite different, and in particular anti-mentalistic, traditions? Here is a tentative hypothesis: in order to *see* that *it is* a mistake, it is necessary to give up a dogma of linguistic philosophy: namely, that *all (not just practical) knowledge is propositional knowledge* (in a wide sense of "propositional knowledge", as *not* opposed to "object-knowledge"). Propositional knowledge is defined as (justified) recognition of objects as truth-makers (as making such or such proposition true). Now reflexive (essential) knowledge does *not* consist in recognitions of *objects*. The being (and the acts) I am aware of are not given to me as the being and the acts of an object. (Medieval philosophy worked out this essential distinction in developing the Aristotelian concept of *esse ut actus*; Husserlian phenomenology renewed it independently). This distinction is what makes "knowledge from inside" different from "objective knowledge". This difference was enough for Frege to affirm that the *original* way in which one is given to oneself is strictly private, i.e. ineffable. Turing follows him on this point. It is argued that original self-consciousness is not propositional knowledge (thus far Frege was right), but not because it is ineffable (thus far he was wrong), but because *it is a constitutive part of propositional knowledge* (concerning "mental" life of other people as well as one's own). More precisely, it is the tacit part of propositional knowledge: where "tacit" (not yet made explicit) *does not* mean "incapable of being made explicit". Making it (more) explicit is one of the main tasks of philosophers (this follows from the "Socratic" version of Jefferson's argument). But even recent defenders of the role and significance of "knowledge from inside" (such as Thomas Nagel, for exemple) seem to share the Cartesian Mistake, which prevent "subjective" knowledge from being recognised as "essential" knowlege, somehow *presupposed* by "objective" knowledge.

Partiamo dalla preistoria dei dibattiti attuali sulla coscienza, e precisamente dal famoso articolo di Turing sul test di Turing, *Computing machinery and intelligence* (1950). La seconda delle obiezioni che Turing considera è l'argomento detto "*from consciousness*", attribuito a un tal Professor Jefferson ma condiviso da molte signore, e che suonerebbe così:

"Fino a quando una macchina non saprà scrivere un sonetto o comporre un concerto sulla base di pensieri pensati e di emozioni sentite, e non per giustapposizione casuale di simboli, fino a quel momento non potremo concedere che la macchina è come il cervello - cioè che non solo sa scrivere un sonetto o un concerto ma *sa di* scriverlo. Nessun meccanismo sarebbe in grado di *sentire* (e non soltanto segnalare, che è un facile espediente) piacere per propri successi, dolore quando gli fondono le valvole, di farsi rianimare dalle lusinghe e buttar giù dai propri sbagli, di farsi sedurre da eros, di arrabbiarsi o deprimersi quando non riesce ad avere quello che vuole".

Lasciamo da parte il cervello, che non si vede cosa c'entri (non s'è mai sentito di un cervello che scrivesse sonetti e fosse soggetto a crisi di rabbia, anche perché l'espressione rabbiosa di un cervello è difficile da immaginare). Per il resto l'obiezione ha il solo difetto che non lascia ben capire *cosa* obietti a Turing, e perciò sembra intollerabilmente dogmatica e circolare. Turing propone un criterio per decidere quando una cosa può essere detta intelligente o pensante, ma il professore sembra sapere già quando una cosa può essere detta tale, e ciò che sa il professore a) non è "catturato" dal criterio di Turing (e questa sarebbe un'obiezione: ma) b) si sottrae apparentemente a qualsiasi criterio oggettivo e pubblico. Messa così, l'obiezione è priva di ogni forza. Peccato, perché al fondo il professore ha ragione, ma non c'è di peggio che aver ragione senza saper spiegare perché un altro ha torto. A unica scusante del professore può dirsi che "l'argomento della coscienza" lui l'aveva sollevato un anno prima che Turing proponesse il suo test.

A dire la verità, questo cosiddetto argomento è stato sollevato non uno ma più di duemila anni prima di Turing, e una volta esposto come si deve conserva - a mio parere - intatta la sua forza, ma questo è un dettaglio di storia della filosofia poco importante nel nostro contesto. Invece, quanto alla questione del perché Turing ha torto, è molto più illuminante la risposta di Turing stesso. Sentiamola:

"Secondo la forma più estrema di questa concezione, l'unico modo in cui uno potrebbe essere sicuro che una macchina pensa è di *essere* quella macchina, e di vivere l'esperienza di pensare (*to feel oneself thinking*). Allora uno potrebbe descrivere al mondo quello che prova, ma ovviamente nessuno potrebbe ragionevolmente tenerne conto. Così in base a questa concezione il solo modo di sapere che *una persona* pensa è di essere quella particolare persona. E, in effetti, il punto di vista solipsistico. Può anche darsi che sia il modo di pensare più logico, ma rende difficile la comunicazione delle idee. A ha il diritto di credere "A pensa ma B non pensa", mentre B crede "B pensa, ma A no". Invece di litigare continuamente su questo punto, è invalso l'uso di adottare la convenzione di cortesia che ognuno pensa".

C'è uno strano e grave errore in questa tesi, tanto più strano e grave in quanto sfiora una verità soltanto per deformarla orribilmente. La verità sfiorata è che, in un senso di "sapere" che chiamerò "*conoscenza essenziale*", per sapere che *un P è Q* bisogna in effetti *essere un P - un P*, non *quel P*. Per sapere che un'altra persona pensa, bisogna essere una persona ed essere pensante; per sapere che una certa macchina pensa, bisogna *essere* una macchina di quel genere, ed essere pensante. Questo, per lo meno, se sapere se *x* pensa *presuppone* sapere *che cosa intendiamo per pensare*. "Sapere che cosa intendiamo per" è il senso di "sapere" che ho chiamato "*conoscenza essenziale*". Questa conoscenza essenziale

- ecco la mia prima tesi - è conoscenza di una componente essenziale del significato di molti predicati. Inoltre - seconda tesi - essa è essenzialmente *conoscenza tacita*. Questa conoscenza tacita - terza tesi - coincide con il tipo di sapere che si chiama *coscienza*. Quarta tesi: la conoscenza tacita o coscienza è tipicamente *conoscenza dall'interno*, e questo vuol dire precisamente che per conoscere essenzialmente una cosa di un certo genere bisogna essere una cosa di quel genere. Queste sono le quattro tesi di cui ho bisogno per mostrare perché Turing ha torto. La verità sfiorata da Turing è la quarta tesi: per conoscere essenzialmente una cosa di un certo genere bisogna essere una cosa di quel genere. La sua orribile deformazione consiste nella particolarizzazione di questo enunciato: per conoscere essenzialmente una cosa bisogna essere *quella* cosa. Ma questa deformazione non risale certo a Turing, è anch'essa un bel po' più vecchia. La chiamerò *l'errore cartesiano*.

Ora, per argomentare le mie quattro tesi ricorrerò dapprima all'aiuto di una signora, che al fondo è d'accordo con il professor Jefferson, ma ha il vantaggio di esprimere l'intuizione vera di fondo con molta maggior chiarezza e concisione. Questa signora è Jeanne Hersch, limpida pensatrice svizzera. In occasione di un dibattito sul problema se un computer (supponiamo, uno che passi il test di Turing per qualsiasi suo raffinamento richiesto, ipotesi che è già lievemente imbarazzante nel nostro ambiente finito, dove ogni prova n lascerà il dubbio sulla prova $n+1$) - se un computer, dunque, "pensi", Jeanne Hersch sostenne recisamente che no. A chi le chiedeva, un po' irritato, "ma *come* lo sa?" ella rispose: "perché, signore, mi è accaduto qualche volta di pensare".

A costo di rovinare questa bella battuta, cercherò di mostrarne la pertinenza e la profondità. Anzitutto, dunque, essa ci dice che l'accesso alla conoscenza che un computer non pensa è una *riflessione sull'esercizio* del pensiero. Perché? Evidentemente, dobbiamo supporre che questa riflessione approdi a qualcosa di essenziale a ciò che chiamiamo "pensare", e che questo qualcosa *non è catturato da alcun criterio empirico di applicazione del predicato "essere pensante"*, vale a dire da alcun criterio in base al quale qualcosa è pensante se e solo se si manifesta capace di tali e tali prestazioni osservabili (ad esempio, risponde in modo così pertinente alle domande di un interrogante, che costui, non vedendolo, non è in grado di decidere se sia un uomo o un computer: in ogni caso risponde come un uomo, ha le sue stesse debolezze di memoria, ecc.). Evidentemente, l'idea è che è il "se e solo se" che è sbagliato: soddisfare un criterio empirico di applicazione del predicato sarebbe al più una condizione necessaria, non anche una condizione sufficiente per sostenere che la cosa in questione è pensante.

Bene: cosa sarebbe allora questo "qualcosa di essenziale"? Essenziale, ho detto, a ciò che chiamiamo "essere pensante", o più in generale "essere intelligente" - non a ciò che la signora Hersch soltanto *intende* con queste parole. A ciò che "si" intende comunemente, a ciò che *si vuol dire* con queste espressioni. Si tratterebbe dunque di una componente essenziale del *significato* di queste espressioni, una componente per l'appunto che sfuggirebbe a ogni criterio empirico di applicazione dei relativi predicati. Come mai?

Cerchiamo anzitutto di capire bene. Il punto è dunque che la riflessione sull'esercizio del pensare - e intendo riflessione nel senso stretto in cui l'esercizio in questione è attuato dall'individuo stesso che riflette - rivelerebbe qualcosa di universale, e per nulla affatto di privato o solipsistico, una componente del senso di espressioni come "essere pensante" o "essere intelligente", o una componente essenziale della nozione di intelligenza.

Se questo è vero, allora è falso che la signora Hersch, poniamo, potrebbe sì descrivere al mondo ciò che apprende dalla considerazione riflessiva dell'esercizio di pensare, ma che nessuno potrebbe ragionevolmente tenerne conto, che nessun altro cioè disporrebbe della base di evidenza pertinente per approvare ciò che essa dice. Ovvero:

a) SE (Riflessione individuale = Conoscenza del significato), ALLORA NON (Esito solipsistico)

Reciprocamente, se questa riflessione individuale non può che approdare a un esito solipsistico, (cosa "ovvia" per Turing), allora è vero che nessuna riflessione individuale può approdare a una conoscenza essenziale relativa al pensare, ovvero a una miglior comprensione di ciò che *intendiamo* per "essere pensante", o intelligente. Cioè:

b) SE (Riflessione individuale = esito solipsistico), ALLORA NON (Riflessione individuale = Conoscenza del significato).

Ora si può procedere in due modi. Rapidamente, a partire da b), per Modus tollens. È palesemente falso che nessuna riflessione individuale può approdare a una miglior comprensione di ciò che *intendiamo* per "essere pensante", o intelligente. La riflessione - in senso stretto - sull'esercizio del pensare è una via maestra di esplicitazione di questa nozione, di tutto ciò che essa - ma solo confusamente e tacitamente - implica per la coscienza ordinaria. I filosofi, da Platone e Aristotele in poi, non hanno fatto praticamente altro. Ma limitiamo pure il pensare a uno dei suoi aspetti - la capacità di fare inferenze logicamente corrette. Provate a insegnare a qualcuno la logica senza fargli *fare esercizi*. Dunque, è falso che una riflessione individuale debba avere necessariamente un esito solipsistico.

Ma a me pare più interessante partire da a). Occorre dunque provare che la riflessione individuale è una fonte di conoscenza essenziale, e più precisamente di comprensione del significato. Ora, questo è particolarmente semplice nel caso di un predicato come "soffrire". In questo caso, è evidente che ciò che intendiamo per "soffrire", e ancora più semplicemente per "soffrire fisicamente", non è affatto interamente "catturato" dai criteri empirici di applicazione del predicato, per cui si dice che qualcosa soffre se manifesta certi comportamenti, detti appunto di dolore. Non solo perché uno potrebbe fingere (gli animali non fingono, eppure li vediamo soffrire, senza ombra di dubbio), ma perché una persona impara a usare il verbo come si deve quando impara ad applicarlo anche a se stessa, più o meno urbanamente, invece di strillare e piangere. Ora, è scarsamente controvertibile che nel mio proprio caso - supponendo che abbia già imparato l'italiano - non procedo per osservazione dall'esterno (ad esempio guardandomi allo specchio: toh, che espressione sofferente. Soffro dunque?), così come non ho in generale bisogno di osservarmi per sapere se sto giocando a tennis, o suonando il piano. Appunto, dicono Turing e Cartesio, nel mio proprio caso ho un accesso alla mia condizione dall'interno, con il vantaggio che non posso sbagliarmi e lo svantaggio che questa conoscenza è privata e solipsistica. Ma questo è sbagliatissimo, perché al contrario: la sofferenza vissuta, lungi dall'essere una condizione incomunicabile, è un paradigma del senso della parola

"soffrire", e una condizione per impararne l'uso. Non è l'unica, perché anche osservare comportamenti di un certo tipo, contorcimenti e urla (oppure stoica fermezza, se è per questo) è una condizione per acquisire un buon dominio di questa nozione. Ma sono certa che se chiedete a qualcuno: "hai ancora mal di stomaco?" e quello risponde: "non so, aspetta un momento che vado a vedere che faccia ho" - dubiterete che non abbia capito la domanda.

Il punto è dunque che l'esperienza vissuta del dolore fisico, la sua conoscenza "dall'interno", è una parte costitutiva di ciò che intendiamo per "soffrire", *di modo che non possiamo affatto attribuire questo predicato senza attribuire questa esperienza vissuta*. Indubbiamente, possiamo sempre sbagliarci nell'attribuirla. Ma non certo perché non sappiamo se nel caso altrui "soffrire" voglia dire altro che manifestare comportamenti di dolore. Sappiamo benissimo che vuol dire anche altro *in ogni caso*, visto che si tratta di una conoscenza concettuale: oppure non sappiamo l'italiano. E questo spiega perché l'esplicazione del predicato in termini esclusivi di criterio di applicazione empirica non è, *in linea di principio e non di fatto*, una esplicazione esaustiva, e tralascia una componente essenziale del suo senso. Supponiamo infatti - a torto, del resto - che io non abbia accesso diretto, empatico e non osservativo, al dolore altrui: anche se questo fosse vero, certamente io rifiuterei qualunque analisi concettuale del senso di "soffrire" che non ne espliciti questa componente costitutiva. E, credo, anche voi, se parlate italiano. Perciò, se anche qualcuno escogitasse un test di Turing della sofferenza, e un computer lo superasse brillantemente, questo non costituirebbe ancora una *ragione sufficiente* per concludere che un computer può soffrire, secondo ciò che intendiamo per "soffrire".

Il caso del pensare è appena più complesso. "Mi è accaduto qualche volta di pensare", e per questo so che un computer, nella misura almeno in cui dipende da un programmatore e non può che obbedire alle istruzioni ricevute, non pensa, o non è intelligente. Lo so sulla base di una riflessione sull'esercizio, attuale o passato, di pensiero: vale a dire che precisamente *ciò che si intende (non: ciò che io intendo)* per "essere intelligente" o pensante non mi è accessibile, almeno in una componente essenziale del senso di questa espressione, che sulla base di questo *paradigma* dell'essere pensante: il mio essere, quale mi è dato in questo esercizio. Sarebbe sbagliato vedere in questa constatazione un peccato di superbia. Per paradigma infatti non si intende, appunto, criterio empirico, ma esemplificazione. Ora ogni conoscenza essenziale, o dall'interno, ha per definizione bisogno di un'esemplificazione individuale, ma non si riduce per questo a una conoscenza dell'essere individuale. Mi concederete per esempio che la condizione che chiamiamo "vita" ha due modi di esserci presente o nota: la vita si manifesta, o si percepisce dall'esterno negli oggetti che ci appaiono viventi, e la vita si vive. Ora per il fatto che una vita - una sola purtroppo - mi è data da vivere, e che perciò la chiamo "mia", io non so soltanto che cos'è avere la mia vita, ma so che cos'è avere una vita *propria*, ed è quella che attribuisco agli oggetti apparentemente animati. So precisamente che una vita è data come propria quando è data da vivere, e non soltanto data da percepire. E so anche che per accedere al senso dell'espressione "una vita umana", si parlasse pure di quella di un aborigeno australiano o di un papa dugentesco, ciascuno non ha che l'esempio della propria, per lavorarci alle più sfrenate variazioni eidetiche.

Ma vediamo dunque *come* il mio essere mi è dato nell'esercizio attuale dell'intelligenza, ad esempio ora che scrivo. Ecco: come un fare, soffrire, volere, progettare, rimpiangere, sperare, ricordare; come un essere presi in una trama di dipendenze, bisogni, desideri, interessi, come un far fronte a situazioni, scelte, decisioni; e come un potere, infine, trascendere tutte queste *motivazioni* nell'assunzione di un atteggiamento puramente cognitivo, in cui il mio solo impegno, liberamente assunto, è quello di accertare la verità di ciò che è in questione. Questa adesione a un valore - il valore della giustificazione razionale - non è d'altra parte, per quanto ne so, che un aspetto dell'esperienza di valore,

cioè della possibilità che ho, in certi momenti, di trascendere le motivazioni radicate nella "cura" di sopravvivere, possibilità senza la quale non si dà pensiero "creativo". Pure, si dà, ed è un aspetto di ciò che intendiamo per "intelligenza".

Il punto è che questo immenso sfondo e fondamento vissuto della nozione di intelligenza, questa *componente interna o paradigmatica del senso del predicato*, non è preservato dall'esplicazione che di questo predicato si propone quando lo si definisce in base a un criterio empirico di applicazione a entità date come oggetti, di forma umana o no, un criterio della forma: "x è intelligente se si manifesta capace di tali prestazioni". Dato il criterio e date le prestazioni non c'è scampo alla conclusione: "x è intelligente". Eppure noi sentiamo che in questa attribuzione ciò che di essenziale intendevamo è andato perduto. In definitiva, con una tale "esternalizzazione" del predicato non arriveremo mai a meglio *comprendere* che cosa *intendevamo*, anche se potremo forse fornire una teoria che *spieghi* i fenomeni (esterni) dell'intelligenza.

Costretta dal poco tempo, ho argomentato tutte insieme le mie quattro tesi. Infatti, data la prima, le altre non sono che ovvie definizioni. Dato che si conceda che noi abbiamo una conoscenza essenziale di che cosa è l'intelligenza, consistente nella conoscenza di una componente essenziale del significato di "essere intelligente" (prima tesi), si concederà facilmente che si tratta al più di conoscenza tacita, o implicita (seconda tesi); ma questa conoscenza implicita, che esplicitiamo sulla base di una riflessione, è (per definizione di riflessione) la coscienza che abbiamo di pensare (come di ogni altra funzione nell'esercizio della quale consiste il nostro esistere. Di questo infatti siamo indubbiamente coscienti, ma esso non ci è dato che nell'esercizio di questa o quella delle funzioni in cui esso consiste: respirare, mangiare, scrivere...) (terza tesi). Infine, "coscienza" è semplicemente un'altra parola per "conoscenza dall'interno" (quarta tesi).

Ma non m'illudo affatto che mi concediate già la prima tesi.

Essa equivale a sostenere che (come dice quel grande filosofo del linguaggio che è Dante), il senso ordinario e comune di numerosi predicati della nostra lingua (e non soltanto di quelli cosiddetti "psicologici") è tale che "intender non lo può chi non lo prova". Perché allora lo si nega molto comunemente, questo principio? In virtù dell'errore cartesiano, naturalmente. E in che cosa consiste questo errore? Nel limitare molto arbitrariamente il ruolo della *coscienza* a quello di modo in cui mi è data o conoscibile la mia esistenza (indubbiamente diverso da quello in cui vengo a sapere che esiste un altro), invece di vedervi correttamente *una componente costitutiva del senso di "essere vivente", "sofferente", "senziente", "amante", "pensante" - o, infine, di "essere un soggetto", o ancora più specificamente, "essere una persona"*. Ora, perché tanta parte della filosofia moderna e contemporanea della mente ha seguito Cartesio in questo errore? Perché, mi pare, per vedere che è un errore bisogna rinunciare a un dogma, magari oggi offuscato, ma mai veramente messo in questione, della filosofia del linguaggio: e cioè, che ogni conoscenza sia conoscenza proposizionale (vale a dire conoscenza per mezzo di proposizioni), o equivalentemente, dato che ogni conoscenza proposizionale porta necessariamente su oggetti (quelli appunto che rendono le proposizioni vere o false), che ogni conoscenza sia conoscenza di oggetti. Ora, la conoscenza dall'interno o coscienza non è conoscenza di un oggetto: in essa, l'essere di cui sono cosciente non mi è affatto dato come l'essere di un oggetto. Certo, io posso sempre assumere su me stesso un punto di vista esterno o

oggettivo, ad esempio studiare il movimento da me vissuto come studio la dislocazione spaziale di qualunque altro corpo. Ma non è così che esso mi è dato originariamente. Ora - ultimo passaggio un po' delicato - quando rifletto, io non faccio che rendere tematico questo modo originario e non oggettuale in cui sono dato a me stesso, di solito non tematicamente. Nel farlo, cesso appunto di considerare la vita, il pensiero etc. come proprietà di oggetti, per volgermi riflessivamente al loro attuale esercizio. Questo è bastato ai filosofi del linguaggio, a partire da Frege, per affermare che la presenza originaria a se stessi è strettamente ineffabile, è un'esperienza mistica o comunque solipsistica. Turing lo segue su questo punto. Il fatto è però che il nostro essere, quale ci è dato dall'interno, non è un oggetto di conoscenza proposizionale semplicemente perché ne è un presupposto o una *fonte*. La presenza originaria non è una conoscenza proposizionale (in questo Frege aveva ragione) - ma non perché sia ineffabile (in questo aveva torto) o solipsistica, dunque non una conoscenza, ma perché è *parte costitutiva di ogni sapere proposizionale* (relativo a persone, compresa la mia in quanto assuma su di essa un punto di vista oggettivo). Precisamente, è quella parte costitutiva del senso dei predicati in questione che è data "dall'interno", o la loro componente *immediata e tacita, paradigmatica*. La nostra comprensione del senso di questi predicati è precisamente, per questa loro componente *essenziale*, conoscenza tacita. "Abbandonare il piano di un discorso su oggetti" non significa affatto darsi alla mistica, ma *risalire* a questa componente tacita del senso di ciò che diciamo, e provare a esplicitarla, fin dove si può. Questo è propriamente quello che si chiama filosofia, o una gran parte di ciò che così si chiama: e questo spiega come mai, circa duemilaquattrocento anni fa, Socrate propose un "argomento della coscienza" che suona così: conosci te stesso - e conoscerai l'essenza delle cose, e avrai conoscenza dell'universale.