

Bruno Cancellieri, ChatGPT e Claude.ai

Fondamenti di agentologia

Per una teoria generale dell'agire

9/7/2026

Copyright © Bruno Cancellieri

Indice

| | |
|--|----|
| Nota..... | 4 |
| Premessa..... | 5 |
| Concetti fondamentali..... | 6 |
| Assiomi..... | 7 |
| Definizioni derivate..... | 8 |
| Prime proposizioni..... | 10 |
| Principio generale..... | 12 |
| Metodo agentologico..... | 13 |
| Applicazione alle relazioni sociali..... | 14 |
| Applicazione psicoterapeutica e autoterapeutica..... | 16 |
| Un caso applicato..... | 18 |
| Osservazione conclusiva..... | 20 |
| Appendice: Articolo di presentazione di questo saggio..... | 21 |

Nota

Questo saggio nasce da un dialogo con ChatGPT, sviluppato secondo un metodo che ho adottato già per altri saggi: ho proposto un titolo e un tema, ChatGPT mi ha suggerito un indice, che ho poi modificato per includere aspetti che ritenevo rilevanti. Concordato l'indice, ho fatto scrivere i capitoli uno alla volta, correggendo di volta in volta la terminologia e i contenuti.

Il quadro teorico specifico di questo saggio — i concetti primitivi di agente, situazione, percezione, programma e azione, l'apparato di assiomi e definizioni derivate che li vincola, e il metodo agentologico che ne deriva — è stato elaborato in questo dialogo. Esso si inserisce tuttavia in una formazione psicologica e filosofica che avevo sviluppato in autonomia ben prima di dialogare con ChatGPT, e di cui un precedente saggio, dal titolo *Psicologia dei bisogni*, che ho scritto senza aiuti esterni, costituisce testimonianza.

Questo saggio dialoga inoltre con *Psicologia dell'interagire*, un saggio che applica una logica affine — situazione relazionale, interpretazione, motivazione, logica psicologica, strategia, transazione, meta-interazione — alle interazioni umane. Il capitolo *Applicazione alle relazioni sociali* esplicita la corrispondenza tra i due apparati concettuali.

Claude.ai è stato utilizzato in questo saggio per correggere i testi prodotti da ChatGPT attraverso una discussione articolata, capitolo per capitolo, nella quale le modifiche sono state proposte, argomentate e concordate insieme.

La responsabilità delle tesi sostenute in questo saggio, e delle loro eventuali debolezze, resta comunque mia.

Bruno Cancellieri

Premessa

L'agentologia è una disciplina che si propone di descrivere e analizzare i fenomeni attraverso il concetto di agente.

Essa parte dall'ipotesi che molti fenomeni possano essere compresi più facilmente se, anziché chiedersi che cosa siano gli enti, ci si domanda quali agenti siano coinvolti, quali azioni compiano, quali obiettivi perseguano e come le loro azioni modifichino le situazioni in cui si trovano.

Il lettore troverà in questo apparato punti di contatto evidenti con tradizioni già esistenti — la teoria generale dei sistemi, la cibernetica, la nozione di agente razionale dell'intelligenza artificiale — che hanno già descritto organismi, macchine e organizzazioni in termini di percezione, decisione e azione. È bene dirlo subito, non solo in chiusura: l'agentologia non rivendica un oggetto nuovo, ma un apparato — pochi concetti primitivi, un insieme dichiarato di assiomi, un criterio esplicito per stabilire quando il linguaggio si applica e quando no. Il suo valore, se ce n'è uno, non sta nel descrivere fenomeni che quelle tradizioni non saprebbero già descrivere, ma nel rendere comune ed esplicito un apparato che esse condividono spesso solo implicitamente, e nel permettere di applicarlo, senza modifiche sostanziali, a domini che di norma restano separati — dagli organismi alle organizzazioni, dalle relazioni personali agli agenti artificiali. Il capitolo Un caso applicato, più avanti, mette alla prova questa pretesa su un esempio concreto.

L'agentologia è costruita come una teoria sistematica, organizzata in quattro livelli: concetti fondamentali, assiomi, definizioni derivate, proposizioni.

I concetti fondamentali (primo livello) non sono definizioni in senso pieno, ma termini primitivi la cui portata esatta è fissata dagli assiomi (secondo livello) che li vincolano reciprocamente. Le definizioni derivate (terzo livello) introducono concetti costruiti a partire dai primitivi. Le proposizioni (quarto livello) sono affermazioni argomentate a partire dagli assiomi e dalle definizioni, con un grado di rigore che il testo dichiara esplicitamente di volta in volta.

L'obiettivo dell'agentologia è offrire un linguaggio generale, sufficientemente astratto da poter essere applicato allo studio degli esseri viventi, delle organizzazioni, delle società e degli agenti artificiali — senza pretendere di sostituirsi alle discipline che già se ne occupano.

Concetti fondamentali

I concetti fondamentali costituiscono il nucleo teorico dell'agentologia. Agente, situazione, percezione, programma e azione sono qui assunti come termini primitivi: non permettono una definizione non circolare. Le formulazioni che seguono sono dunque esplicitazioni intuitive, utili a orientare la lettura; il significato operativo di ciascun termine è fissato soltanto dagli assiomi che lo vincolano.

Agente

Un agente è un'entità dotata di uno stato interno e di almeno un programma, capace di percepire almeno in parte la situazione in cui si trova e di compiere azioni che tendono a modificarla o a conservarla.

Situazione

Una situazione è la configurazione, in un determinato istante, dell'ambiente esterno e dello stato interno degli agenti coinvolti.

Percezione

La percezione è il processo mediante il quale un agente acquisisce informazioni sulla situazione. Le percezioni modificano lo stato interno dell'agente.

Programma

Un programma è una struttura interna, parzialmente stabile nel tempo, che vincola l'insieme delle azioni che un agente può produrre a partire da una data percezione: esclude alcune azioni, ne rende altre più probabili, senza necessariamente determinarne una sola in modo univoco.

Questa definizione ha una conseguenza precisa: un'entità le cui azioni fossero del tutto indipendenti da qualunque struttura interna — puramente casuali rispetto a ciò che percepisce — non possiederebbe un programma e, per A4, non sarebbe un agente. La nozione di programma non è dunque vuota: esclude dei casi.

Costituiscono esempi tipici di programmi i meccanismi genetici, i riflessi, gli istinti, le emozioni, le abitudini, le conoscenze, le strategie, i ragionamenti, gli algoritmi. L'elenco è aperto, ma il criterio che definisce l'appartenenza a questa categoria — vincolare in modo relativamente stabile il rapporto tra percezione e azione — non lo è.

Azione

Un'azione è qualsiasi modificazione della situazione prodotta da un agente, oppure qualsiasi comportamento finalizzato a conservarla. Ogni azione produce uno o più effetti, intenzionali o non intenzionali: la distinzione tra i due tipi di effetto è ripresa nella definizione di responsabilità (D13).

Assiomi

Gli assiomi che seguono non hanno tutti la stessa funzione logica. I primi quattro sono costitutivi: insieme ai concetti fondamentali, fissano che cosa si intende per agente e per situazione, e la loro negazione equivarrebbe a cambiare argomento più che a proporre una tesi falsa. I successivi cinque sono operativi: affermano dipendenze, dinamiche e regole di composizione che avrebbero potuto, in linea di principio, essere diverse. Il termine assioma è mantenuto per entrambi i gruppi — in un sistema costruito su primitivi non definiti è la funzione ordinaria di un assioma vincolarne il significato — ma la distinzione tra le due funzioni viene qui dichiarata esplicitamente.

Assiomi costitutivi

- A1.** Esistono agenti.
- A2.** Ogni agente si trova sempre in una situazione.
- A3.** Ogni agente percepisce almeno una parte della situazione in cui si trova.
- A4.** Ogni agente possiede almeno un programma.

Assiomi operativi

- A5.** Il comportamento di un agente in un dato istante dipende dalla situazione percepita, dal suo stato interno e dai suoi programmi.
- A6.** I programmi di un agente possono modificarsi nel tempo (programmi plastici) o restare invariati (programmi rigidi).
- A7.** Un agente può, mediante le proprie azioni, modificare la situazione, altri agenti oppure se stesso.
- A8.** Un insieme di agenti costituisce, esso stesso, un agente di ordine superiore se e solo se a livello dell'insieme emergono una percezione e un'azione non riducibili alla semplice somma delle percezioni e delle azioni dei singoli componenti — per esempio una procedura di decisione comune, un confine riconoscibile rispetto all'esterno, una persistenza nel tempo distinta da quella dei singoli componenti. In assenza di queste condizioni, l'insieme resta una pluralità di agenti, non un agente.
- A9.** Un agente può appartenere contemporaneamente a più agenti di ordine superiore.

Definizioni derivate

D1. Informazione — Qualsiasi differenza acquisita mediante una percezione e registrata nello stato interno di un agente.

D2. Stato interno — Insieme delle informazioni, dei programmi e delle altre caratteristiche interne possedute da un agente in un determinato istante.

D3. Decisione — Processo mediante il quale un programma seleziona un'azione tra quelle che la situazione percepita rende disponibili (cfr. D12, Libertà).

D4. Obiettivo — Situazione che un agente tende a realizzare, mantenere oppure evitare.

D5. Memoria — Parte dello stato interno che conserva informazioni nel tempo.

D6. Apprendimento — Modificazione di uno o più programmi in conseguenza dell'esperienza; è il caso, previsto da A6, di un programma plastico che si modifica.

D7. Interazione — Sequenza di azioni mediante la quale due o più agenti modificano reciprocamente le rispettive situazioni.

D8. Comunicazione — Interazione (D7) nella quale almeno un'azione ha come obiettivo specifico (D4), e non come semplice effetto collaterale, la modificazione dello stato interno di un altro agente mediante un segnale regolato da un codice condiviso tra gli agenti coinvolti. Questa clausola distingue la comunicazione dalla generalità delle interazioni: non ogni interazione che modifica uno stato interno altrui è comunicazione, ma solo quella in cui tale modificazione è perseguita intenzionalmente attraverso un codice.

D9. Cooperazione — Interazione (D7) nella quale due o più agenti perseguono almeno un obiettivo comune (D4) e le rispettive azioni sono reciprocamente funzionali al suo raggiungimento.

D10. Competizione — Interazione nella quale il raggiungimento dell'obiettivo di un agente riduce la probabilità che un altro agente raggiunga il proprio.

Nota. Cooperazione e competizione non sono categorie reciprocamente esclusive: una stessa interazione può essere cooperativa rispetto a un obiettivo e competitiva rispetto a un altro (interazione mista) — è il caso ordinario delle relazioni tra organizzazioni che collaborano su un progetto e competono su un mercato.

D11. Potere — Capacità di un agente A di modificare in modo rilevante e duraturo lo stato o il comportamento di un agente B, anche contro le preferenze di B o superando la resistenza che B eventualmente oppone. Questa clausola distingue il potere dalla generica capacità di azione già garantita dall'assioma A7: non ogni azione che modifica un altro agente è esercizio di potere, ma solo quella che prevale su una preferenza o resistenza contraria.

D12. Libertà — Insieme delle azioni che un agente percepisce come concretamente disponibili nella situazione in cui si trova e che è in grado di eseguire con i programmi di cui dispone. La libertà così definita è relativa alla percezione dell'agente: un'azione oggettivamente possibile ma non percepita come tale non rientra, per questo agente, nella sua libertà. La divergenza tra le azioni percepite come disponibili e quelle oggettivamente disponibili a un osservatore esterno è essa stessa un fenomeno agentologicamente rilevante — per esempio nei casi di ignoranza o di autoinganno.

D13. Responsabilità — Attribuzione a un agente degli effetti di una propria azione, graduata in base a due fattori: il grado di intenzionalità dell'effetto (cfr. Azione) e l'ampiezza delle alternative che gli erano concretamente disponibili al momento della decisione (D12, Libertà). Un effetto pienamente intenzionale, prodotto in presenza di alternative disponibili, fonda un'attribuzione di responsabilità piena; l'assenza di alternative o di intenzionalità la attenua o la esclude.

D14. Relazione — Legame relativamente stabile tra due o più agenti, costituito dal sedimentarsi nel tempo di una molteplicità di interazioni (D7) tra essi, tale da orientare in modo riconoscibile le interazioni successive. Una relazione non è un'ulteriore interazione, ma lo stato interno — condiviso o reciprocamente rappresentato — che le interazioni ripetute depositano negli agenti coinvolti (cfr. D2, Stato interno).

Prime proposizioni

Nota metodologica. Le dimostrazioni che seguono non impiegano un calcolo logico formalizzato: mostrano, passo per passo, che ciascuna proposizione è conseguenza diretta degli assiomi e delle definizioni citati esplicitamente, senza introdurre contenuto ulteriore. Il lettore è invitato a verificare, per ciascun passaggio, che le premesse citate bastino davvero a produrre la conclusione.

P1. Ogni agente agisce sempre all'interno di una situazione.

Dimostrazione. Per A2, ogni agente si trova sempre in una situazione. Per definizione di agente, un agente è capace di compiere azioni. Le due premesse, congiunte, danno la tesi.

P2. Ogni agente contribuisce continuamente a trasformare oppure a conservare la situazione in cui si trova.

Dimostrazione. Per la definizione di Azione, ogni azione modifica oppure tende a conservare una situazione. Per la definizione di Agente, un agente è definito dalla sua capacità di agire. Dunque, finché agisce, ogni agente contribuisce all'evoluzione o alla conservazione della situazione.

P3. Due agenti possono reagire diversamente alla stessa situazione percepita.

Dimostrazione. Per A5, il comportamento dipende dalla situazione percepita, dallo stato interno e dai programmi dell'agente. Due agenti possono avere stati interni o programmi differenti (A6). A parità di situazione percepita, se stato interno o programmi differiscono, A5 non impone lo stesso comportamento: comportamenti diversi sono dunque compatibili con gli assiomi.

P4. Ogni apprendimento modifica il comportamento potenziale dell'agente.

Dimostrazione. Per D6, l'apprendimento è la modificazione di uno o più programmi. Per A5, i programmi concorrono a determinare il comportamento. Modificare un programma modifica quindi, per A5, l'insieme dei comportamenti che quel programma può produrre.

P5. Ogni fenomeno collettivo che coinvolga una pluralità di agenti può essere descritto come una rete di interazioni tra essi.

Dimostrazione. Per definizione, un fenomeno collettivo coinvolge più agenti. Le relazioni tra agenti che si modificano reciprocamente sono, per D7, interazioni. Una pluralità di agenti in relazione reciproca costituisce quindi, per costruzione, una rete di interazioni. Si noti che questa proposizione descrive una possibilità di descrizione, non afferma che ogni fenomeno collettivo sia riducibile a tale descrizione senza perdita di informazione: cfr. la clausola di demarcazione del Principio generale.

P6. Non ogni aggregato di agenti costituisce un agente di ordine superiore.

Dimostrazione. Per A8, un aggregato costituisce un agente di ordine superiore solo se a livello dell'aggregato emergono percezione e azione non riducibili alla somma dei componenti. Esistono aggregati di agenti — per esempio una folla priva di qualunque procedura di decisione comune — in cui questa condizione non è soddisfatta. Per tali aggregati, A8 nega lo status di agente di ordine superiore.

P7. Modificando la percezione di un agente si possono modificare le sue azioni senza modificare i suoi programmi.

Dimostrazione. Per A5, il comportamento di un agente dipende dalla situazione percepita, oltre che dal suo stato interno e dai suoi programmi. Affermare una dipendenza genuina da un fattore comporta che esista almeno una coppia di percezioni diverse che, a parità di stato interno e di programmi, producano azioni diverse: se così non fosse per nessuna coppia di percezioni, la percezione non giocherebbe alcun ruolo nel determinare il comportamento, e A5 sarebbe vuoto rispetto a quel fattore. Ne segue che è

possibile modificare l'azione di un agente intervenendo sulla sua percezione, senza toccare né il suo stato interno né i suoi programmi.

Principio generale

Ogni fenomeno che esibisca una configurazione non arbitraria di percezione, decisione e azione (A3-A5) può essere descritto come il risultato delle azioni di uno o più agenti che, sulla base delle proprie percezioni, dei propri programmi e del proprio stato interno, modificano continuamente le situazioni in cui sono immersi.

Questo principio ha lo statuto di un principio regolativo, non di un'ipotesi empirica generale: indica un modo possibile e spesso utile di descrivere i fenomeni, non impone che ogni fenomeno debba essere descritto in questi termini, né afferma che la descrizione agentologica sia sempre la più informativa.

Il principio ha inoltre un confine esplicito: non si applica ai fenomeni per i quali non sia possibile identificare, se non arbitrariamente, un confine, una percezione e un repertorio di azioni — per esempio i processi fisici privi di retroazione, come l'erosione di una roccia per opera del vento. In questi casi l'agentologia non ha nulla da aggiungere alla descrizione fisica ordinaria, e non pretende di sostituirla.

Metodo agentologico

L'analisi agentologica di un fenomeno richiede anzitutto di verificare che il fenomeno soddisfi le condizioni del Principio generale — che sia cioè identificabile una configurazione non arbitraria di agenti, percezioni e azioni. Se questa condizione non è soddisfatta, il metodo non si applica.

Se la condizione è soddisfatta, l'analisi consiste nel rispondere ordinatamente alle seguenti domande.

- Quali sono gli agenti coinvolti?
- Qual è la situazione iniziale?
- Quali informazioni percepisce ciascun agente?
- Quali programmi risultano rilevanti?
- Quali obiettivi persegue ciascun agente?
- Quali azioni compie?
- Come tali azioni modificano la situazione?
- Quale nuova situazione si genera?

L'analisi procede quindi secondo il ciclo fondamentale dell'agentologia:

Situazione → Percezione → Programma → Azione → Nuova situazione

Applicazione alle relazioni sociali

I capitoli precedenti hanno costruito l'apparato descrittivo generale dell'agentologia. Questo capitolo e il successivo ne propongono due applicazioni a domini diversi: le relazioni tra agenti distinti (questo capitolo) e i rapporti tra i programmi interni a un medesimo agente (il capitolo successivo). In entrambi i casi si tratta di estensioni applicative, non di ulteriori assiomi: introducono concetti — qui, il benessere legato alla qualità delle relazioni; nel capitolo successivo, il conflitto e la disfunzione — che il nucleo assiomatico, deliberatamente descrittivo, non contiene.

Il legame con Psicologia dell'interagire

Questo saggio dialoga da vicino con un lavoro precedente, Psicologia dell'interagire, dedicato specificamente alle interazioni umane. Quel saggio sostiene, nella propria premessa, che buona parte delle nostre gioie e dei nostri dispiaceri dipende dalla qualità delle relazioni sociali, che a loro volta dipendono dalla qualità delle interazioni quotidiane che le costruiscono. È una tesi ampiamente sostenuta dalla ricerca sul benessere soggettivo, ma va intesa come l'indicazione di un fattore rilevante tra altri — personalità, condizioni materiali, salute — piuttosto che come una gerarchia rigida tra le cause della felicità.

Corrispondenza tra i due apparati concettuali

L'apparato di Psicologia dell'interagire — situazione relazionale, interpretazione, motivazione, logica psicologica, strategia, transazione — può essere riletto come un'istanza specifica, propria delle interazioni umane, del vocabolario più generale qui costruito.

| Psicologia dell'interagire | Fondamenti di agentologia |
|---|---|
| Situazione relazionale | Situazione, ristretta alla configurazione di ruoli, aspettative e rapporti di potere (D11) tra agenti |
| Interpretazione | Informazione (D1) fornita dalla percezione, elaborata dai programmi cognitivi dell'agente |
| Motivazione (bisogni, desideri, valori) | Sottoinsieme dei programmi che, incontrando l'interpretazione, genera gli obiettivi (D4) |
| Logica psicologica | Programma specializzato nel selezionare come perseguire un obiettivo dato (D3, Decisione) |
| Strategia | Sequenza di azioni che tale programma seleziona |
| Transazione | Una singola azione, nel caso in cui sia percepibile da un altro agente |
| Interazione | Coincide quasi letteralmente con D7 |
| Relazione | D14 |

Questa corrispondenza non toglie nulla alla specificità di Psicologia dell'interagire: la fenomenologia del vissuto immediato, la distinzione tra bisogni, desideri e valori, le configurazioni triadiche restano contenuti propri di quel saggio, che l'agentologia non pretende di ricostruire. Ciò che l'agentologia aggiunge è un linguaggio comune capace di collocare quel modello, insieme ad altri, dentro un apparato applicabile senza modifiche sostanziali a organizzazioni e ad agenti artificiali.

Relazione e rete sociale

Psicologia dell'interagire distingue un livello ulteriore, la relazione: il legame che si sedimenta quando le stesse interazioni si ripetono nel tempo tra gli stessi individui. L'agentologia formalizza questo livello nella definizione D14: una relazione è lo stato interno che una molteplicità di interazioni ripetute deposita negli agenti coinvolti, capace di orientare riconoscibilmente le interazioni successive.

Una rete sociale — l'insieme delle relazioni che collegano più agenti — non costituisce automaticamente un agente di ordine superiore. Vale qui lo stesso criterio già introdotto in A8 e verificato in P6: perché una rete di relazioni acquisti essa stessa lo statuto di agente occorre che emergano, a livello della rete, una percezione e un'azione non riducibili alla somma delle relazioni che la compongono — per esempio una procedura di decisione collettiva o un confine riconoscibile. In assenza di queste condizioni, la rete resta ciò che P5 chiama una rete di interazioni, non un agente.

La meta-interazione riletta in termini agentologici

Psicologia dell'interagire propone infine la meta-interazione: un'interazione che ha come oggetto un'altra interazione, mediante la quale i partecipanti ricostruiscono insieme le interpretazioni, le motivazioni e le logiche che hanno generato un'interazione precedente. Nel vocabolario di questo saggio, una meta-interazione è una comunicazione (D8) il cui obiettivo condiviso (D4, D9) è ricostruire congiuntamente le percezioni, i programmi e le decisioni (D1, D3) che hanno generato una precedente interazione (D7) tra gli stessi agenti.

Definita in questi termini, la meta-interazione non è più soltanto una pratica pensata per le relazioni umane: è un caso particolare di un'operazione generale — la comunicazione riflessiva su un'interazione passata — applicabile in linea di principio a qualunque coppia o gruppo di agenti, umani o artificiali, capaci di comunicazione (D8) e di un accesso, per quanto parziale, al proprio stato interno passato (D5, Memoria). Se buona parte del benessere individuale dipende dalla qualità delle relazioni (D14), e la qualità di una relazione dipende dalla qualità delle interazioni che la costruiscono, la meta-interazione — nella sua forma generale qui proposta — è lo strumento che l'agentologia mette a disposizione per intervenire deliberatamente su tale qualità.

Applicazione psicoterapeutica e autoterapeutica

Il capitolo precedente ha applicato l'agentologia ai rapporti tra agenti distinti. Questo capitolo applica lo stesso apparato ai rapporti tra i programmi interni a un medesimo agente, per costruire un linguaggio utile alla comprensione — e, quando possibile, al mutamento — dei conflitti motivazionali. Come nel capitolo precedente, si tratta di un'estensione applicativa: introduce concetti valutativi (conflitto, disfunzione) assenti dal nucleo assiomatico, e li dichiara come tali.

Conflitto intra-agente

T1. Conflitto intra-agente — Situazione in cui due o più programmi di uno stesso agente, attivati dalla medesima percezione, tendono verso azioni incompatibili tra loro. È la generalizzazione di D10 (Competizione) al caso in cui i termini in competizione non sono agenti distinti, ma componenti interne a un solo agente.

Programmi e sub-agenti interni

Non tutti i programmi in conflitto meritano lo stesso trattamento teorico. Un riflesso o un'abitudine elementare sono programmi semplici: non c'è, nel trattarli come tali, alcuna pretesa ulteriore. Alcune strutture interne, tuttavia — schemi consolidati, disposizioni affettive persistenti, cioè che tradizioni cliniche diverse chiamano parti, stati dell'Io o sotto-personalità — sono abbastanza stabili, autoreferenziali e orientate a un proprio obiettivo (D4) da poter essere trattate, utilmente, come se fossero piccoli agenti che perseguono un fine proprio all'interno dell'agente maggiore.

T2. Sub-agente interno — Un programma trattato, per utilità descrittiva, come se possedesse una propria percezione e un proprio obiettivo, in ragione della stabilità e dell'autonomia relativa che mostra rispetto alle altre componenti del programma dell'agente.

Questa mossa ha uno statuto dichiaratamente metaforico, non ontologico: attribuire a un programma una percezione e un obiettivo propri è un dispositivo modellistico — vicino a ciò che altrove si è chiamato "stance intenzionale" — utile a rendere pensabile il conflitto, non la scoperta che dentro un agente esistano letteralmente più agenti nel senso di A1. Il criterio per decidere quando questa mossa è utile ricalca, in scala ridotta, quello di A8: un programma merita di essere trattato come sub-agente quando mostra, con una certa persistenza, una propria linea di condotta difficilmente riducibile alle altre componenti del programma dell'agente.

Potere intra-agente e acrasia

T3. Potere intra-agente — Caso particolare di D11 (Potere) in cui A e B sono due programmi, o due sub-agenti interni (T2), di uno stesso agente: un programma prevale in modo rilevante e duraturo sulle azioni che un altro programma dello stesso agente avrebbe altrimenti prodotto, anche contro l'obiettivo (D4) che quest'ultimo persegue.

Questa nozione dà un linguaggio formale al fenomeno classico, discusso da Aristotele sotto il nome di acrasia e in filosofia contemporanea come debolezza della volontà: un programma — un impulso, un'abitudine, una compulsione — esercita potere (T3) su un altro programma, quello che l'agente riconoscerebbe, riflettendo, come il proprio obiettivo prevalente, senza che a ciò corrisponda l'intervento di un agente esterno. L'assioma A7, che garantisce a ogni agente la capacità di modificare se stesso, non specifica se tale modificazione sia sempre coerente al suo interno: T3 rende conto proprio dei casi in cui non lo è.

Il cambiamento terapeutico come apprendimento diretto

Il cambiamento che una psicoterapia persegue può essere riletto come un'istanza intenzionale di D6 (Apprendimento): la modificazione di uno o più programmi in conseguenza dell'esperienza — qui,

dell'esperienza specifica costituita dal lavoro terapeutico stesso. Per A6, un programma può essere plastico o rigido; la terapia lavora sui programmi plastici, e una parte non secondaria del lavoro terapeutico consiste proprio nello stabilire quali programmi siano, in un dato momento, effettivamente plastici e quali no.

Ciò che distingue questo apprendimento da un apprendimento qualunque è il suo punto di partenza: non l'acquisizione di un'informazione nuova sull'ambiente esterno, ma una percezione più accurata (D1) di un conflitto (T1) che il programma precedente teneva implicito. Rendere esplicito un conflitto intra-agente — dargli il nome dei suoi termini, riconoscere quali sub-agenti interni (T2) vi siano coinvolti e quali obiettivi perseguano — è spesso, di per sé, la condizione che rende un programma altrimenti rigido di nuovo plastico.

Il metodo applicato a sé stessi

Le domande del metodo agentologico possono essere riformulate in prima persona, e rivolte non a un fenomeno esterno ma ai propri sub-agenti interni: quali sub-agenti sono coinvolti in questo conflitto? Qual è la situazione che lo attiva? Che cosa percepisce ciascuno di essi? Quali programmi, in ciascuno, risultano rilevanti? Quale obiettivo persegue ciascuno? Quale azione tende a produrre? Come queste azioni, in competizione tra loro, modificano il mio comportamento effettivo? Quale nuova situazione — interna, non solo esterna — ne risulta?

Applicato a sé stessi, questo esercizio non richiede necessariamente un contesto clinico: può essere condotto per iscritto, come pratica riflessiva ordinaria.

Il problema specifico dell'autoterapia

L'applicazione a sé stessi incontra però una difficoltà che l'applicazione a un fenomeno esterno non incontra: l'osservatore e l'osservato coincidono. La percezione su cui l'autoanalisi si fonda è prodotta dagli stessi programmi che l'autoanalisi vorrebbe esaminare, ed è quindi esposta a essere distorta esattamente nei punti più rilevanti. Questo saggio ha già incontrato il problema, a proposito di D12 (Libertà): la divergenza tra le azioni che un agente percepisce come disponibili e quelle oggettivamente disponibili è, si è detto, il caso tipico dell'autoinganno. Nell'autoterapia questo rischio non è un'eccezione, ma la condizione ordinaria di partenza.

Alcuni accorgimenti possono attenuare, senza eliminare, il problema: l'esternalizzazione tramite scrittura, che rende la propria ricostruzione un oggetto stabile da poter rileggere e mettere in discussione a distanza di tempo; la rilettura differita, che permette di confrontare la ricostruzione fatta a caldo con una lettura più distaccata; il confronto con un terzo — un professionista o, con i limiti che comporta, una persona di fiducia — che non condivide gli stessi programmi e può quindi percepire ciò che l'agente, da solo, sistematicamente non percepisce.

Confini d'uso

Come il Principio generale ha una clausola di demarcazione esplicita, anche questa applicazione ne richiede una propria. L'agentologia, in questo capitolo, offre un linguaggio per descrivere e riflettere su conflitti motivazionali ordinari — l'ambivalenza, la procrastinazione, gli schemi relazionali disfunzionali di intensità moderata. Non offre, e non pretende di offrire, uno strumento diagnostico o terapeutico per condizioni che richiedono un trattamento clinico — trauma, psicosi, depressione maggiore, disturbi che compromettono la stessa capacità di percepire accuratamente la propria situazione (A3). In questi casi la mediazione di un professionista non è un'opzione tra le altre, ma la condizione perché qualunque analisi, agentologica o no, possa dirsi affidabile.

Un caso applicato

I due capitoli precedenti hanno introdotto concetti applicativi in astratto. Questo capitolo li mette alla prova su un unico esempio, per mostrare concretamente che cosa l'apparato agentologico permette di vedere — non in linea di principio, ma in un caso specifico — che una descrizione generica in termini di sistemi e retroazioni lascerebbe implicito.

Marco lavora da tre anni nel reparto diretto da Giulia. Da alcuni mesi consegna sistematicamente in ritardo i rapporti settimanali, pur sapendo che il ritardo irrita Giulia e pesa sulla sua valutazione annuale. Giulia, dal canto suo, ha reagito ai ritardi intensificando i controlli e chiedendo aggiornamenti quotidiani — cosa che Marco vive come sfiducia, e che rafforza, invece di attenuare, la sua tendenza a rimandare. La relazione tra i due (D14), costruita in tre anni di collaborazione perlopiù cooperativa (D9), si è progressivamente caricata di elementi specifici di segno negativo.

Applicando il metodo agentologico: gli agenti coinvolti sono Marco e Giulia — il reparto nel suo complesso non soddisfa, in questo caso, il criterio di A8, perché non esiste una procedura decisionale che coinvolga il reparto come tale, indipendente dalle singole persone, e resta dunque una pluralità, non un agente di ordine superiore, come previsto da P6. La situazione iniziale è una relazione (D14) sedimentata su una storia di collaborazione, ma con un'interpretazione recente asimmetrica: Marco percepisce i controlli di Giulia come un giudizio sulla propria affidabilità; Giulia percepisce i ritardi di Marco come segno di disimpegno.

I programmi rilevanti in Marco includono sia un obiettivo professionale esplicito — mantenere una buona valutazione — sia un'abitudine di evitamento che si attiva di fronte a compiti percepiti come fonte di giudizio negativo: i due programmi sono in conflitto (T1), e l'abitudine di evitamento esercita di fatto potere (T3) sul programma professionale, prevalendo su di esso nonostante l'obiettivo dichiarato di Marco resti quello di consegnare puntualmente — un caso di acrasia nel senso qui definito. L'azione osservabile — il ritardo ripetuto — è l'esito di un conflitto intra-agente che rimane implicito finché nessuno lo nomina.

Dal lato di Giulia, l'azione di intensificare i controlli è un esercizio di potere (D11) che, incontrando l'interpretazione di Marco, produce l'effetto opposto a quello inteso: anziché ripristinare l'affidabilità, rafforza il programma di evitamento, perché conferma a Marco l'interpretazione che teme. L'interazione tra i due (D7), pur radicata in una relazione a base cooperativa (D9), assume localmente una dinamica competitiva (D10) — esattamente il caso di interazione mista già previsto nella nota a D9-D10.

Il punto di svolta arriva quando Giulia propone un confronto diretto su ciò che sta accadendo tra loro: una meta-interazione (D8), nel senso definito nel capitolo Applicazione alle relazioni sociali. Nel corso del confronto emerge che Marco non aveva mai percepito, tra le azioni a sua disposizione, quella di spiegare apertamente le difficoltà di carico di lavoro che stavano alla base dei ritardi: non perché gli fosse oggettivamente preclusa, ma perché non rientrava nella sua libertà percepita (D12) — un caso, appunto, di divergenza tra azioni percepite come disponibili e azioni oggettivamente disponibili, di cui la definizione di D12 già avvertiva.

L'esito immediato del confronto è un cambiamento di percezione, non ancora un cambiamento di programma: Marco reinterpreta i controlli di Giulia come preoccupazione anziché come sfiducia, e Giulia reinterpreta i ritardi come segnale di sovraccarico anziché di disimpegno. Per P7, questo basta a modificare le azioni successive di entrambi — Marco ricomincia a consegnare puntualmente, Giulia allenta i controlli — senza che i rispettivi programmi (l'abitudine di evitamento di Marco, lo stile di controllo di Giulia) siano stati ancora toccati alla radice. Se il nuovo schema si stabilizza nel tempo, si avrà inoltre, per D6, un vero apprendimento: i programmi stessi, non solo le azioni contingenti, risulteranno modificati; ma questo è un esito ulteriore, non garantito dal solo cambiamento di percezione.

Che cosa aggiunge l'apparato

Una descrizione puramente sistemica di questo episodio — un ciclo di retroazione negativa tra i comportamenti di Marco e Giulia — non sarebbe stata falsa, ma avrebbe lasciato indistinti alcuni elementi che qui risultano invece separabili: la differenza tra l'esercizio di potere (D11) e la semplice capacità di agire che l'assioma A7 garantisce a chiunque; la differenza tra ciò che un agente può fare oggettivamente e ciò che percepisce di poter fare (D12), che è il punto preciso su cui il confronto ha agito; il conflitto tra due programmi dello stesso agente (T1-T3), che spiega perché Marco non riuscisse a fare ciò che lui stesso avrebbe voluto fare, senza dover invocare altro che l'apparato già introdotto; e la constatazione, niente affatto scontata, che la meta-interazione ha funzionato modificando percezioni (P7), non necessariamente i programmi sottostanti. Sono distinzioni che un lessico di soli input, output e retroazione non è attrezzato a rendere altrettanto esplicite.

Osservazione conclusiva

L'agentologia propone un cambiamento di prospettiva: anziché considerare gli enti come elementi fondamentali della descrizione del mondo, considera fondamentali gli agenti e le loro azioni.

Questo punto di partenza non è isolato, e la Premessa lo ha anticipato: l'agentologia condivide le proprie fonti con la cibernetica, la teoria generale dei sistemi e la nozione di agente razionale dell'intelligenza artificiale. Il capitolo Un caso applicato ha cercato di mostrare, su un esempio concreto, che cosa l'apparato assiomatico aggiunge a quelle tradizioni: non un oggetto nuovo, ma un linguaggio comune ed esplicito, verificabile passo per passo.

Ogni fenomeno che soddisfi le condizioni del Principio generale viene così interpretato come il risultato dinamico dell'interazione tra agenti dotati di programmi, inseriti in situazioni mutevoli e capaci di modificare il mondo, gli altri agenti e se stessi.

L'agentologia non si propone di sostituire le discipline esistenti, ma di offrire loro un linguaggio concettuale comune, e un criterio esplicito — la configurazione non arbitraria di percezione, decisione e azione — per stabilire quando questo linguaggio si applica e quando no.

I capitoli applicativi che precedono questa osservazione — sulle relazioni sociali, sull'uso psicoterapeutico e, soprattutto, il caso applicato che li ha ripresi entrambi in un solo esempio — sono un primo tentativo di rispondere a una domanda più esigente di quella a cui finora si è risposto: non solo se l'apparato sia internamente coerente, ma se renda visibile qualcosa che una descrizione più generica, in termini di sistemi e retroazioni, lascerebbe implicito. Il lettore è nella posizione migliore per giudicare se questo tentativo sia riuscito.

Appendice: Articolo di presentazione di questo saggio

Articolo pubblicato nel blog *Il mondo visto da me* (<https://blog2.cancellieri.org/dagli-enti-agli-agenti-nasce-lagentologia/>)

Marco rimanda da mesi la consegna dei suoi rapporti settimanali. Sa che questo infastidisce Giulia, la sua responsabile, e sa che pesa sulla sua valutazione. Eppure continua a rimandare. Giulia, dal canto suo, ha reagito stringendo i controlli — il che, paradossalmente, non fa che peggiorare le cose.

È una scena da ufficio qualunque. Ma è anche un piccolo laboratorio per una domanda più generale: come descriviamo ciò che accade quando qualcuno percepisce una situazione, la interpreta, decide come muoversi, agisce — e nel farlo modifica, spesso senza volerlo, la situazione in cui si trova insieme a un altro?

È da una domanda come questa che parte *Fondamenti di agentologia*, il mio nuovo saggio, appena concluso dopo un lungo lavoro a più voci — mio, di ChatGPT e di Claude.

Un cambio di prospettiva minimo, ma non banale

L'idea di partenza è semplice da enunciare: invece di chiederci che cosa siano le cose — enti, oggetti, sostanze — proviamo a chiederci quali agenti sono coinvolti in un fenomeno, che cosa percepiscono, quali obiettivi perseguono, quali azioni compiono e come quelle azioni trasformano la situazione. Un'azienda, una coppia, un animale, un programma informatico: in tutti questi casi, secondo l'agentologia, conviene guardare non alla “natura” della cosa, ma a ciò che fa e a ciò che gli succede intorno mentre lo fa.

Non è un'idea del tutto nuova — chi conosce la cibernetica, la teoria dei sistemi o l'intelligenza artificiale vi riconoscerà un'aria di famiglia, e il saggio lo dice esplicitamente fin dalle prime pagine, senza far finta di inventare qualcosa dal nulla. Quello che il saggio prova ad aggiungere non è l'idea in sé, ma la sua costruzione: pochi concetti di partenza — agente, situazione, percezione, programma, azione — da cui derivare, con passaggi dichiarati uno per uno, tutto il resto: che cosa sia il potere, che cosa sia la libertà, che cosa distingua una relazione da una singola interazione, quando un gruppo di persone diventi davvero “un” agente collettivo e quando resti soltanto una pluralità.

Un impianto che si autolimita

La cosa a cui tengo di più, in questo saggio, non è tanto la teoria in sé, quanto il modo in cui è costruita: dichiara sempre dove finisce il rigore e dove comincia l'estensione più libera, e non pretende di spiegare tutto. C'è persino una clausola esplicita che dice quando l'agentologia *non* si applica — ai fenomeni fisici privi di una vera retroazione, per esempio, dove parlare di “agenti” sarebbe solo un modo di dire.

Da questo impianto derivano due applicazioni concrete, pensate per essere utili, non solo eleganti.

La prima riguarda le relazioni sociali, ed è il punto di contatto con un saggio precedente, *Psicologia dell'interagire* (https://cancellieri.org/doc/Psicologia_dell_interagire.pdf): buona parte del nostro benessere dipende dalla qualità delle relazioni che costruiamo giorno per giorno, e l'agentologia offre un modo per parlarne — con qualche cautela in più su quanto davvero questo fattore pesi rispetto ad altri.

La seconda riguarda l'uso, per così dire, su noi stessi: i nostri comportamenti nascono spesso da programmi interni — abitudini, impulsi, schemi — che entrano in conflitto tra loro, un po' come se dentro di noi convivessero più “piccoli agenti” con obiettivi diversi. È esattamente ciò che succede a Marco: non gli manca la volontà di consegnare in tempo, gli succede che un'abitudine di evitamento prende il sopravvento sul suo stesso obiettivo dichiarato. Il saggio prova a dare un nome a questo

genere di conflitto, e propone qualche accorgimento per chi voglia usarlo su di sé — ricordando però, con altrettanta chiarezza, dove finisce l'utilità di un esercizio riflessivo e dove comincia il bisogno di un aiuto professionale vero.

Che cosa succede a Marco e Giulia

Il saggio si chiude proprio con il caso di Marco e Giulia, ripreso per intero: si scopre che il punto di svolta non è, in un primo momento, un cambiamento del carattere o delle abitudini di nessuno dei due, ma un semplice cambio di percezione reciproca, ottenuto parlandosi apertamente di ciò che stava succedendo tra loro. Marco non aveva mai davvero considerato, tra le cose che poteva fare, quella di spiegare a Giulia le proprie difficoltà; una volta che l'ha fatto, il suo comportamento è cambiato prima ancora che la sua abitudine di fondo fosse davvero superata.

È un esempio piccolo, ma è anche la miglior risposta che so dare alla domanda più scomoda che qualcuno mi ha posto leggendo una versione precedente del saggio: perché dovrebbe servire un vocabolario nuovo, se già ne abbiamo tanti? La mia scommessa è che un linguaggio comune ed esplicito, capace di descrivere allo stesso modo un conflitto d'ufficio, una relazione di lunga data e persino un agente artificiale, valga la fatica di costruirlo — sapendo fin dall'inizio che si tratta, appunto, di una scommessa, non di una certezza.

Il saggio completo, “Fondamenti di agentologia. Per una teoria generale dell'agire”, è disponibile in https://cancellieri.org/doc/Fondamenti_di_agentologia.pdf.

(Articolo scritto in collaborazione con Claude.ai)