 Etica & Politica / Ethics & Politics, XXVI, 2024, 1, pp. 33-45
ISSN: 1825-5167

RICERCARE L'INTESA CON UNA MACCHINA
COMPATIBILITA FRA UMANI E SOGGETTI ARTIFICIALI NELLA TEORIA DEL DISCORSO DI HABERMAS

PAOLO CAPRIATI
Università degli Studi di Palermo
Dipartimento di Giurisprudenza
paolo.capriati@unipa.it

ABSTRACT
Is the emergence of artificial subjectivities endowed with a certain autonomy compatible with Habermas' theory of discourse? To answer this question, it is necessary, preliminarily, to address the question of the Habermasian subject. In doing so, one observes how machines stand as potential interlocutors. As interlocutors, like the other protagonists in Habermasian discourse, they must aim at understanding and must have a legitimate interest. This opens the field to two orders of problems: (a) are machines capable of understanding? (b) can machines have an interest? Arguments borrowed from the Chinese Room debate show how, in principle, there are no obstacles to the entry of machines into discourse. If machines can participate in discourse, then in what language do they talk to humans? To communicate, humans and machines must resort to translation techniques, known as natural language processing. Translation already places itself outside language, deferring to a shared idea of reality in which to root universal meanings. By placing communication outside language, the entry of machines into discourse seems to checkmate Habermasian theory. However, two further arguments – "the Others mind reply" and the private language argument – make the case that communication between humans and machines poses no different problems than communication between humans alone.

KEYWORDS
Habermas, discourse theory, artificial entities, Chinese room, natural language processing.

1. INTRODUZIONE

Questo scritto intende analizzare i profili di compatibilità fra l'emergere di soggettività artificiali e la teoria del discorso di Habermas.

Tale questione verrà analizzata da due prospettive diverse.

La prima prospettiva sarà volta a valutare la tenuta *esterna* del pensiero habermasiano: la teoria del discorso può essere adottata per descrivere *anche* la comunicazione fra umani e macchine?

Prima di rispondere a questa domanda, occorre chiedersi se la comunicazione intersoggettiva (o senza soggetto) di Habermas – desoggettivizzando il processo discorsivo dell'intesa – ci risparmi dall'onere di indagare la compatibilità fra teoria del discorso e lo scambio comunicativo umano-macchina. Come si vedrà, Habermas non può fare a meno di interlocutori per dare sostanza al suo discorso. La domanda di partenza, quindi, si rivela rilevante.

Verrà, in secondo luogo, delineato il profilo delle macchine cui ci stiamo riferendo. Non si tratta di qualsiasi tipo di macchina, ma di quelle dotate di una certa autonomia. Qui riteniamo "autonoma" una macchina il cui comportamento non è perfettamente prevedibile o spiegabile a posteriori da un osservatore umano.

Una volta superate le questioni preliminari, approfondiremo alcuni profili problematici. Se per Habermas partecipano al discorso «i possibili interessati» e se il discorso è volto al raggiungimento dell'intesa reciproca, allora possiamo sollevare due tipi di obiezione.

La prima obiezione è volta a fare chiarezza sul concetto di "intesa". Una macchina è in grado di intendere? La seconda obiezione, invece, riguarda la questione dell'interesse. Si può dire che una macchina possieda degli interessi? Entrambe queste criticità verranno affrontate mutuando argomenti dal dibattito sulla stanza cinese.

Affrontate queste due obiezioni, entreremo nel merito della questione, chiedendoci in quale lingua comunichino umani e macchine. Escludendo che condividono un comune linguaggio, l'impianto habermasiano – che fonda la legittimità del diritto su processi linguistici – rischia di entrare in crisi qualora ammettessimo l'ingresso nel discorso di soggetti artificiali.

La seconda prospettiva di osservazione è quella interna. Attraverso tale prospettiva si intendono considerare eventuali incompatibilità comunicative fra umani e macchine. Ricorrendo ad argomenti come quello della "Other Minds" e del linguaggio privato, vedremo come non ci siano in linea di principio ragioni sufficientemente forti per sostenere che umani e macchine comunichino in un linguaggio così diverso da quello in cui gli umani comunicano gli uni con gli altri.

2. UNA QUESTIONE PRELIMINARE. IL PROBLEMA DEL SOGGETTO

La teoria del discorso è compatibile con le novità apportate dall'avvento delle tecniche di *machine learning*? Occorre definire i due termini della questione e poi riscrivere la domanda affinché sia meno vaga. Che cos'è la teoria del discorso? E quali sono queste novità che potrebbero mettere in discussione la teoria del discorso?

Per teoria del discorso, in riferimento a decisioni giuridicamente vincolanti, intendiamo il principio che «riconduce la legittimità del diritto a quei processi e

presupposti comunicativi capaci di fondare – dopo la loro istituzionalizzazione giuridica – la supposizione che i processi della produzione e applicazione giuridica conducano a risultati razionali» (Habermas 2013: 493).

Le nuove tecnologie che qui interessano riguardano forme definite, in maniera generica, di intelligenza artificiale. Le forme che rilevano sono quelle dotate di un'autonomia tale per cui si può porre una questione di soggettività.

La domanda si rivela così più chiara: la teoria del discorso è compatibile con l'emergere di soggettività artificiali dotate di un certo grado di autonomia? Per quale ragione non dovrebbe esserlo?

La prima questione da affrontare è quella relativa alla soggettività, in generale.

Si potrebbe in prima battuta sostenere che la teoria del discorso, affrontando la questione del soggetto in modo originale, riesce a superare eventuali problemi di compatibilità e competenza con entità artificiali. A fondamento della teoria del discorso, per Habermas, ci sarebbe la ragione comunicativa.

Essa «si distingue dalla ragion pratica anzitutto perché non è più riferita a singoli attori, o a macro-soggetti di natura statale e sociale. Ciò che rende possibile la ragione comunicativa è il *medium* linguistico, attraverso cui s'intrecciano interazioni e si strutturano forme di vita» (Habermas 2013: 33).

Prendendo le distanze tanto da Rousseau – con il riferimento al macro-soggetto dotato di potere legislativo – quanto da visioni che intendono la formazione legislativa come aggregazione di volontà individuali, Habermas sembrerebbe proporre non la via del soggetto, ma una via che passa fra i soggetti, intersoggettiva. Come è stato suggerito dallo stesso Habermas, potrebbe parlarsi di una «comunicazione senza soggetto»[1].

> La teoria del discorso prende finalmente congedo da quelle figure di pensiero – ancora legate alla filosofia della coscienza – per cui bisognava o ascrivere la prassi d'autodeterminazione dei cittadini a un soggetto sociale collettivo oppure ricondurre l'anonimo "dominio delle leggi" alla concorrenza dei soggetti individuali. Nel primo caso, si vedeva la cittadinanza come un attore collettivo che rispecchiava e agiva in nome della totalità. Nel secondo caso, i singoli attori funzionavano da rotelline, o variabili dipendenti, dei processi del potere – processi funzionanti alla cieca, in quanto al di là degli individuali atti di scelta potevano esserci solo decisioni aggregate, non liberamente consapevoli (Habermas 2013: 362).

Pur superando la dicotomia macro-soggetto/pluralità di individui, Habermas riesce davvero a fare a meno di un soggetto politico di riferimento?

[1] Il concetto di "comunicazione senza soggetto" (Habermas 2013; Habermas 2017) rivela affinità con quello di "a*nonymous public conversation*", che troviamo in Benhabib (1996). Altrove (Dryzek 2006), tale concetto è stato valutato insufficiente per analizzare le "*divided societes*". Sarebbe, infatti, troppo amorfo considerando che l'identità dei soggetti che prendono parte al discorso è una questione cruciale.

> La teoria del discorso punta sull'intersoggettività di grado superiore caratterizzante i processi d'intesa che si compiono nelle procedure democratiche oppure nella rete comunicativa delle sfere pubbliche. All'interno e all'esterno del complesso parlamentare, queste comunicazioni senza soggetto formano arene in cui può prender piede una formazione più o meno razionale dell'opinione e della volontà circa questioni politiche, vale a dire, circa materie socialmente rilevanti e bisognose di disciplina. Il flusso di comunicazione che si instaura tra a) pubblico formarsi dell'opinione, b) risultati elettorali istituzionalizzati, e c) decisioni legislative, serve a garantire che l'influenza dei mass media e il potere comunicativo si trasformino – attraverso la funzione legislativa – in un potere amministrativamente esercitabile (Habermas 2013: 362).

In altri punti, il riferimento alla comunicazione senza soggetto si fa ancora più sfuggente.

> Subjectless and anonymous, an intersubjectively dissolved popular sovereignty withdraws into democratic procedures and the demanding communication presuppositions of their implementation. It is sublimated into the elusive interactions between culturally mobilized public spheres and a will-formation institutionalized according to the rule of law. Set communicatively aflow, sovereignty makes itself felt in the power of public discourses (Habermas 1997: 58-59).

In proposito, si è osservato come tali riferimenti finiscano per essere connotati da un certo lirismo e siano supportati da argomenti poco fondati (Goodin 2008: 261).

Da quanto visto finora, espressioni come "senza soggetto" e "anonima" hanno più che altro una carica evocativa. Non è l'assenza di soggetto ciò che Habermas inaugura, ma, al massimo, la difficoltà di catturarlo.

Infatti, seppur rarefatto, un soggetto politico di riferimento nella ricostruzione di Habermas è comunque presente. In questo senso, nonostante il superamento della soggettività dicotomica appena descritta, ad Habermas sono necessari degli interlocutori che portino avanti il discorso. Tali interlocutori si caratterizzano per la loro eterogeneità. Non si tratta, infatti, né genericamente di cittadini, né di soggetti collettivi dotati di legittimità a decidere. Il soggetto che si intuisce tra le righe della teoria del discorso resta indefinito. Sviluppandosi nel medesimo medium linguistico, a questo soggetto è richiesto di comunicare nella stessa lingua degli altri partecipanti al discorso. Comunicare nella stessa lingua è necessario per raggiungere l'intesa.

Sull'intesa reciproca, infatti, si fonda la legittimità di una teoria che rompe il legame volontà/potere. Il potere, in una democrazia, secondo Habermas è legittimato dall'intesa che i partecipanti al discorso possono raggiungere e non dall'aggregazione delle loro volontà.

3. A QUALI MACCHINE CI RIFERIAMO?

Per portare avanti il suo discorso, Habermas non può fare a meno di interlocutori. La loro interlocuzione è volta all'intesa reciproca, pertanto ciò che è loro richiesto è l'utilizzo della stessa lingua. Sembra chiaro che Habermas non voglia riferirsi solo agli individui né solo a macro-soggetti o a centri di potere. Per questa ragione, occorre domandarsi se tra gli interlocutori al discorso di Habermas si possano far rientrare soggetti artificiali.

Bisogna anzitutto domandarsi se una macchina può comunicare. Questa domanda non appare, però, completa. Si tratta di una semplice comunicazione? Ciò che Habermas richiede ai suoi interlocutori è una comunicazione volta all'intesa reciproca. Al di là del comunicare, le macchine sono capaci di intesa e di farsi intendere?

La domanda si rivela tutt'altro che semplice. Si tratta, in altre parole, dell'annosa questione sulla coscienza delle macchine. Dal momento che chiedersi se una macchina sia in grado di intendere o di essere intesa non è una domanda che può essere in questa sede approcciata, verranno prese vie traverse.

Proviamo a immaginare che una macchina partecipi al discorso. Che tipo di partecipazione dovrebbe avere per essere considerata un soggetto e non solo uno strumento? Nell'impossibilità di definire quando un soggetto è in grado di intendere, dobbiamo sostituire i termini della questione.

Per poter essere considerata un soggetto, la macchina in questione deve essere dotata di una certa autonomia. "Autonomia" in questo contesto vuol dire che essa agisce senza essere subordinata alla volontà di altri. Non essere soggetti alla volontà altrui significa rispondere alla propria volontà. Siamo nuovamente caduti in un'annosa questione, che è quella relativa alle capacità di volere di una macchina. Anche a questa domanda non possiamo dare una risposta: occorre percorrere un'altra strada.

Rimanendo sulla via dell'autonomia, un ente può dirsi autonomo quando agisce in base a regole proprie. In ipotesi, dunque, una macchina che agisce in base a regole proprie può considerarsi in una certa misura autonoma e tale autonomia la eleva al rango di soggetto e non di semplice strumento. Ma quando di una macchina si può dire agisca in base a regole proprie?

Il fenomeno del *machine learning* sembra essere particolarmente rilevante a questo proposito. Si può definire il *machine learning* come l'insieme di tecniche e di metodi che utilizzano dati per trovare nuovi *pattern* e per generare nuova conoscenza e modelli utili per l'effettiva produzione sui dati (Van Otterlo 2013). In particolare, ci si riferisce a quegli algoritmi le cui azioni sono difficili da predire per gli esseri umani o la cui logica decisoria è difficile da spiegare a posteriori (Mittelstast et al. 2016).

Seguendo quest'ultima osservazione, le "regole proprie", in base alle quali una macchina agisce, sono tali perché producono comportamenti che sono

incomprensibili e imprevedibili per gli umani. L'autonomia di una macchina, in altre parole, è legata alla sua capacità di essere una "scatola nera": oggetto insondabile e inconoscibile per un osservatore umano. Al contrario, se consideriamo una macchina il cui comportamento è perfettamente prevedibile o spiegabile a posteriori, ci stiamo riferendo a uno strumento. In relazione a una macchina del genere, non ha senso chiedersi se essa sia o meno capace di intendere e di essere intesa: se essa è a priori perfettamente conoscibile non vi sarà alcuna tensione verso l'intesa, giacché questa si dà per presupposta. Pertanto, non possiamo sostenere che una macchina siffatta partecipi al discorso.

Paradossalmente, dunque, le macchine candidate ad interloquire sono proprio quelle il cui funzionamento resta per noi sconosciuto. A differenza delle macchine-strumento, per queste macchine "*black-box*" si pone un problema di ricerca dell'intesa.

Se non è possibile rispondere direttamente alla domanda "le macchine sono capaci di intendere o di essere intese?", si può almeno definire per quale classe di macchine ha senso porre questa domanda. In ragione di ciò, abbiamo escluso tutte quelle macchine la cui autonomia ridotta non permette di classificarle come soggetti.

Da ciò discende che il criterio per definire la soggettività è quello dell'autonomia. In altre parole, un soggetto è tale se è autonomo. Il criterio dell'autonomia è quindi requisito necessario per permettere l'accesso al discorso.

Da quanto visto finora, evitando di definire il concetto di intesa, la teoria del discorso di Habermas non sembra precludere a una macchina l'accesso alla partecipazione.

4. OBIEZIONI E CONTRO-OBIEZIONI PER L'ACCESSO DELLE MACCHINE AL DISCORSO

Si è osservato come Habermas non possa fare a meno di soggetti: egli ha bisogno di interlocutori che prendano parte al discorso. Tali interlocutori sono diversi da quelli che vengono solitamente identificati come i protagonisti del discorso politico. Non si tratta né di individui né di macro-soggetti costituitisi come comunità o come istituzione, ma di soggetti la cui azione è votata all'intesa reciproca. L'azione politica, ridotta a comunicazione, si sostanzia in uno scambio tra attori eterogenei che condividono la stessa lingua. Considerate queste premesse, non sembra esserci alcuna barriera per l'ingresso delle macchine al discorso.

A questa conclusione si possono avanzare due ordini di obiezioni.

La prima è già emersa e riguarda la capacità di intendere di una macchina. Alla tesi che non limita il processo discorsivo dell'intesa ad attori esclusivamente umani, si può obiettare che una macchina non è capace di intendere. Un'obiezione del genere ci rimanda al dibattito sulla stanza cinese di Searle (1999).

Searle immagina di essere in una stanza assieme ad un computer per rispondere a delle domande in cinese che gli vengono passate attraverso una fessura sotto la porta. Searle non parla il cinese, ma consultando le indicazioni del computer, invia risposte corrette in cinese fuori dalla porta, e questo induce chi è fuori a supporre erroneamente che ci sia qualcuno che parli cinese nella stanza.

L'esperimento della stanza cinese intende dimostrare che una macchina è in grado di dare una parvenza di comprensione della lingua, ma questo non significa che possa realmente comprendere. In altre parole, Searle sostiene che i computer semplicemente usano regole sintattiche per manipolare stringhe di simboli, ma non hanno comprensione del significato o della semantica.

L'esperimento della stanza cinese non è stato risparmiato da critiche. Quella che ai nostri fini risulta più rilevante è nota come "the Brain Simulator reply", sostenuta fra gli altri da Paul e Patricia Churchland (1990). Secondo questa critica, bisogna immaginare un cervello artificiale che simula il funzionamento di quello umano: anziché operazioni su stringhe di simboli, questo cervello artificiale imita le sequenze di attivazione nervosa che si verificano nel cervello di un soggetto che parla cinese quando lo comprende. Dal momento che un computer del genere avrebbe lo stesso funzionamento del cervello di un essere umano che parla cinese, processando le informazioni alla stessa maniera, si può sostenere che questo computer sia in grado di comprendere il cinese[2].

Una tale critica mira a riconsiderare le possibilità di comprensione – e quindi di intesa – di una macchina: mettere in discussione l'assunto che le macchine non sono in grado di pensare mostra come non abbiamo argomenti sufficienti per sostenere che le macchine non sono in grado di intendere.

Si tratta di un capovolgimento dell'onere della prova: ciò che occorre dimostrare non è la capacità delle macchine di pensare, ma la loro incapacità. In altre parole, il dibattito sulla stanza cinese mostra come non ci siano argomenti sufficientemente solidi per escludere a priori la capacità di pensare di una macchina. In assenza di tali argomenti, la prima obiezione sulle possibilità di una macchina di partecipare al discorso non può essere accettata.

La seconda obiezione ha a che fare con il concetto di interesse. Come precisa Habermas «le norme e le regole che possono pretendere legittimità sono soltanto quelle che tutti i possibili interessati potrebbero approvare in quanto partecipanti a discorsi razionali» (Habermas 2013: 546). Occorre interrogarsi, dunque, sulla capacità di una macchina di avere degli interessi. Se si dimostrasse, infatti, che le macchine non possono avere interessi, allora non sarebbero legittimate a partecipare al discorso di

[2] Oltre ai Churchland, si sono occupati di questa tesi Chalmers (1996), Cole & Foelber (1984), Pylyshyn (1980).

Habermas. L'obiezione può essere formulata nei seguenti termini: un'entità incapace di comprendere non può avere coscienza di se stessa, quindi dei suoi interessi.

Questo tipo di obiezione, in realtà, si rivela ancora più debole della precedente. Il concetto di interesse, infatti, può essere sostituito con quello di obiettivo. Se è difficile accettare completamente che una macchina possieda una coscienza, non si può escludere che essa sia in grado di perseguire degli obiettivi, che corrispondono ai suoi interessi. Avendo disgiunto il concetto di interesse da quello di coscienza, le resistenze nell'accettare l'incapacità di una macchina di pensare non si dimostrano argomenti rilevanti in questa seconda obiezione

Riusciamo a escludere completamente che una macchina possa perseguire un determinato interesse? O che sia interessata a che le cose vadano in una certa maniera anziché in un'altra? Non trovando argomenti che persuadano sull'inesistenza di interessi da parte di una macchina, una tale obiezione non può essere tenuta in considerazione.

A questo punto, la strada verso l'accesso al discorso da parte di una macchina sembra più spianata: queste due obiezioni non si sono rivelate un reale ostacolo.

5. IN CHE LINGUA PARLANO UMANI E MACCHINE?

Superate le obiezioni relative all'impossibilità di considerare le macchine come parti attive del discorso, si proverà a indagare quale sia la lingua in comune fra umani e macchine.

Umani e macchine riescono a comunicare fra loro attraverso tecniche di *natural language processing* (NLP). Si tratta, in altre parole, di strumenti per l'elaborazione del linguaggio naturale. Il NLP non crea una nuova lingua, ma rende intelligibili per una macchina proposizioni del linguaggio naturale. Si tratta, in altre parole, di un'operazione di traduzione.

Non esiste, dunque, una lingua comune ad umani e macchine. Essi possono comunicare attraverso processi di traduzione da una lingua all'altra.

L'assenza di una lingua comune, però, non significa incomunicabilità. Svincolare la comunicazione dalla lingua significa trovare un nuovo fondamento per la pretesa di validità delle norme. Se linguaggi diversi non rappresentano più un limite per la comunicazione, allora uno spazio centrale viene acquisito dalle regole di traduzione. Tali regole non sono la somma dei due diversi linguaggi, ma ciò che permette di comunicare a soggetti che parlano lingue diverse. Tale processo di mediazione finisce per delegittimare e devalorizzare il "discorso" per come l'aveva inteso Habermas. Se ciò che permette a due agenti di comunicare non è il linguaggio in sé, ma le regole di traduzione, l'accordo non è più radicato nel discorso, che smette di essere condiviso.

Esso viene anticipato – o posticipato – allo stadio che riguarda la fase di incontro fra due linguaggi diversi.

Se con la teoria del discorso Habermas fissa nella procedura discorsiva la legittimità delle decisioni, svincolandola in questo modo da predeterminati contenuti etici, nel dialogo umano-macchina, la legittimità acquisisce una dimensione extra-linguistica. Il dialogo umano-macchina, definito in questi termini, impone di trovare un punto di incontro diverso da quello meramente linguistico.

A questo punto, l'accordo sulle regole di traduzione può essere inteso come una fase del processo discorsivo di intesa. Ciò tuttavia potrebbe apparire come una forzatura. L'accordo sulle regole di traduzione da una lingua a un'altra è esterno al linguaggio cui si riferisce Habermas e ha bisogno di un parametro terzo di riferimento. Tale parametro si deve radicare su una concezione condivisa di realtà che va oltre il linguaggio stesso. Un radicamento del genere priva la teoria del discorso della sua forza pervasiva.

6. IL PROBLEMA DEL LINGUAGGIO

Se ammettiamo l'accesso di una macchina al discorso, occorre riformare la definizione di linguaggio. Se non riuscissimo in questa operazione, dovremo ammettere che l'emergere di soggettività artificiali non è compatibile con la teoria del discorso.

La definizione di linguaggio di Habermas principia da una rottura con i postulati metafisici di Kant – sulla contrapposizione tra intelligibile e fenomenico – e la dialettica speculativa di Hegel – tra essenza e fenomeno.

Se «"Reale" è ciò che si lascia simbolicamente formulare» (Habermas 2013: 44), il rapporto fra linguaggio e mondo viene completamente riconfigurato. Il linguaggio per Habermas è il mezzo che rende il reale "reale". Agire all'infuori di esso vuol dire perdere contatto con la realtà.

Se il dialogo umano-macchina è fra interlocutori che parlano lingue diverse, non si può non riferirsi ad una universalità dei significati. Tale universalità trascende il linguaggio e rende gli sforzi di ancorare la legittimità politica a un processo linguistico vana.

Ma riformando in questo modo la definizione di linguaggio la teoria del discorso sembra perdere significato. Al contempo, non si può non rilevare come umani e macchine parlino lingue diverse, tanto che ci sia bisogno di particolari tecniche – si veda il NLP – per metterli in comunicazione.

D'un tratto l'accesso di soggetti artificiali al discorso si complica e sembra mettere in crisi la teoria del discorso stessa.

7. UNA CONCEZIONE DIVERSA DI LINGUAGGIO

Abbiamo osservato come non ci siano ostacoli insuperabili per l'accesso di macchine al discorso habermasiano. Le obiezioni sollevate non risultano preclusive. Tuttavia, la questione si è rivelata più complessa da quando ci si è domandati se esiste un linguaggio in comune fra umani e macchine.

La domanda di partenza riguardava la compatibilità fra a teoria del discorso e l'emergere di soggettività artificiali. A questa domanda abbiamo dato una risposta positiva. Ciononostante, dei problemi di compatibilità continuano ad esserci.

Si proverà a questo punto ad adottare una prospettiva interna: esistono profili di incompatibilità nella comunicazione fra umani e macchine?

Se l'incompatibilità non è legata a presunte deficienze del soggetto artificiale – come l'incapacità di intendere e di essere inteso e l'impossibilità per una macchina di provare interessi –, altre problematiche ricompaiono in riferimento al linguaggio.

Assumere che umani e macchine parlino due linguaggi diversi mette in crisi il percorso fondativo della legittimità di Habermas su base linguistica. A quel punto diventerebbero dirimenti le regole di traduzione e, dunque, sarebbe necessario il riferimento ad una realtà extra-linguistica che si fondi su una universalità dei significati. È solo facendo appello a tale realtà che si può immaginare una autentica possibilità di intesa per soggetti che comunicano in lingue diverse. Ma l'appello a tale realtà extra-linguistica vanifica gli sforzi fatti da Habermas che fonda il suo discorso su un piano puramente linguistico.

La necessità di riferirsi a una realtà extra-linguistica è il frutto di un assunto di partenza: gli umani comunicano tutti nello stesso linguaggio, mentre le macchine comunicano necessariamente in un'altra lingua.

Anche in questo caso, possiamo derivare una obiezione a questa conclusione dal dibattito sulla stanza cinese. In questo caso, ci rifacciamo alla critica nota come "the Other Minds Reply"[3].

Questa critica è formulata nei seguenti termini: come si fa a sapere che qualcuno è in grado di comprendere il nostro linguaggio? Solo dal suo comportamento. Se una macchina è in grado di superare un test comportamentale (in questo caso riuscendo a comprendere efficacemente il nostro linguaggio) tanto quanto qualsiasi altra mente e se si intende attribuire la capacità di comprensione ad altre persone, in linea di principio la si deve attribuire anche alle macchine.

Sempre in linea con una lettura anti-solpsistica della comprensione, l'argomento del linguaggio privato di Wittgenstein (1967) intende dimostrare come una lingua che si rivela incomprensibile per chiunque, eccetto che per il suo utilizzatore originario, è

[3] Si sono occupati di questa critica Dennett (1987), Moravec (1999).

impossibile. La ragione sta nel fatto che una tale lingua risulterebbe incomprensibile anche per il suo utilizzatore originario, giacché egli non potrebbe stabilire un significato per i suoi presunti segni.

Da ciò risulta come, anche attraverso l'obiezione linguistica, il tentativo di invalidare l'accesso delle macchine al discorso si è rivelato fallimentare – o non particolarmente incisivo. Che tra umani e macchine ci sia una irriducibile difformità linguistica non può essere efficacemente asserito.

Non si può, quindi, sostenere che l'emergere di soggettività artificiali sia in linea di principio incompatibile con la teoria del discorso, così come non siamo riusciti a sostenere che umani e macchine parlino due lingue a tal punto diverse da dover trovare alla realtà una dimensione extra-linguistica.

8. CONCLUSIONI

Questa analisi ha inteso affrontare gli eventuali problemi di incompatibilità fra la teoria del discorso di Habermas e l'emergere di soggettività artificiali dotate di una certa autonomia.

Prima di approcciarsi a questa domanda è stato necessario sgomberare il campo da alcune questioni preliminari. La prima ha riguardato il problema del soggetto. Benché Habermas proclami una comunicazione senza identificare precisamente quali siano i soggetti coinvolti, il suo discorso ha bisogno di specifici interlocutori per essere portato avanti. Si è dimostrato, in questo modo, come la domanda sulla compatibilità umano-macchina sia perfettamente rilevante. In secondo luogo, è stato definito il profilo delle macchine che potrebbero far sorgere problemi di compatibilità. Si tratta di macchine che rispondono a regole proprie e il cui comportamento non è per noi perfettamente prevedibile.

Sono poi emerse due ordini di obiezioni diverse. Il primo ha a che fare con il concetto di intesa e con l'annosa questione "se le macchine possono pensare". Il secondo, connesso al primo, riguarda la capacità di una macchina di avere degli interessi. In entrambi i casi, si sono mutuati argomenti dal dibattito sulla stanza cinese di Searle. In particolare, è emerso come obiezioni del genere non siano sufficientemente incisive da escludere l'accesso per le macchine al discorso.

Stabilito che, in linea di principio, non ci sono particolari problemi di compatibilità, è stata affrontata la questione della lingua. In quale lingua macchine e umani parlano? Se si esclude che essi parlino la stessa lingua e si accetta che la teoria del discorso radichi la sua legittimità proprio nei processi linguistici, verrebbe da concludere che l'accesso delle macchine al discorso manda in crisi la teoria del discorso stessa. Tuttavia, alcune posizioni anti-solipstiche, valide sia per le altre menti biologiche che per le macchine,

hanno mostrato come non ci sarebbero problemi nell'assumere una definizione così ampia di linguaggio da far rientrare sotto lo stesso concetto tanto il linguaggio umano quanto quello artificiale.

Seguendo questa ricostruzione, la teoria di Habermas non preclude una pacifica convivenza fra umani e macchine nel discorso che mira all'intesa.

BIBLIOGRAFIA

Benhabib, S. 1996. *Toward a Deliberative Model of Democratic Legitimacy*. In *Democracy and Difference: Contesting the Boundaries of the Political*, a cura di S. Benhabib, S. Princeton: Princeton University Press, pp. 67-94.

Chalmers, D. 1996. *The Conscious Mind*. Oxford: Oxford University Press.

Churchland, P.M., & Churchland, P.S. 1990. *Could a Machine Think?* In «Scientific American», 262(1), pp. 32-39.

Cole, D. J., & Foelber, R. 1984. *Contingent Materialism*. In «Pacific Philosophical Quarterly», 65(1), pp. 74-85.

Dennett, D.C. 1987. *'Fast Thinking'*. In *The Intentional Stance, a cura di Dennet, D.C.* Cambridge: MIT Press, pp. 324-337.

Dryzek, J.S. 2006. *Deliberative Global Politics: Discourse and Democracy in a Divided World*. Cambridge: Polity.

Goodin, R.E. 2008. *Innovating Democracy: Democratic Theory and Practice After the Deliberative Turn*. Oxford: OUP.

Habermas, J. 1997. *Popular Sovereignty as Procedure*. In *Deliberative Democracy: Essays on Reason and Politics*, a cura di Bohman, J., & Rehg, W. Cambridge: MIT press, pp. 35-67.

Habermas, J. 2013. *Fatti e norme*. Bari: Laterza.

Habermas, J. 2017. *Three Normative Models of Democracy*. In *Constitutionalism and Democracy*, a cura di Bellamy, R. New York: Routledge, pp. 277-286.

Mittelstadt, B.D., Allo, P., Taddeo, M., Wachter, S., Floridi, L. 2016. *The Ethics of Algorithms: Mapping the Debate*. In «Big Data & Society», 3 (2).

Moravec, H. 1999. *Robot: Mere Machine to Transcendent Mind.* New York: Oxford University Press.

Pylyshyn, Z.W. 1980. *The 'Causal Power' of Machines*. In «*Behavioral and Brain Sciences*», 3(3), pp. 442-444.

Searle, J. 1999. *The Chinese Room*.

Van Otterlo, M. 2013. *A Machine Learning View on Profiling*. In *Privacy, Due Process and the Computational Turn-Philosophers of Law Meet Philosophers of Technology*, a cura di Hildebrandt, M., & de Vries, K. Abingdon: Routledge, pp. 41-64.

Wittgenstein, L. 1967. *Ricerche filosofiche*. Torino: Einaudi.