

Bruno Cancellieri, ChatGPT and Claude.ai

Foundations of Agentology

From entities to agents

29 August 2026

Copyright © Bruno Cancellieri

“He’s a procrastinator.” “She’s intrusive.” “That’s just how I am.” These are phrases that seem to explain a behaviour and in fact close the discussion: they attribute a stable property, and from a stable property nothing follows except the possibility of noting it once again.

Agentology changes the questions. Not what a thing is, but who acts, what each one perceives, which goal he pursues, which figures he recognizes between himself and the others, according to what logics he decides, which actions he takes himself to have available. These are questions whose answers might be other than they are, and each of them indicates a point at which something can be changed. The same vocabulary describes a person, a company, a cell and a computer program.

Three things follow from this that are usable even by a reader with no interest in conceptual apparatuses. When one cannot do what one has decided to do there is no defect of the will: there are two parts of the same person pursuing different goals, and knowing how they are made says where to intervene. Emotions are not states to be endured but messages, with a sender, an object and a request: a request can be answered instead of obeyed. And when a relationship goes wrong, almost never do the two recognize the same figures in it: one believes he is giving a service, the other receives a command, and neither of the two is lying.

This essay does not teach how to treat others. It invites one not to stop at the first explanation that comes to mind, which is usually the most convenient and almost always useless.

Contents

Note	5
Preface.....	6
Fundamental concepts.....	8
Agent	8
Consciousness	8
Situation.....	8
Perception.....	8
Program	9
Action	9
Axioms	10
Definitions.....	11
General principle and agentological method	13
Agentological questionnaire.....	13
Emotions	14
The <i>meaning</i> of emotions	14
Emotion as a message.....	14
The force of the message: requests and commands	14
Attraction and repulsion	15
What consciousness can do with the message.....	15
What changes when they are treated in this way.....	16
Relational figures	17
The relational situation and the figures	17
The relational figure as an input to decision	17
The recognition of relational figures	18
Repertoire of relational figures	19
What to expect from the list	19
Summary of the relational figures	19

Relational figures concerning goals	20
Relational figures concerning programs and the knowledge of them	21
Relational figures concerning communication.....	21
Relational figures of asymmetry	22
Relational figures concerning valence and effects	23
Selection	24
The third agent.....	24
Misalignment.....	25
Application to social relationships.....	26
What <i>Psychology of Interacting</i> says	26
What agentology adds	26
Meta-interaction	27
Application to psychotherapy and self-therapy	29
Intra-agent conflict	29
The levels of conflict.....	29
Intra-agent power and akrasia	30
Therapeutic change as directed learning	30
The method applied to oneself	30
The specific problem of self-therapy.....	31
A warning on the use of agentology as a therapeutic instrument.....	31
An example of application	32
Conclusion	34

Note

This essay arose from a dialogue with ChatGPT and Claude.ai, following a method I have also adopted for other works: I propose a topic, discuss its framing and its content with these language models, and have the texts written and then corrected through an articulated exchange in which the changes are proposed, argued for and discussed together.

The specific theoretical framework is developed in these dialogues, but it belongs to a psychological and philosophical background that I built up on my own well before I turned to these tools.

Responsibility for the theses advanced here, and for whatever weaknesses they may have, remains mine.

Bruno Cancellieri

Preface

Of a person who always delivers late we say that he is a procrastinator. Of a company that does not honour its contracts, that it is dishonest. Of a mother who telephones three times a day, that she is intrusive. These are sentences that seem to describe and that in fact close the matter: they ascribe to someone a stable property, and from a stable property nothing follows except the possibility of observing it once again. Whoever utters them stopped enquiring at the very moment he believed he had understood.

The observation is not mine. Alfred Korzybski, in *Science and Sanity* (1933), attributed to a particular use of the verb to be — the one that identifies an individual with a category, “X is a Y” — a considerable share of the confusion with which we speak about the world; a pupil of his, D. David Bourland, went so far as to propose writing without ever employing that verb. Karl Popper, in section 10 of *The Poverty of Historicism*, called methodological essentialism the attitude of those who begin by asking “what is matter?”, “what is justice?”, and methodological nominalism that of those who ask instead “how does this piece of matter behave?”; and he observed that physics and biology made their progress by moving to the second question, while the sciences of man have remained attached to the first.

Agentology takes that shift seriously and carries it where it has not yet fully arrived: to the behaviour of living beings and of their organizations. For the question of what a thing is it substitutes four preliminary questions about what it does: which agents are in play, what each of them perceives, which goal it pursues, which actions it takes to be available. In this it has an illustrious precedent. In *Mind and Nature* (1979) Gregory Bateson does not say what mind is: he lists six conditions a process must satisfy for it to be called mental, and the first is that it be an aggregate of interacting parts whose interactions are triggered by differences and organized in circular chains of determination. A definition made up entirely of what happens, and in which the verb *to be* is not used.

The advantage of the shift from being to acting lies in the way the two descriptions end. The one by identity or property ends in a verdict, and a verdict has no sequel: it names something supposed to be constant, and what is constant offers no handholds. Description by agents cannot end in that way, because each of its four questions has an answer that could be different from the one it is: one can perceive something else, want something else, have available actions one had not seen. Where the verdict delivered a character trait, the analysis delivers four points at which something can be changed; and in the case where none of the four can be modified, it says so, which is in any event more than a verdict grants.

Two things follow. The first is that a description by agents remains open to refutation. “He is aggressive” presents itself as a fact about a person, and admits of no reply; “he perceives my questions as attacks” is a hypothesis about a perception, which can be checked by asking him and corrected if need be. The second is that the same vocabulary cuts across domains that usually remain separate: the same words can describe a cell, a person, a company and a computer program, and what one has learned to see in one of these things one can try to see in the others.

None of this makes agentology the best way of describing anything whatever. For phenomena in which no boundary, no perception and no repertoire of actions can be identified — the erosion of a rock, the fall of a heavy body — description by agents makes no sense and is not applicable; the chapter *General principle and agentological method* says so at length. The thesis is narrower: when the question concerns a behaviour, that is, why someone does what he does and what might make him do otherwise, description by agents is preferable to description by essences, and it is so for the reason just given.

Agentology is a conceptual apparatus of my own devising which sets out to describe and analyse, by means of the concept of agent, the phenomena that involve living beings, their organizations and

machines endowed with a program. To the reader who has no interest in conceptual apparatuses, it offers three things.

The first: when one cannot do what one would like to do, or has decided to do, this is not a defect of the will. Two parts of the same person are pursuing different goals, and that structure indicates where one can intervene — on the situation that activates the conflict, on the signal by which one of the two parts makes itself heard, or on the program that generates that signal. Instead of thinking of a defect of character, one can consider three points on which to reflect.

The second: emotions are not states of consciousness to be endured, but messages, with a sender, an object and a request. A request can be answered instead of obeyed; and of a message one can establish which part is intelligible — what one feels almost always is, why one feels it almost never.

The third: when a relationship goes wrong, almost never do the two recognize the same figures in it, and almost never is each doing what the other believes he is doing; and there is an operation that makes it possible to discover together what each believed himself to be doing: *meta-interaction*.

The centre of this essay is the repertoire of relational figures. The chapters that precede it serve to build it — they say what an agent is, how it perceives, how it decides, and where its goals come from; the one that follows puts it to the test on a case. Without the repertoire the apparatus would remain general and hard to employ; without the apparatus the repertoire would be a list of names.

It must be said at once, however, that the apparatus is general and the repertoire is human: a cell does not lie, does not manipulate and does not feel itself the victim of an injustice. The apparatus serves nonetheless, and is indeed what makes it possible to recognize the same figures where one does not expect to find them — dependence between two companies, command between two programs, selection between two organisms.

The reader will find in this apparatus points of contact with traditions that already exist — general systems theory, cybernetics, the notion of rational agent in artificial intelligence — which have already described organisms, machines and organizations in terms of perception, decision and action. Agentology does not claim a new field of study. With respect to those traditions it does two things: it makes explicit an apparatus that they often share only implicitly and in fragmentary fashion, and it makes it possible to apply that apparatus to domains which at first sight seem to have nothing in common.

Fundamental concepts

The following concepts constitute the theoretical core of agentology. *Agent*, *consciousness*, *situation*, *perception*, *program* and *action* are here taken as primitive terms.

Agent

An agent is an entity capable of perceiving the situation it is in and of performing actions that tend to modify or to preserve it. That it possesses an internal state and at least one program is not part of this formulation: it is established by the axioms.

An internal component deserves the name of sub-agent, and not merely of program, when it displays with a certain persistence a line of conduct of its own, not reducible to that of the other components, and maintains it as situations vary. The criterion is a matter of degree and not always decidable: between a very stable program and a barely autonomous sub-agent the difference is one of degree, and in many cases there is no way of establishing on which side a particular case falls.

An agent may be made up of agents of a lower order, which may be called *sub-agents* with respect to the first. Each sub-agent offers services to its superior agent in terms of information (warnings, requests for intervention and answers to questions). A sub-agent may also address messages to other sub-agents of the same agent.

Consciousness

Consciousness is that by virtue of which certain information, instead of merely producing effects, is *felt*. The essay does not say what consciousness is, nor how it is produced: it assumes it as primitive, on a par with agent and with situation, and makes use of a single property of it — that what reaches it is felt. Its use is constrained by axiom A1, which attributes to every human being a single agent endowed with consciousness and an indeterminate number of agents that lack it.

An agent is said to be conscious if it is endowed with consciousness, unconscious if it is not. In this essay the two terms are always used in this sense, and never in the weaker sense of “being aware of something”: an unconscious agent may possess extremely accurate information about its own situation, and act on the basis of it with great effectiveness.

Nothing that follows depends on knowing what consciousness is. It depends only on the fact that, in a human being, some information is felt and some is not: which is an observation, not a theory. On this single fact rests the definition of emotion given in the chapter Emotions.

Situation

A *situation* is the configuration, at a given instant, of the external environment and of the internal states of the agents involved. Since the internal state comprises information, memories and programs, the situation also includes the relationships among the agents, which have no existence of their own but consist in what past interactions have deposited in the agents themselves. Interactions, on the other hand, are not included: a situation is instantaneous, whereas an interaction is a sequence of actions, and is what leads from one situation to the next.

Perception

Perception is the process by which an agent acquires information about the situation and messages from other agents. Perceptions modify the agent's internal state.

Program

A *program* is an internal logical structure, partially stable over time, which constrains the set of actions an agent can produce from a given perception: it excludes some actions and makes others more probable or necessary.

This definition has a precise consequence: an entity whose actions were wholly independent of any internal structure — purely random with respect to what it perceives — would not possess a program and, by axiom A4 (see the following section), would not be an agent.

Typical examples of programs are genetic mechanisms, reflexes, instincts, the mechanisms that generate emotions, habits, items of knowledge, strategies, reasonings, algorithms, logics, and so on.

Action

An *action* is any gesture, movement or expression — or sequence of these — on the part of an agent, aimed at modifying a situation or at preserving it. Every action produces one or more effects, intentional or unintentional: the distinction between the two kinds of effect is taken up in the definition of Responsibility, in the chapter *Definitions*.

Actions do not remain isolated. When the actions of two agents answer one another, reciprocally modifying their internal states, we have an ***interaction***. When interactions between the same agents are repeated over time in recognizable forms, what they deposit in those agents' internal states is a ***relationship***. The three terms lie on a scale of duration: the single act, the exchange in which the act is set, the sediment that repeated exchanges leave behind.

Since a relationship is a deposit in internal states, and internal states are part of the situation, the relationships built by past actions and interactions are part of the situation in which subsequent actions are performed.

Axioms

The statements that follow are all that the essay assumes. None of them says what an agent wants, nor why it pursues one goal rather than another: such questions are dealt with in the chapter Emotions.

A1. There are agents: for example living organisms, their organs, cells, organizations and machines endowed with a program. Every human being comprises one conscious agent — his consciousness — and an indeterminate number of unconscious agents.

Note. That every human being comprises a single conscious agent and an indeterminate number of unconscious agents is a thesis about the human being, not falsifiable as things stand, for which I take responsibility. I place it among the axioms because the chapter Emotions rests entirely upon it.

A2. Every agent is always in a situation.

A3. Every agent perceives at least a part of the situation it is in.

A4. Every agent possesses at least one program.

A5. The behaviour of an agent at a given instant depends on the perceived situation and on its internal state.

A6. The programs of an agent may change over time (plastic programs) or remain invariant (rigid programs).

A7. An agent can, by means of its own actions, modify the situation, other agents or its own internal state, and in particular its programs.

A8. A set of agents constitutes an agent of higher order when, at the level of the set, there emerge perceptions and decisions not reducible to the mere sum of the perceptions and decisions of the individual components — for example a coordinated decision procedure, a boundary recognizable with respect to the outside, a persistence over time distinct from that of the individual components. In that case the agents may be considered *sub-agents* with respect to the higher-order agent. In the absence of these conditions, the set remains a plurality of agents, not an agent.

A9. An agent may belong simultaneously to several higher-order agents.

Definitions

Information — Any difference detected and acquired through a perception and recorded in the internal state of an agent, which may give rise to differences of behaviour or may be emitted into the environment towards other agents.

Internal state — The set of the information, the programs and the other internal characteristics possessed by an agent at a given instant.

Decision — The operation by which a program selects an action from among those that the perceived situation makes available (cf. *Freedom*). The program is the structure, the decision is the operation that this structure performs: this is why they appear as two distinct moments in the fundamental cycle. Decisions are taken according to logics which in their most elementary form consist of the instruction “if X, then Y”, where X is a recognized situation and Y is the action to be carried out.

Goal — A situation that an agent tends to bring about, to maintain or to avoid.

Memory — The part of the internal state that preserves information over time.

Learning — Modification of one or more programs as a consequence of experience; it is the case, provided for by A6, of a plastic program that changes. It is not an exclusively human phenomenon, nor an exclusively psychological one: a memory lymphocyte modifies its own program as a consequence of an exposure and preserves the modification for years.

Interaction — A sequence of actions by which two or more agents reciprocally modify their respective internal states.

Communication — An interaction in which at least one action has as its specific goal, and not as a mere side effect, the modification of the internal state of another agent by means of a signal governed by a code at least partly shared among the agents involved. This clause distinguishes communication from interactions in general: not every interaction that modifies the internal state of another is communication, but only that in which such modification is pursued as a goal by means of a code.

Command — A communication whose goal is that the receiver perform an action aimed at obtaining a certain situation.

Cooperation — An interaction in which two or more agents pursue at least one common goal and their respective actions are reciprocally functional to its attainment.

Competition — An interaction in which the attainment of one agent's goal reduces the probability that another agent will attain its own.

Note. Cooperation and competition are not mutually exclusive categories: one and the same interaction may be cooperative with respect to one goal and competitive with respect to another (mixed interaction) — this is the ordinary case of relations between organizations that collaborate on a project and compete in a market.

Power — The capacity of an agent to modify in a significant and lasting way the state or the behaviour of another agent, even against the latter's preferences or overcoming whatever resistance it may offer. This clause distinguishes power from the generic capacity to act: not every action that modifies another agent is an exercise of power, but only one that prevails over a contrary preference or resistance.

Freedom — The set of the actions that an agent takes to be available in the situation it is in and that it is able to perform with the programs at its disposal. Freedom so defined is relative to the judgement of

the agent: an action objectively possible but not taken to be such does not fall, for that agent, within its freedom. The divergence between the actions taken to be available and those objectively available to an external observer is itself an agentologically relevant phenomenon — for example in cases of ignorance or of self-deception.

Responsibility — The attribution to an agent of the effects of one of its own actions, graduated according to two factors: the degree of intentionality of the effect (cf. Action) and the range of the alternatives concretely available to it at the moment of decision. A fully intentional effect, produced in the presence of available alternatives, grounds an attribution of full responsibility; the absence of alternatives or of intentionality attenuates or excludes it. The attribution may have as its object another agent or the agent itself. It is the operation on which moral judgements and practical evaluations are founded: one reacts differently to the same harm according to whether one judges it intentional or accidental.

Note. Here “available” does not have the same sense as in the definition of Freedom: there it was a matter of the actions that the agent took to be such, here of those that an observer recognizes as possible for him. That the agent did not see them may attenuate the attribution, but is not enough to exclude it.

Relational situation — The part of the situation that concerns two determinate agents: the configuration of their goals, of their reciprocally relevant programs, of the asymmetry between them and of the effects each produces on the other, at a given instant. It is instantaneous like every situation, and it obtains also between agents meeting for the first time; what successive relational situations deposit in the internal states is the relationship.

Relationship — A relatively stable bond between two or more agents, constituted by the sedimentation over time of a multiplicity of interactions between them, such as to orient subsequent interactions in a recognizable way. A relationship is constituted by the interactive habits — shared or reciprocally represented — deposited in the internal states of the agents involved. In this sense, a relationship might be regarded as the “pattern” (that is, the set of the characteristic and repetitive forms) of the habitual interactions between two agents. The forms that recur in it are relational figures, defined in the following entry. It is to be distinguished from the relational situation, of which it is the sediment.

Relational figure — A recognizable configuration that two agents form together, in a given respect: the goals of the two, their programs, the asymmetry between them, the affective valence, the effects each produces on the other. Cooperation, dependence, command and obedience are relational figures. Since the respects are many, one and the same relational situation presents several of them at once — two colleagues may cooperate, depend on each other and not trust each other; and since each recognizes it on his own account, one and the same figure gives rise to two readings, one per agent, which may diverge. It is the recognized figures, and not the relational situation, that enter into decisions.

Note. A figure is recognized in a single action, in an interaction or in a relationship; the figures that recur in a relationship are the characteristic forms of which its pattern is made. In this essay relational figure is written in the definitions, in the titles and at the first occurrence in each section, and figure where the context is clear.

Conflict — A situation in which two or more agents, or two or more programs of the same agent, tend towards actions or goals that are mutually incompatible or antagonistic. Competition is the case in which the terms of the conflict are distinct agents; intra-agent conflict is the case in which they are components of the same agent.

General principle and agentological method

Any phenomenon that exhibits a non-random configuration of perception, decision and action can be described as the result of the actions of one or more agents which, on the basis of their own perceptions and of their own internal state — which includes programs and memory — modify or maintain, according to certain logics, the situations in which they are immersed.

This principle does not apply to processes devoid of feedback, in which nothing of what happens depends on what anyone perceives — for example the erosion of a rock by the wind. In such cases agentology has nothing to add to the ordinary physical description, and does not claim to replace it.

The **agentological method** consists in answering the questions contained in the following questionnaire, and others more detailed that can be derived from them.

Agentological questionnaire

- Which agents are involved?
- What is the current situation?
- What information does each agent perceive?
- Which figures does each agent recognize between himself and each of the others?
- Which programs are relevant?
- Which goals does each agent pursue?
- Which actions does it perform?
- By what criteria and according to what logics does each agent decide its own actions?
- In what ways do these actions modify the situation?
- Which new situations may be generated?

The questions follow the order of the fundamental cycle of agentology:

Situation → Perception → Program → Decision → Action → New situation

The last question leads back to the first: the situation produced by the actions is the one from which the analysis begins again. The method is therefore not a list of answers, but a circuit to be travelled as many times as there are situations that follow one another.

Emotions

The axiomatic core defined above attributes goals to every agent, but does not say what determines a goal, nor why an agent pursues one rather than another. Axiom A1, moreover, states that every human being comprises one conscious agent — his consciousness — and an indeterminate number of unconscious agents, but so far nothing has been said about how these communicate with one another. The two questions have, in my view, the same answer, and this is why the chapter addresses them together: the goals that consciousness pursues are largely proposed to it, or imposed on it, by the unconscious agents, and the means at their disposal for doing so are the emotions. Emotions are, in this sense, the principal mechanism by which a human being has the goals he has and not others.

The *meaning* of emotions

I here call *emotion* only that which is felt. An affective process that does not reach consciousness — a physiological activation, an automatic appraisal, a preference that orients behaviour without the subject noticing it — is not, in this vocabulary, an emotion: it is a program, or the working of an unconscious agent. This is a terminological choice, not an empirical thesis. It does not deny the existence of the affective processes that do not reach consciousness, which are on the contrary the presupposition of the whole chapter; it merely reserves the name of emotion for the signal that reaches consciousness from those processes.

Emotion as a message

Emotion — Information that an unconscious agent transmits to consciousness, and that is felt by the latter in terms of pleasure or pain, attraction or repulsion. It has three attributes: an object (a situation or an event present, remembered or imagined), a positive or negative valence towards that object, and an intensity.

An emotion is therefore a communication, that is, an action whose specific goal, and not mere side effect, is to modify the internal state of the agent “consciousness”. It is, however, a communication whose code is only partly shared. Valence and intensity reach consciousness with relative clarity: one feels that one is attracted or repelled, and one feels how strongly. The object and the reason may arrive distorted, not clearly recognized, or not arrive at all. This is the ordinary experience of feeling something without knowing why, and it is the predictable consequence of the fact that the sender is not conscious and is not able to construct and send to consciousness clearer and more eloquent messages.

Reconstructing the complete message — which agent emits it, about which object, with a view to which goal — is an operation of abduction, the same that in the chapter *Application to psychotherapy and self-therapy* is equated with *reverse engineering*. Its fallibility is not an accident but a structural feature, and it must be borne in mind whenever anyone — the subject himself or an observer — claims to know why a person feels what he feels.

The force of the message: requests and commands

An emotion is not neutral information. It is accompanied by a promise: of pleasure, if consciousness goes along with the attraction or avoids what repels it; of pain, otherwise. It is the promise that gives the message its force, and that turns information into a solicitation to act.

Depending on its intensity, an emotion functions as a request or as an imperative command, in the sense in which command was defined in the chapter *Definitions*. At low intensity it leaves consciousness the time and the room to deliberate, to postpone, to decide otherwise: it enters into deliberation as one of its elements, and in this sense it is a request. At high intensity deliberation is ruled out, or else it takes place but its outcome is irrelevant because the action is already under way: it

is a command, and it is the case in which the unconscious agent exercises power over consciousness. Between the two extremes there is no boundary but a continuum.

This distinction supplies the description that *akrasia* — the lack of self-mastery — otherwise lacks. To describe it, as will be done in the chapter *Application to psychotherapy and self-therapy*, as the prevailing of one program over another is correct, but incomplete: programs, taken in themselves, have no preferences to assert nor means of asserting them. What prevails is a program that has at its disposal a channel for reaching consciousness and a promise with which to press upon it.

The intensity of an emotion does not always have the same effect. The same intensity is resisted differently by different people, and by the same person at different moments — depending on tiredness, on habituation to that particular signal, on the presence of a contrary signal. Intensity is a necessary factor in accounting for the force of an emotion, not a sufficient one: how much a given intensity manages to obtain depends also on the internal state of the one who receives it.

Attraction and repulsion

Every emotion involves an attraction towards, or a repulsion from, its own object. In the vocabulary of the essay: it proposes a goal to consciousness — to bring about or maintain the situation of which the object forms part, in the case of attraction; to avoid it or interrupt it, in the case of repulsion.

Mixed emotions — those in which the same object attracts and repels at once — are no exception to this thesis: they are a consequence of it. They are two messages, emitted by two different unconscious agents, which reach consciousness simultaneously with opposite valences; they are, that is, a case of intra-agent conflict, as described in the following chapter. They have, however, a peculiarity that makes them precious: they are the only kind of internal conflict both of whose terms present themselves to consciousness. The other intra-agent conflicts have to be reconstructed by abduction; ambivalence is felt. This is probably the reason why it is advisable for self-analysis to begin there.

There remain cases that this description does not cover. *Surprise* does not seem to have a valence of its own: it precedes appraisal rather than expressing it, and it resembles a warning — the information by which a sub-agent signals that something has changed, without thereby requesting any determinate action — more than an emotion in the sense defined here. Diffuse moods, which seem to be without an object, admit of two readings: either the object is there but is so general as not to be identifiable, or they are not emotions but the background noise produced by the overall internal state. I leave the question open.

What consciousness can do with the message

If the relation were one-way — commands that arrive, execution that follows — there would be nothing to say about change, and the chapter *Psychotherapeutic and self-therapeutic application* would be left without an object. The actions available to consciousness are at least these: to execute; to postpone execution, which is already a form of resistance; to act not on the requested action but on the situation that generates the signal, withdrawing from the contexts that activate it; to interrogate the signal, that is, to attempt the reconstruction of the complete message; and finally, to modify the program that emits it, which is the slowest way and the only lasting one.

Each of these actions is in turn perceived by the unconscious agents, which recalibrate their own signals accordingly. The relation is therefore a cycle, not a chain of command: it is the same structure Situation → Perception → Program → Decision → Action → New situation, applied within a single person.

One consequence should be pointed out, because it touches on a point in the chapter *Fundamental concepts*, where it was said that every sub-agent offers services to its own superior agent in the form of information. If the unconscious agents command consciousness, consciousness is not their superior

agent, but a peer of theirs: the superior agent is the person as a whole, of which both consciousness and the unconscious agents are sub-agents, and emotions are horizontal communications between sub-agents, not services rendered upwards. Consciousness is not the summit of the person: it is the person's sub-agent specialized in deliberation and in long-range action, and as such it is the addressee of the messages of the others.

What changes when they are treated in this way

The difference between a state and a message is not verbal. Of a state one can only wait for it to change, or try to suppress it: these are the two things commonly done with fear, with anger, with inopportune desire. To a message, on the other hand, one can reply, and the available replies are those listed in the preceding section. None of them is available to anyone who regards his own emotion as a condition he is in.

Two things follow from this. The first is that one can distinguish the reliable part of the signal from the part that is not. What one feels — attraction or repulsion, and with how much force — almost always arrives correct: it is the information the sender knows how to give. What one attributes it to often arrives wrong. “I am anxious” is almost always true; “I am anxious about tomorrow's meeting” is a hypothesis, and asking oneself whether it is the right one is an operation that makes sense only for someone who has admitted that the emotion has an object, and that the object might be another.

The second is that ambivalence stops being a defect. Those who treat emotions as states regard with embarrassment not knowing whether they want a thing or do not want it, and try to resolve the matter by deciding, that is, by silencing one of the two sides. Those who treat them as messages read in them a piece of information: two different senders, with opposite requests, each of them protecting something. The useful question is no longer which of the two is right, but what each is trying to protect — and this, unlike the other, is a question one sometimes knows how to answer.

Relational figures

The chapter *Definitions* has distinguished three things that ordinary usage confuses: the relational situation, which is the instantaneous configuration between two agents; the relationship, which is what successive relational situations deposit in their internal states; and the relational figure, which is a recognizable configuration that the two form together in a given respect. This chapter lists the figures that seem to me most relevant to decisions, and shows what follows from the fact that each of the two agents recognizes them on his own account.

The relational situation and the figures

Of one and the same pair of agents different descriptions can be given. One is that of a third observer, who compares the goals of the two and finds, for example, that the attainment of the one reduces the probability that the other will attain its own: this is competition as the chapter *Definitions* defines it, and it is a configuration of the pair, not a property of either of the two. The others are the relational figures that the two agents recognize in it: how each of them reads that configuration. The three descriptions may not coincide, and as a rule they do not coincide entirely; nor is even the first a description made from nowhere, since the third observer too is an agent who perceives, better informed but no less situated.

This is not a difficulty peculiar to figures. It is the application to the relational situation of what the essay has already established for the situation and for freedom. Axiom A5 says that the behaviour of an agent depends on the perceived situation, and the definition of perception explains why: to the situation no agent has access, it has access to the information that its perception delivers to it. The relational situation is a part of the situation, and is no exception: it enters behaviour only through the figures recognized in it. The divergence between the relational situation and the recognized figure is the same as the one that, in the definition of freedom, separates the objectively available actions from those that the agent takes to be available.

From this comes a consequence worth stating separately: the two readings of one and the same figure may be asymmetrical, and their asymmetry is not a pathological case but the ordinary condition. One believes he is giving a service, the other receives a command. One cooperates, the other feels exploited. Neither of the two is lying.

The relational figure as an input to decision

To recognize a relational figure is not a conclusion, it is an entry. In the fundamental cycle — Situation → Perception → Program → Decision → Action → New situation — the recognition of the figure does not add a phase: it is part of what perception delivers to the programs. The rules of the form «if X, then Y» that the chapter *Definitions* attributes to decision often have for X not a physical situation but a figure: if this man is a competitor, then I do not tell him what I know; if he has authority over me, then I comply; if I depend on him, then I do not protest.

There follows a property that distinguishes figures from other perceptions, and it is the chief reason why they deserve a chapter. The recognized figure tends to make itself true. He who recognizes that he is in competition acts accordingly, the other perceives those actions and responds competitively in turn, and the relational situation ends by resembling the figure that preceded it. It is the same cycle as before, gone round once more; but its effect is that an error of recognition does not remain an error of description: it becomes an action that modifies what it was describing. In the chapter *An example of application* this is what happens between Marco and Giulia: the intensification of controls, decided on the basis of a figure in which Giulia reads disengagement, produces the disengagement it presumed.

This is also the answer to the objection the chapter invites. To say «ours is a relation of competition» looks like the same move that the *Preface* declares sterile: to attribute a stable property and close the

inquiry. The difference lies in what one supposes oneself to have recognized. A figure is not a property of a person: it is a hypothesis about a configuration between two, which can be verified by asking and if need be corrected; and unlike a character it has a date of beginning, a history, and at least two readings. A verdict is sterile when it believes itself unique and definitive, not by virtue of being pronounced.

The recognition of relational figures

To perceive a relational figure and to recognize it are not the same operation. Perception, as the chapter *Fundamental concepts* defines it, is the process by which an agent acquires information about the situation; recognition is the comparison of that information with a repertoire the agent already possesses. That repertoire is made of programs, in the sense of the corresponding entry, which lists among its examples knowledge, strategies and logics; and since axiom A6 distinguishes rigid programs from plastic ones, some figures are recognized without ever having been learned and others only through having been met.

A figure is recognized at three levels, and not with the same reliability: in a single action — a tone, a correction, a silence; in an interaction, that is in an exchange as it unfolds; and in a relationship, where what is recognized is what recurs. To attribute a figure to a relationship on the evidence of a single act is the ordinary way of getting it wrong, and it is the same leap that the *Preface* reproaches in one who, from a few late deliveries, concludes that a person is a procrastinator. The scale on which one recognizes is therefore, alongside those already named, a source of error.

There follows a second reason why two agents may diverge, independent of the first. So far it has been said that each has access to his own intentions and to the effects he undergoes, and not to those of the other. But the two may also have different repertoires: a figure that one recognizes may be altogether missing in the other, who does not read it badly but does not read it at all. It is the case in which a meta-interaction is not enough, because it is not a matter of correcting an attribution but of showing a figure the other has never learned to see.

Finally, figures are not only recognized: they are pursued. Since a figure is part of a situation, it can be a goal in the sense of the corresponding entry — a situation that an agent tends to bring about, to maintain or to avoid. This is what makes mechanical the cycle stated in the preceding section: he who recognizes a figure of competition sets himself the goal of bringing it about or of winning it, acts accordingly, and the other recognizes it in turn. Figures are thus at once an input and an output of the decision process: they are recognized in the situation and pursued in the action.

Repertoire of relational figures

The relational figures that follow are twenty-two. They are not a classification, and they claim to exhaust nothing: they are a repertoire, that is the list of what, between two agents, it is worth being able to recognize. The preceding chapter said what a figure is and how it is recognized; this one says which.

What to expect from the list

The entries that follow are not mutually exclusive. A pair of agents almost always recognizes several of these relational figures at once, and others successively: two colleagues may cooperate on a project, compete for a promotion, depend on each other asymmetrically and not trust each other, all in the same month. Some entries are independent of one another, others are not, and where the link exists it is indicated: dependence is the basis of power, sharing is a requirement of cooperation, attraction orients selection without determining it.

The list is open. It is not an exhaustive classification and does not derive from a principle that would guarantee its completeness: it gathers the figures that seem to me most frequently relevant to decisions, and others may be added to it provided they pass the criterion stated in the chapter *An example of application* — that a distinction is worth what it costs when something changes, either the action one chooses or what one expects to happen.

Finally, not all figures are applicable to all kinds of agent. Each requires of the agent that recognizes it certain capacities, and where these are lacking the figure is not rare: it is devoid of sense. To attribute an offence requires being able to attribute intentions; to trust requires representing to oneself the future behaviour of another; to imitate requires plastic programs. A bacterium competes and cooperates, it does not take offence; a thermostat does neither the one nor the other. Where the restriction matters, the entries indicate it.

Summary of the relational figures

- Cooperation, competition, indifference
- Hostility
- Sharing, non-sharing
- Commitment
- Imitation, differentiation
- Knowledge of the other
- Sincerity, lying
- Truth, falsity
- Manipulation
- Confrontation and avoidance
- Dependence, independence
- Power exercised, power undergone
- Command, obedience
- Recognized superiority and inferiority
- Institutional superiority and inferiority
- Giving and receiving service
- Debt and gratitude
- Attraction, repulsion
- Trust, distrust
- Harm, offence
- Justice, injustice
- Choosing and discarding, being chosen and being discarded

In the sections that follow the figures are grouped by category, and defined.

Relational figures concerning goals

Cooperation, competition, indifference — The first two are defined in the chapter *Definitions*, together with the observation that they are not mutually exclusive: one and the same interaction may be cooperative with respect to one goal and competitive with respect to another. Here the third case must be added, which without it the list would leave uncovered: indifference, in which the goals of the two do not interfere at all. It is the commonest figure between agents of any kind, and it is the one first lost sight of, because he who feels himself in competition tends to attribute to the other an interest in him that the other does not have. The decision that changes is whether it is worth acting on the other: from an indifferent agent there is nothing to be obtained, and to believe the contrary is one of the costliest errors.

Hostility — A figure in which harming the other is not an effect of the actions but their goal. It is distinguished from competition by the structure of the goal and not by its intensity: in competition the other's loss is the consequence of one's own gain and ceases once the gain is obtained; in hostility it is the diminution of the other that is pursued, and it can continue when nothing more is to be got from it. War is its institutional and organized case, but the figure occurs between two persons and between two organizations that fight with no weapons. It must also be kept distinct from *Harm, offence*, which concerns an effect already produced and its attribution: hostility is a goal, and can exist before any harm has been done. The typical misalignment is between one who feels himself in competition and one who feels himself attacked, and by the property stated above it tends to resolve itself in the second sense: he who is treated as an enemy behaves as one. The decision that changes is whether the conflict will end when the disputed object is assigned — in competition yes, in hostility no.

Sharing, non-sharing — Two agents may share information, resources, criteria of judgement, values, or membership of a third agent. Membership must however be distinguished according to its object: of two members of the same organization a certain behaviour is expected because a common decision procedure binds them; of two members of the same category — which is not an agent and decides nothing — something is expected only because they have been placed in a class, and this second expectation almost always has a weaker foundation. Sharing is a requirement of cooperation — one does not cooperate on a goal that one of the two is unaware of, nor with incompatible criteria of success — but it is not sufficient to produce it. And it may produce the opposite: when what is shared is the desire for an object that admits of no two possessors, sharing generates competition, and the more bitterly the more the two resemble each other. This is the thesis René Girard developed in *Deceit, Desire and the Novel* (1961): desire is born not from the object but from the imitation of another who desires it, and the model thereby becomes a rival. The decision that changes concerns what to put in common: increasing sharing improves cooperation where the objects are divisible, and worsens it where they are not.

Commitment — A figure in which an agent binds its own future behaviour towards another: the promise, the contract, the given word. In the terms of the essay it is an agent that voluntarily restricts its own freedom — the set of the actions it takes to be available — in exchange for the other's predictability; and when the commitment is reciprocal, both buy predictability with freedom. It must be kept distinct from trust, which is the expectation of the one who receives and not the act of the one who binds himself: one commits oneself also to those one does not trust. The misalignment is daily and costly, because commitments have degrees that words do not distinguish: one believes he has said «in principle», the other has already counted on him and given up alternatives. The decision that changes is whether it is still legitimate to reconsider, and on what terms: before an intention one changes one's mind, before a commitment one negotiates or defaults.

Relational figures concerning programs and the knowledge of them

Imitation, differentiation — An agent may take as its own the programs of another, or deliberately move away from them. Imitation requires plastic programs, in the sense of axiom A6, and presupposes a judgement: one imitates whom one judges successful, and this is why imitating is at once a technical act and a declaration of rank. Differentiation has two motives that it is well not to confuse: avoiding overlap, when resemblance produces competition for the same resources, and asserting an identity, when what is wanted is not to be confused with another. The two sides often see the same thing in opposite ways: he who imitates intends it as recognition, he who is imitated may read it as appropriation. The decision that changes is whether to approach or to move away from the programs of the other, and the answer depends on which of the two motives is in play.

Knowledge of the other — It is the degree to which an agent has at its disposal a model of the programs, the perceptions and the goals of another. It is not binary and it is not symmetrical: one may know very well an agent by whom one is known very little, and this is one of the most stable forms of advantage. This relational figure requires no resemblance and no sympathy — an adversary is modelled precisely in order not to do what he would do — and must be kept distinct from imitation, which involves adopting the programs of the other and not merely representing them to oneself. It is of models of this kind that the internalized sub-agents are made of which the chapters *Application to social relationships* and *Application to psychotherapy and self-therapy* speak, and it is on them that meta-interaction operates. The decision that changes is how far to trust one's own predictions about the behaviour of the other.

Relational figures concerning communication

Sincerity, lying — A figure in which what an agent transmits corresponds or not to what he himself believes. It concerns the sender and not the content: a sincere agent may transmit the false, if it is he who is in error or who deceives himself — the self-deception that the chapter *Definitions* names in connection with freedom and to which the chapter *Application to psychotherapy and self-therapy* returns. It includes silence: one lies by keeping back what one knows to be relevant to the other, and omission is the commonest form and the least recognizable. It must be kept distinct from distrust, which is the expectation of the one who receives and not the act of the one who transmits: one may trust a liar and distrust a sincere man. It is one of only two figures in the list in which the divergence between the two readings is not the error of either but the result intended by one. The decision that changes is whether it is better to question the other or to verify on one's own account: with a liar questions are of no use.

Truth, falsity — A figure in which what passes between the two corresponds or not to how things stand. It is independent of the preceding one, and all four combinations occur, including that of the liar who happens to say the truth. It is the only figure in the list whose two readings may coincide and both be wrong: two agents perfectly sincere with each other may share a false representation, and their relationship will suffer from it without either having anything to reproach himself with. The decision that changes is where to look when something does not add up: in the other's sincerity, or in the facts. It is a distinction that concerns the instrument too, because meta-interaction serves the first and not the second: it reconstructs what each believed, not what was true.

Manipulation — A figure in which an agent obtains a behaviour from another by acting on his decision process instead of on his reasons, and in such a way that the other believes he has chosen. It does not require lying: one manipulates also with true statements, by choosing what to say, when to say it and in what order, and by exploiting what one knows of the other's programs. It is distinguished from command, which is declared — he who receives it knows that something has been asked of him — and from power, which prevails over a resistance in broad daylight: manipulation is power that hides inside the other's choice. In the terms of the preceding chapter it operates by producing in the one who undergoes it the recognition of a figure that suits the one who exercises it. The misalignment

is maximal and deferred: one recognizes a persuasion or a help, the other recognizes a manipulation only afterwards, and often never. The decision that changes is whether the other is to be answered or withdrawn from: to an argument one replies, to a manipulation one does not.

Confrontation and avoidance — A figure in which the two can, or cannot, speak of what happens between them. It does not depend on the other figures: confrontation is avoided in cooperative relationships and undertaken in hostile ones. It is the figure that opens or closes access to the instrument this essay proposes, since a meta-interaction is possible only where confrontation is; and what makes it impossible is almost always a prediction — that speaking of it will make things worse — whose truth is never put to the test, because avoiding confrontation also prevents discovering that it would have gone otherwise. The typical misalignment is that one of the two considers the question settled and the other considers that it is never spoken of. The decision that changes is whether it is worth raising a question or better to act by other routes.

Relational figures of asymmetry

Dependence, independence — An agent depends on another to the extent that the loss of the relationship costs it: because it receives from it resources, information or services that it does not find elsewhere, or because it has no alternatives. Dependence is rarely equal, and it is the foundation of power rather than a consequence of it: one can prevail over the preferences of the other to the extent that breaking off the relationship costs the other more. It applies to any agent that receives something from another, and describes a supplier and a client as it does a cell and the tissue that nourishes it. The decision that changes is how far one can insist, and whether it is worth procuring an alternative: he who reduces his own dependence modifies the figure of power without having acted on it directly.

Power exercised, power undergone — Power is defined in the chapter *Definitions* as the capacity to modify in a significant and lasting way the state or the behaviour of another agent, even against its preferences. As a relational figure it must be recognized from two sides, and the two sides rarely agree: he who exercises it hardly ever notices it in the right measure, because he sees his own actions as normal and has no access to the effort it costs to undergo them; he who undergoes it tends to attribute to it a more precise intention than it has. The decision that changes is where to look for the cause of one's own behaviour when it cannot be explained: in one's own will, or in the margin that someone else has narrowed.

Command, obedience — Command is defined in the chapter *Definitions* as the communication whose goal is that the receiver perform an action. There is a figure of command and obedience when performance is expected in a stable way, and it is here that authority must be introduced, which the notion of power does not cover: authority is the right to command, granted by him who obeys or established by a third party, whereas power prevails over a resistance without need of any right. The difference is seen when authority falls away and power remains, or the reverse. The decision that changes is whether, faced with a command, the question is to obey or to negotiate: against power one can only resist or withdraw, whereas authority can be contested by invoking the very source that instituted it.

Recognized superiority and inferiority — It is the comparative judgement each of the two makes of the worth of the other, and it makes sense only if one specifies with respect to what: competence in a determinate field, merit, capacity to obtain results. Recognized means that it is a judgement and not a fact: the two may make opposite ones, and may make different ones field by field. Between organizations the judgement takes the name of reputation and has the same effect: it determines from whom standards and procedures are accepted. The decision that changes is whom to ask for a correction and from whom to accept one.

Institutional superiority and inferiority — It is the position established by a third party — an organization, a community, a norm — independently of the judgement of the two. It is distinguished from the preceding one because it does not need to be recognized in order to operate, and the

divergence between the two is a typical source of conflict: the ordinary case is that of one who is superior in grade and inferior in competence in the eyes of his subordinate. The decision that changes is by what route something can be obtained: towards an institutional superiority one goes by the procedure, towards a recognized superiority one goes by the argument.

Giving and receiving service — An agent performs for another an action that the other would not perform by itself, or not as well. It is the figure already described in the chapter *Fundamental concepts* between a sub-agent and its superior agent, but it occurs also between equals and between agents of any kind: an organ for the organism, a supplier for a client, a program for its user. The typical misalignment is that the service given is received as a command or as control, and the reason is that every service involves deciding, in another's place, some part of what concerns him. The decision that changes is whether an offer of help should be accepted or refused, and the useful question is not whether it is well intentioned — usually it is — but how much freedom it takes from the one who receives it.

Debt and gratitude — A figure in which what an agent has received from another leaves an obligation. It stands to *Giving and receiving service* as the relationship stands to the interaction: the act and its sediment. It is asymmetrical in a characteristic way: he who has given remembers longer and values higher than he who has received, and sometimes he who receives finds himself with a debt that the giver did not mean to create — it is the ordinary way in which a gift becomes a burden. The decision that changes is whether something is owed, and therefore whether a subsequent request is a request or a collection.

Relational figures concerning valence and effects

Attraction, repulsion — The chapter *Emotions* defines them as the two valences with which an emotion reaches consciousness. When the object of the emotion is another agent, they constitute a relational figure, and it is the only one in the list that does not have to be reconstructed: it is felt. It requires, however, an agent endowed with consciousness, and to others it does not apply. Like every emotion it may be mixed, and the co-presence of attraction and repulsion towards the same person is an intra-agent conflict, not an uncertainty to be resolved. The decision that changes is how much weight to give one's own affective orientation in the choices concerning that agent — which is a decision and not a datum: one chooses to work with those one does not like, and avoids those one does.

Trust, distrust — It is the reliability attributed to the future behaviour of the other: how far one expects him to do what he has said, and not to take advantage of what has been granted him. It must be kept distinct from attraction, with which it is easily confused: one trusts people one does not like, and does not trust people one likes very much. It presupposes some degree of knowledge of the other, and it is asymmetrical by construction, because each has of himself an evidence he does not have of the other. The decision that changes is how far to expose oneself: what to entrust, how much to verify, what guarantees to ask. Between organizations it is the entry that decides whether a contract must provide for everything or may leave something open.

Harm, offence — Harm is an unfavourable effect produced in an agent by the actions of another. An offence is a harm to which the one who undergoes it attributes an intention. The distinction is the one the chapter *Definitions* fixes in connection with responsibility, and it is here that misalignment reaches its maximum: he who has produced the harm often has not noticed it, and almost always judges it unintentional, because he has access to his own intentions; he who undergoes it has access only to the effect, and for him the attribution of an intention is a hypothesis that nothing in immediate experience contradicts. The decision that changes is what reparation to ask and what to offer: to a harm one responds by repairing it, to an offence one does not, and this is why so many material reparations settle nothing. What a harm or an offence deposits over time is resentment, and its deliberate interruption is forgiveness: they are not separate figures but the sediment of this one, and the decision they change is whether the account is closed — as long as it stays open, every subsequent interaction is added to it.

Justice, injustice — A figure in which each of the two measures what happens between them against a criterion he supposes common: the proportion between what each gives and receives, the correspondence between what one has done and what one obtains, or the equal application of a rule to both — this last is the figure recognized when the other forgives himself what he reproaches in us. It presupposes the sharing of criteria of judgement, in the sense of the entry *Sharing*: without a supposedly common criterion there is no injustice but only disadvantage. It must be kept distinct from *Harm, offence*: one may suffer an injustice without any harm having been meant, and even without any harm having been done, as in the unequal division of an advantage. Misalignment is here the rule and not the exception, because each applies the criterion from his own position and remembers his own contributions better than the other's: two agents may both feel wronged, sincerely and at the same time. The decision that changes is the largest in the list, because the sense of injustice turns a disadvantage into a claim: he who judges himself at a loss considers whether it is worth changing, he who judges himself wronged considers whether to demand, and authorizes himself actions he would otherwise forbid himself. The essay does not say what is just: it describes the figure, that is the fact that each measures against a criterion and that the criteria may diverge.

Selection

Choosing and discarding, being chosen and being discarded — An agent seeks, maintains, loosens or breaks off a relationship, and changes its choice over time. The positions are four and must be kept distinct, because each changes different things: choosing, discarding, being chosen, being discarded. It is a decision more than a relational figure, but it is recognized as a figure from all four sides, and it is asymmetrical as no other entry is, because the choice may not be reciprocated. It is always relative to a determinate role: one chooses someone as a colleague and not as a friend, and to confuse the two planes is an ordinary way of wounding without meaning to. Of all the figures it is the one with the largest consequences: for a human being, being discarded is among the most heavily charged recognitions, and the emotional signal that accompanies it has the intensity that the chapter *Emotions* attributes to the messages consciousness cannot ignore.

Selection is the point at which all the preceding entries converge, and it is the reason why this essay treats them as inputs to a decision process. None of them determines it. Attraction orients it in most cases, but not necessarily; dependence can keep alive a relationship that all the other entries would advise against; an institutional superiority can remove the choice altogether, because a subordinate does not choose his superior. What the agentological method can do is to make explicit the terms on which the decision is taken, and in particular to distinguish what belongs to the relational situation from what belongs only to the way in which it has been perceived.

The third agent

So far relational figures have been considered within a pair. But a pair is rarely alone: around two agents there are others, and each of them, on entering the picture, modifies the figure between the two. Every dyadic figure is embedded in as many triads as there are third parties that count, and the principle is that the figure between A and B depends also on the figures that A and B have with C.

The triad as the unit of analysis has a well-known precedent: Fritz Heider's theory of cognitive balance (*Attitudes and Cognitive Organization*, 1946), which considers the three relations of a triad as represented by one of the three, and predicts a pressure towards the configurations in which their signs agree. From Heider I take the idea that the triad is perceived by someone and described by no one; I do not take the calculus of signs, because it presupposes figures that are signed and symmetrical, and most of the entries listed above — commanding, depending, giving service — are neither the one nor the other.

The simplest case derives from what the essay already has, without adding anything. If A wants to command B, and C too wants to command B, A and C are in competition: the attainment of the goal of the one reduces the probability that the other will attain his own, which is the definition given in the

chapter *Definitions*. Neither of the two has done anything to the other, and they may never have met: the figure between them is born of the figures each has with a third.

The rule does not hold for every figure, however, and the condition is the one already indicated in the entry *Sharing*: it depends on whether what they want from the third admits of two recipients. If A and C both wanted to cooperate with B they would not thereby be in competition, because B's cooperation is divisible; commanding him is not, or is so only badly, and indeed he who serves two masters serves both badly. The question to ask before a triad is therefore not whether the two have the same figure with the third, but whether what they want from him is divisible.

Other recurrent cases, by way of example. If A depends on B and B depends on C, A depends on C without knowing it, and his predictions about B's behaviour will remain wrong until he notices. If A and B both obey C they find themselves in the same position with respect to him, and this lowers the cost of cooperating with each other: it is the ordinary way in which alliances between subordinates are born. If A imitates C and B too imitates C, A and B will find themselves alike without having sought each other, and the resemblance will produce competition or recognition according as the resources it brings them near are divisible or not. The possible combinations are in the hundreds — twenty-two entries for each of the three sides — and are not to be enumerated: what can be given is the principle and the way of applying it case by case, which is what the agentological method offers in general.

There remains finally, and here more strongly than elsewhere, what the preceding chapter established: the triad is perceived. A's triad is A's, and may correspond to no other. C may be unaware that A exists, and find himself obstructed by someone of whom he knows nothing; A may attribute to B a figure with C that B does not have; two agents may act against each other with determination, each inside a triad the other would not recognize. The recognition of the third is no more reliable than that of the second, and is indeed less verifiable, because the third is usually not present to be questioned.

Misalignment

The yield of this chapter lies not in the list, which is ordinary, but in what is obtained by going through it twice: once from one's own point of view, once from the other's. The cases in which the two recognitions diverge have a recurrent form. One of the two gives a service, the other receives a command: it is the mother who telephones three times a day with which the *Preface* began, and the reason why calling her intrusive explains nothing. One cooperates, the other feels exploited, because the common goal the first speaks of is the first's goal. One imitates out of recognition, the other undergoes an appropriation. One controls out of concern, the other receives distrust: it is the case of Giulia and Marco, where the divergence did not concern the facts, which both knew, but the relational figures each recognized in them.

In all these cases the recognition of the two is incompatible, and neither of the two is lying. There is nothing, in the experience of either, that could correct it: each perceives his own intentions and the effects he undergoes, not the intentions of others nor the effects he produces. What is needed is an operation that sets the two recognitions against each other, and it is the one the chapter *Application to social relationships* calls meta-interaction. The present chapter specifies its object: what a meta-interaction sets against each other are not the facts, about which there is usually no disagreement, but the figures, which the facts are not enough to establish. With two exceptions, which the list now contains: in lying and in manipulation the divergence is not the error of both but the result intended by one, and the remedy is not to reconstruct together what each believed, but to stop taking at face value what is said.

Application to social relationships

The preceding chapter applied the agentological conceptual apparatus to the inner life of a person. This chapter applies it to what happens between persons, and introduces a concept that the axiomatic core does not contain — the well-being linked to the quality of relationships.

What *Psychology of Interacting* says

This essay takes up ideas formulated in an earlier essay, *Psychology of Interacting*, devoted specifically to human interactions. That essay maintains that a good part of our joys and of our sorrows depends on the quality of our social relationships, which in turn depend on the quality of the everyday interactions that build them. It is a thesis widely supported by research on subjective well-being, but it should be understood as pointing to one relevant factor among others — personality, material conditions, health — rather than as a rigid hierarchy among the causes of happiness.

Psychology of Interacting has interacting as its object: not personality, not communication, not relationships taken in isolation, but the process by which two or more people, through a succession of transactions, modify, maintain or restore a relational situation — the configuration of roles, mutual expectations, power relations, constraints and emotional climate that obtains between them at a given moment. Its central thesis is that human beings do not react to the relational situation, but to the interpretation they construct of it. The chapter *Definitions* takes up the term in a narrower sense — the configuration of goals, of reciprocally relevant programs, of asymmetry and of reciprocal effects — leaving out the roles, the expectations and the emotional climate that *Psychology of Interacting* includes in it.

The interpretation of the situation meets the motivational system of the person — needs, desires and values — and out of this a goal is born. The goal calls up a psychological logic, that is, a relatively stable set of rules of the type “if I am in a situation of this kind and I want to obtain this result, then it is best to behave in this way”. The logic orients a strategy, the strategy is translated into transactions, and every transaction modifies the relational situation, from which the round begins again.

The essay distinguishes four levels: the transaction, elementary unit of interacting and behaviour perceptible by the other; the interaction, that is, a succession of transactions within one and the same relational situation; the relationship, that is, the bond that sediments over time between particular individuals; the social network, that is, the set of the relationships in which a person is embedded. And it observes that a relational situation may also be internalized: the other — a dead parent, a distant person, an imagined figure — produces no real transactions, but is internally represented and orients all the same the behaviour of the one who internalizes him.

What agentology adds

The conceptual apparatus of *Psychology of Interacting* — relational situation, interpretation, motivation, psychological logic, strategy, transaction — can be reread as an application of the agentological vocabulary to interactions between persons. The rereading is not an exercise in translation: at four points it brings to light something that in the earlier essay remained implicit.

That one reacts to the interpretation and not to the situation corresponds to axiom A5 together with the definition of perception: the behaviour of an agent depends on the perceived situation, and not on the situation, for the reason that to the situation no agent has access — it has access instead to the information that its perception delivers to it. The gain is not one of abstract generality: if the divergence between the situation and the perceived situation is a condition of every agent, then it is a place to intervene even where one is not accustomed to look for it, in an organization that does not see its own market as much as in a program that does not see the case for which it was not written.

Psychological logic is, in the agentological vocabulary, a *program*: an internal structure that constrains the actions an agent can produce from a given perception. *Psychology of Interacting* observes that logics resemble software in their conditional structure, but differ from it because they are modified from within, through the very experience they help to orient. In agentology that difference no longer runs between man and machine: it is the distinction, established by axiom A6, between plastic programs and rigid programs, which holds for any agent whatever. There are rigid programs in human beings — and it is on this that the following chapter builds its notion of therapeutic change — and plastic programs in machines.

The place that that essay avowedly leaves empty, the place of the emotions, is occupied by the chapter *Emotions* of the present essay. The immediate experience of attraction and repulsion is not, there, a datum to be recorded: it is a message that an unconscious agent sends to consciousness, with an object, a valence and an intensity. And the co-presence of attraction and repulsion towards the same person, which *Psychology of Interacting* notes without explaining, becomes the simultaneous reception of two messages coming from different senders, that is, an intra-agent conflict. That essay says that the experience is really twofold; agentology says why it is.

It happens that one behaves in relation to someone who is not present: a dead parent, a distant person, an imaginary judge. *Psychology of Interacting* records the phenomenon and treats it as a particular case of relational situation, observing that unlike the others it is not verifiable, because the other is not there to contradict us. Agentology describes it differently: that someone is not a situation, he is an agent — a sub-agent of the person, built out of the interactions actually had with the original, and one that goes on evaluating, approving and reproaching according to the criteria that have been attributed to him. The difference is not one of words. A sub-agent can be modified, and it is modified without the real person knowing anything about it; and it goes on operating after that person has died. This is the reason why a therapy can concern itself with the relationship with someone deceased, and the following chapter returns to it.

Finally the social network, which in that essay is the set of the relationships in which an individual is embedded. Agentology adds a question to it: whether the network is itself a higher-order agent. By axiom A8 it is, when at the level of the network there emerge perceptions and decisions not reducible to the sum of those of its members — a common decision procedure, a boundary recognizable with respect to the outside, a persistence distinct from that of the individuals. A family may satisfy those conditions; a group of acquaintances almost never. It is the distinction that makes it possible to establish when it makes sense to attribute a goal to a group, and when instead this is merely a manner of speaking.

These correspondences take nothing away from the specificity of *Psychology of Interacting*: the phenomenology of immediate experience, the distinction between needs, desires and values, the more or less balanced triadic configurations remain contents proper to that essay, which agentology does not claim to correct. What agentology adds is a language capable of placing that model, together with others, within a more general conceptual apparatus, applicable also to organizations and to artificial agents.

Meta-interaction

The essay *Psychology of Interacting* proposes *meta-interaction*: an interaction that has another interaction as its object. The goal is to reconstruct together the interpretations, the motivations and the logics that generated an earlier interaction, or to negotiate a future one.

In this essay meta-interaction is no longer merely a practice conceived for human relationships: it is a particular case of a general operation — reflexive communication about a past or possibly future interaction — applicable in principle to any pair or group of agents, human or artificial, capable of communicating with one another and of some access, however partial, to their own past internal state (memory). If a good part of individual well-being depends on the quality of relationships, and the

quality of a relationship depends on the quality of the interactions that build it, meta-interaction — in the general form proposed here — is an instrument for intervening deliberately on that quality.

Which disagreements make it necessary is stated in the section *Misalignment* of the chapter *Repertoire of relational figures*: they do not concern the facts, about which there is usually no disagreement, but the figures each recognizes in it.

Application to psychotherapy and self-therapy

The preceding chapter applied agentology to the relations between persons. This chapter applies the same conceptual apparatus to the relations between the programs internal to one and the same person, in order to facilitate the understanding — and, where possible, the overcoming — of motivational conflicts.

Intra-agent conflict

Intra-agent conflict — The case of conflict in which the parties in contrast are not distinct agents, but components of the same agent: two or more programs, or two or more sub-agents, which, activated by the same perception evaluated in different ways, tend towards actions antagonistic to or incompatible with one another.

The levels of conflict

Conflict occurs at three levels, which are distinguished by where the boundary between the parties in contrast falls.

The first level is conflict between distinct agents: two persons, but also two organizations, or a person and an institution. It is the level dealt with in the chapters *Repertoire of relational figures* and *Application to social relationships*, under the names of competition, power, dependence and mixed interaction.

The second is conflict between sub-agents of one and the same person. Among these must be counted the internalized models of other people. An internalized parent is a sub-agent built out of the interactions had with him, who goes on evaluating, approving and reproaching according to criteria that have been attributed to him. Two consequences follow that matter for a therapy: that sub-agent can be modified without the real person changing, and it goes on operating after the death or the departure of the original.

The third is conflict between programs of one and the same agent which do not have enough autonomy for them to be called sub-agents: a habit against a resolution, a reflex against a learned rule. The boundary between this level and the preceding one is the one indicated in the chapter *Fundamental concepts*, and it is a matter of degree: in many cases it is not decidable, and for the purposes of a therapy it may not be necessary to decide it, because in both cases what is at stake is modifying programs. The distinction does matter, however, for the difficulty of the undertaking: modifying two programs that reside in the same agent is presumably easier than modifying two that reside in different sub-agents, precisely because the latter are more autonomous.

The three levels can influence one another. A conflict between persons may be sustained, in each of them, by an internal conflict that leads them to interpret the other in one way and not in another. An internalized sub-agent is the sediment of past interpersonal conflicts. And an internal conflict may discharge itself outwards, seeking outside an adversary easier to fight than the one it has within.

The therapeutic instrument for dealing with the first level has already been defined: meta-interaction, that is, the joint reconstruction of what the participants perceived, wanted and decided in a past interaction. For the second and the third level the procedures described in the sections that follow apply.

A word of caution, finally. The level to which a conflict belongs is not always evident: it is itself the object of a reconstruction, and hence of a possible error. To attribute to an internal contrast what arises from a real contrast with another person, or the reverse, is no rare eventuality: it is one of the most common forms of error in the analysis of oneself.

Intra-agent power and akrasia

Intra-agent power — A program exercises power over another program of the same agent when it prevails in a significant and lasting way over the actions that the second would otherwise have produced, even when this conflicts with the goal towards which the second tends. It is the particular case of power in which the one who exercises it and the one who undergoes it are not distinct agents, but components of the same agent: two programs, or two sub-agents.

A person knows what he wants to do, and does the opposite. Aristotle called this akrasia; contemporary philosophy calls it weakness of the will. In the terms of this essay the two poles are two programs: one pursues the goal that the person, on reflection, declares to be his own; the other is an impulse, a habit, a compulsion that runs in the opposite direction. It is the second that prevails, and that prevails stably. No one compels him from outside: the compulsion comes from within.

That this should be possible does not contradict the axiom according to which every agent can modify itself. That axiom guarantees the capacity to act upon oneself, not the coherence of what results from it: an agent can work for itself and against itself.

The chapter *Emotions* has indicated the channel through which this power is exercised: the emotional signal, and the promise of pleasure or of pain that accompanies it.

Therapeutic change as directed learning

The change that a psychotherapy pursues can be reread as intentional learning: the modification of one or more programs as a consequence of the experience constituted by the therapeutic work itself.

Obviously, the programs to be corrected that we are speaking of cannot be read as those of a computer. One can nevertheless try to reconstruct them starting from the behaviours and the emotions of the subjects: it is an activity of *abduction* which in computing is called *reverse engineering*.

A program may be plastic or rigid; therapy works on plastic programs, and a not inconsiderable part of the therapeutic work consists precisely in establishing which programs are, at a given moment, effectively plastic and which are not.

What distinguishes this learning from learning in general is its starting point: not the acquisition of new information about the external environment, but a more accurate perception of a conflict that the previous program implied. Making an intra-agent conflict explicit — recognizing which sub-agents are involved in it and which goals they pursue — is often, in itself, the condition that makes an otherwise rigid program plastic again. This is not a consequence of the conceptual apparatus: it is a hypothesis that the apparatus makes it possible to formulate with precision, and that only clinical work can confirm or refute.

The method applied to oneself

The questions of the agentological questionnaire can be formulated in the first person, and addressed not to an external phenomenon but to one's own agents and sub-agents: which of them are involved in this conflict? What is the situation that activates it? What does each of them perceive? Which figures does each of them recognize between itself and the others? Which programs, in each of them, turn out to be relevant? Which goal does each of them pursue? Which action does it tend to produce? How do these actions, in competition with one another, modify my actual behaviour? What new internal or external situation results from it?

Applied to oneself, this exercise does not necessarily require a clinical setting: it can be carried out in writing, as an ordinary reflective practice.

The specific problem of self-therapy

Application to oneself, however, meets a difficulty that application to an external phenomenon does not meet: the observer and the observed coincide. The perception on which self-analysis is founded is produced by the very programs that self-analysis would like to examine, and is therefore liable to be distorted at the most relevant points. This essay has already met the problem, in connection with freedom: the divergence between the actions that an agent perceives as available and those objectively available is the typical case of self-deception. In self-therapy this risk is not an exception, but the ordinary starting condition.

Some expedients can attenuate, without eliminating, the problem: externalization through writing, which makes one's own reconstruction a stable object that can be reread and called into question after some time; deferred rereading, which makes it possible to compare the reconstruction made in the heat of the moment with a more detached reading; comparison with a third party — a professional or, with the limits this entails, a trusted person — who does not share the same programs and can therefore perceive what the agent, on his own, systematically fails to perceive.

A warning on the use of agentology as a therapeutic instrument

Agentology offers a language for describing and reflecting on ordinary motivational conflicts — ambivalence, procrastination, dysfunctional relational patterns of moderate intensity. It does not offer a diagnostic or therapeutic instrument for conditions that require clinical treatment — trauma, psychosis, major depression, disorders that compromise the very capacity to perceive one's own situation accurately. In such cases the mediation of a professional is not one option among others, but the condition for any analysis, agentological or not, to be called reliable.

An example of application

The two preceding chapters introduced applicative concepts in the abstract. This chapter puts them to the test on an imaginary example, which proves nothing but makes it possible to verify on a case what was asserted in the Preface. The criterion is this: a distinction is worth what it costs when something changes — either the action one chooses, or what one expects to happen. In the story that follows the distinctions introduced change both the one and the other, and at the end of the chapter it is said where.

Marco has worked for three years in the department headed by Giulia. For some months he has been systematically late in delivering the weekly reports, although he knows that the delay irritates Giulia and weighs on his annual appraisal. Giulia, for her part, has reacted to the delays by intensifying controls and asking for daily updates — something Marco experiences as distrust, and which reinforces, instead of attenuating, his tendency to put things off. The relationship between the two, built over three years of largely cooperative collaboration, has progressively taken on specific elements of a negative sign.

Applying the agentological method: the agents involved are Marco and Giulia — the department as a whole does not satisfy, in this case, the criterion of A8, because there exists no decision procedure involving the department as such, independent of the individual persons, and it therefore remains a plurality, not a higher-order agent. The initial situation is a relationship sedimented on a history of collaboration, but with a recent asymmetrical interpretation: Marco perceives Giulia's controls as a judgement on his own reliability; Giulia perceives Marco's delays as a sign of disengagement. In the terms of the chapter *Relational figures*, the two recognize different figures in the same interaction: Giulia sees herself on the side of one who gives a service — attention, support, verification — and Marco receives a command; and it is the divergence between the two readings, not one or the other taken alone, that explains the deterioration.

The relevant programs in Marco include both an explicit professional goal — to maintain a good appraisal — and a habit of avoidance that is activated in the face of tasks perceived as a source of negative judgement: the two programs are in conflict, and the habit of avoidance in fact exercises power over the professional program, prevailing over it even though Marco's declared goal remains that of delivering on time — a case of *akrasia* in the sense defined here. The observable action — the repeated delay — is the outcome of an intra-agent conflict that remains implicit as long as no one names it.

On Giulia's side, the action of intensifying controls is an exercise of power which, meeting Marco's interpretation, produces the opposite effect to the one intended: instead of restoring reliability, it reinforces the avoidance program, because it confirms to Marco the interpretation he fears. The interaction between the two, though rooted in a relationship with a cooperative basis, locally takes on a competitive dynamic — exactly the case of mixed interaction already provided for in the note on cooperation and competition.

The turning point comes when Giulia proposes a direct discussion of what is happening between them: a meta-interaction, in the sense defined in the chapter *Application to social relationships*. In the course of the discussion it emerges that Marco had never perceived, among the actions at his disposal, that of openly explaining the workload difficulties that lay behind the delays: not because it was objectively barred to him, but because it did not fall within his freedom, in the sense defined here — a case, precisely, of divergence between the actions that an agent takes to be available and those that are objectively available to him.

The immediate outcome of the discussion is a change of perception, not yet a change of program: Marco reinterprets Giulia's controls as concern rather than as distrust, and Giulia reinterprets the delays as a signal of overload rather than of disengagement. This is enough to modify the subsequent actions of both — Marco starts delivering on time again, Giulia relaxes the controls — without their respective

programs (Marco's habit of avoidance, Giulia's style of control) having yet been touched at the root. If the new pattern stabilizes over time, there will moreover be genuine learning: the programs themselves, and not only the contingent actions, will turn out to have been modified; but this is a further outcome, not guaranteed by the change of perception alone.

A purely systemic description of this episode — a negative feedback loop between the behaviours of Marco and Giulia — would not have been false, but it would have left undistinguished a number of elements that here turn out to be separable: the difference between the exercise of power and the mere capacity to act; the difference between what an agent can objectively do and what it perceives itself as able to do, which is the precise point on which the discussion acted; the conflict between two programs of the same agent or two sub-agents, which explains why Marco could not do what he himself would have wanted to do, without having to invoke anything beyond the apparatus already introduced; and the observation, by no means obvious, that the meta-interaction worked by modifying perceptions, not necessarily the underlying programs. These are distinctions that a lexicon of inputs, outputs and feedback alone can certainly express, but does not name separately; and having named them separately, here, has produced the two effects announced. It has shown Giulia where to act: not on Marco's will, which was not at issue, but on the list of the actions he believed he had at his disposal — and this is what the discussion did, by adding one to it. And it has told them both what to expect: since it is the perception that has changed and not the program, the improvement is real but fragile, and under pressure the old behaviour will return, if in the meantime nothing else has happened. Without that distinction the affair would have been filed away as resolved, and the relapse would have come as a surprise.

Conclusion

In order to understand something we are accustomed to ask what it is or what substance it is made of: what kind of person he is, what character he has, what that organization is like. Agentology proposes changing the questions. Not what a thing is, but who acts, what he perceives, what he wants, according to what logics he takes his decisions, which actions he believes he has at his disposal, and so on.

Saying that a person is lazy closes the discussion: nothing remains but to observe it again and confirm it. Asking instead what the person sees when he is faced with a task, what he fears, which options he believes he has, yields differentiated answers, and each of them could be different if the situation changed. A judgement on character leaves nothing to be done; agentological questions bring out areas in which one can intervene. They do not guarantee that the intervention will succeed: they say where to attempt it.

Emotions, the relations between people, internal conflicts. At first sight these topics seem to have nothing whatever in common, but in each of them agentology sees agents that interact according to perceptions, logics and goals, and it proposes the same questions for understanding why what happens happens.

This essay does not say how one ought to treat others, but it urges us not to dismiss a person by saying that he is made in a certain way as an explanation of his behaviour. This does not mean agreeing with everyone, nor forgiving everything: whoever has intentionally caused harm, being able to act otherwise, answers for it, and this is what is here called responsibility. It means that one cannot stop at the first explanation that comes to mind, which is usually the most convenient and almost always useless.

I leave it to the reader to judge whether the conceptual apparatus proposed here is useful. The next time he sees someone do something that seems inexplicable, he can try to answer the questions of the agentological questionnaire. If afterwards he understands something he did not understand before, the time he has spent reading this essay will not have been wasted.