

Bruno Cancellieri, ChatGPT, and Claude.ai

The Psychology of Judgment

How Social Behavior Is Governed by the Judgments
we anticipate and those we form

August 4, 2026

Copyright © Bruno Cancellieri

We behave based on judgments that almost no one actually voices. Before we speak, reveal ourselves, or choose what to reveal and what to keep to ourselves, our mind assesses how we will be judged—and reacts to that assessment more than to the situation itself.

That assessment is not based on an objective reading of others. It is constructed primarily based on how we see ourselves and our relationship with others; what others reveal does play a role, but it carries less weight than we believe.

Our judgments are observable, and we unconsciously anticipate being judged based on the way we judge others: that anticipation thus becomes part of the criteria we use to evaluate others.

The resulting judgment is perceived and judged by others even when it is not explicitly expressed. This is the *recursion* from which this essay takes its starting point: an evaluation that becomes the object of evaluation, and which, for this reason, is never merely a description of what is being evaluated. On the other hand, judgments tailored to our listeners in order to gain approval or avoid conflict are not merely a pretense: by expressing them repeatedly, we end up believing what we have said.

The mechanism, in and of itself, is not flawed: it makes trust, cooperation, and reputation possible, as well as competition and conflict. The problem is not how it works, but how it is calibrated—it is calibrated for an environment in which those who judged also made decisions, and today it continues to react to judgments that will have no consequences.

The essay does not teach us how to get rid of it. It proposes a model for recognizing when it kicks in, a criterion for distinguishing between actions that might better calibrate it and those that fuel it, and two questions to keep in mind: “Can this person change anything in my life?” and “What would I think if my judgment were not observable by anyone?” They do not prevent the system from acting; they help us observe a mechanism that normally operates unnoticed.

Table of Contents

Abstract	5
Note	6
Introduction	7
PART ONE — The Mechanism	11
1 To evaluate is to assign value	12
2 When Evaluation Becomes Recursive	17
3 What Is the Purpose of Judgment?	20
4 Presumed Judgment	24
5 Being Judged: Where Does the Presumed Judgment Come From?	27
6 The observer we carry within us	32
PART TWO — The Two Sides	36
7 To Judge Is to Expose Oneself	37
8 Expressing a judgment changes us	43
9 A Mechanism Calibrated for a World That No Longer Exists	49
PART THREE — What to Do With It	53
10 Why Intuitive Remedies Make Things Worse	54
11 The Principle of Recalibration	59
12 Two Fundamental Questions	65
13 The Decline of Recalibration	70
14 When to Stop the Observation	73
Conclusion	77
Appendix I — References	79
Appendix II — What the Model Asserts	85
Appendix III — Questions for Further Reflection	89
Appendix IV — Glossary	92

“Society is an interweaving and interplay of minds. I imagine your mind, and above all what your mind thinks of mine, and what your mind thinks of what mine thinks of yours. I present my mind to yours, and I expect you to present yours to mine.”

Charles Horton Cooley, 1902

Abstract

Psychology of Judgment, by Bruno Cancellieri, written in collaboration with ChatGPT and Claude.ai, proposes a model of judgment built around its recursiveness: we evaluate someone or something, and that evaluation can in turn be evaluated.

The central thesis concerns the representation each of us forms of others' judgments—the presumed judgment—and argues that it is constructed primarily based on how we see ourselves, while feedback from others carries a reduced and variable weight.

Two independent parameters govern its construction and maintenance: presupposition—that is, the extent to which an estimate is already fixed before observing—and discounting—that is, the extent to which feedback that contradicts it is devalued. Each can be high or low, and the four combinations correspond to four distinct profiles: (1) social anxiety, where both are high; (2) narcissism, where only discounting is high; (3) an unstable self-image, where only presupposition is high and the sense of how one is perceived fluctuates with every piece of feedback; and (4) the best-calibrated functioning, where both are low.

The essay then turns to the active side: expressing a judgment exposes the person making it, and a judgment tailored to the interlocutor tends to become sincere. The mechanism, in and of itself, is not flawed—it enables trust, cooperation, and reputation, as well as competition and conflict—but it is calibrated for an environment in which the judge was also the decision-maker, whereas today people react even to judgments that will have no consequences.

The final section shows why spontaneous remedies—from self-restraint to seeking confirmation—tend to fuel the mechanism rather than correct it, and proposes a criterion for evaluating any intervention. This leads to two questions to keep in mind: “Can this person change anything in my life?” and “What would I think if my judgment were not observable by anyone?”—the first addressing the passive aspect, the second the active one. They do not prevent the system from reacting: they make observable a mechanism that operates beyond one's own scrutiny.

The essay does not claim theoretical originality: the framework already exists in the clinical study of social anxiety and in predictive processing models. The stated contribution is the generalization from the pathological to ordinary functioning and the connection between the passive and active aspects.

Note

This article stems from a dialogue with ChatGPT and Claude.ai, following a method I have also adopted for other works: I propose a topic, discuss the structure and content with these tools, and have them write and then revise the texts through a detailed exchange, in which changes are proposed, justified, and discussed together.

The specific theoretical framework develops through these dialogues, but it is rooted in a psychological and philosophical background that I developed independently long before turning to these tools.

The responsibility for the arguments put forward—and for any weaknesses they may have—remains mine, however.

Bruno Cancellieri

Introduction

What This Essay Argues

It is worth dispelling two expectations, both reasonable and neither of them relevant. This is not an essay on judgment as a decision-making process—heuristics, biases, probability estimates—which is another area of research. Nor is it an essay that urges us to judge less.

On this last point, it is worth noting that the Italian language is less unambiguous than the word might suggest. “Do not judge” treats judgment as a flaw; but “to have good judgment,” “a judicious person,” and even the “wisdom tooth”—so called because it erupts at the age when judgment is supposed to have matured—treat it as a virtue, the lack of which is a reproach. The language captures both aspects, and not by accident: judgment is a necessary faculty, and like any faculty, it can be exercised well or poorly.

The judgment discussed in these pages is an ordinary, largely automatic process, without which social life would not function. It is described, not deplored—and not praised.

Every day we make judgments. We say that a movie is boring, that a person is trustworthy, that a political decision is wrong. And at the same time, we constantly imagine how others judge us. Much of social life unfolds within this web of expressed or presumed judgments.

It is worth distinguishing right away between two words that will have different meanings in what follows. *Evaluating* and *judging* are both acts of assigning value, and in this sense, judgment is a type of evaluation. But evaluating means estimating how much something is worth to us, based on our needs, while judging means transforming that estimate into a verdict that influences a social relationship. When the object of judgment is a person, the verdict influences how we will treat that person or our expectations of how they will treat us; when it is a thing, it nonetheless exposes the person who expressed the judgment to the judgment of others. The first chapter will explore this distinction in greater detail; for now, it suffices to know that the two terms are not synonyms.

We are spontaneously inclined to think that these two processes are independent. We believe that, when we judge something, our judgment depends primarily on the characteristics of what we are judging. And we believe that, when we imagine others’ judgment of us, our representation depends above all on what they do, say, or show.

The thesis of this essay is different.

In social situations, the judgments we form are also influenced by the consequences those judgments will have for us. And the representation we construct of others’ judgments derives, to a large extent, not from direct observation of others, but from our own self-perception.

The result is a system that uses, as its primary source of information, what it has itself produced.

In summary: the system performs a recursive processing of input that is not observed but inferred, constructing it primarily from its own self-perception.

This formulation summarizes the theoretical principle of the essay.

By “*recursive*,” we do not mean a conscious reflection of the type “I think that you think that I think.” Research shows that spontaneous recursion is surprisingly limited: in most situations, it does not exceed one or two levels. It is a rapid, automatic, and largely unconscious process.

By “inferred input,” we do not mean information observed with low precision. The problem does not concern the quality of the measurement, but its origin. In most social interactions, we do not truly know the judgments others form about us. The mind fills this gap by constructing a representation that derives primarily from how we see ourselves and from previous experiences with others in general.

What This Essay Contributes

This principle did not originate with this essay.

In my previous works, particularly **Presumed Judgment** and **Social Self-Governance**, I had already introduced the concepts of presumed judgment, the social supervisor, expected relational cost, and evolutionary maladjustment, along with their connection to *predictive processing*. This work, therefore, does not propose a theory formulated from scratch, but rather represents the consolidation of a research path that has already been initiated, reorganized, corrected, and subjected to systematic critical review.

The scientific literature also already contains a significant portion of the mental architecture described in these pages. Clark and Wells’ model of social anxiety describes a system that constructs an image of how one appears to others primarily based on internal information rather than on observation. *Predictive processing* models formalize a very similar architecture in computational terms. Other lines of research have studied, separately, metaperceptions, reputation, judgment as a social signal, the falsification of preferences, and the transformation of judgments through communication.

The essay’s most significant contribution lies in synthesizing these traditions within a unified theoretical framework, integrating them with the approach developed in previous works.

The Most Risky Proposal

The most important proposal, however, does not concern this synthesis.

It consists of a prediction.

Clinical models describe the mechanism just outlined as characteristic of social anxiety and other conditions in which the representation of others’ judgments is strongly dominated by self-perception. This essay, however, proposes that the same mechanism is part of ordinary social functioning and that social anxiety represents its extreme case.

This is the model’s riskiest claim, and for this reason it is not presented as an established fact but as an empirical prediction.

Some research findings are consistent with this prediction. Taxometric studies suggest that social anxiety is distributed along a continuum rather than organized into clearly separate categories. However, this concerns the distribution of symptoms, not that of psychological mechanisms. A phenomenological continuum does not prove that people with different levels of social anxiety use

the same process with varying intensity. This is precisely what the model proposes and what will need to be tested.

What the paper adds

In addition to this prediction, the paper introduces several extensions and corrections to previous work.

The first concerns what determines when the judgment of others becomes truly important. In previous works, this was formulated in terms of the observer's social weight and the probability of future interactions. In this essay, it is redefined as the power that the other person has over us. The difference is not merely terminological. Two people may be destined to interact daily, yet one may have no reason to manage their reputation with the other: the relationship between master and slave is a clear example of this. The probability of interaction is not enough to explain this asymmetry; power is.

The second correction stems from a prediction of the model that was falsified during its design. An analysis of the literature on narcissism has shown that metaperception cannot be described as a simple consequence of self-perception. The model therefore distinguishes between two independent parameters: the extent to which metaperception is constructed from self-perception and the weight attributed to information that could correct it.

The third correction involves introducing a prediction regarding the depth of recursion: the number of recursive levels should increase with social anxiety. This, too, is not presented as an established finding, but as a testable hypothesis.

Finally, the paper incorporates the corrections that emerged during the verification process. Some hypotheses were abandoned, others scaled back, and still others reformulated.

The Scope of the Model

The model does not claim to explain the entirety of social life.

Its scope is more limited. It concerns situations in which an evaluation can, in turn, become the subject of evaluation. In these cases, the judgment not only describes what is being evaluated but also reveals something about the person making the judgment. It is this recursive structure—and not the mere act of evaluating—that constitutes the subject of the essay.

Its implications, however, may be broader than they appear.

If this mechanism influences the judgments we express and those we attribute to others, it might also influence what we are willing to consider. Even before asking ourselves whether an idea is true or false, the mind might implicitly evaluate its relational consequences: what would change in my social life if I took this idea seriously? This hypothesis will be discussed in the second part of this essay.

What This Essay Does Not Promise

Understanding this mechanism does not mean learning to better control one's social image.

In fact, one of the most robust findings from research on social anxiety shows that excessive self-focus contributes to perpetuating the problem. The aim of this essay is the opposite: to help the reader recognize that presumed judgment is often a construct of one's own mind, and that intuitive remedies—thinking more, constantly monitoring oneself, seeking reassurance—tend to reinforce the mechanism rather than correct it.

How It Is Organized

The essay is divided into three parts.

The first describes how the system works: what an evaluation is, what happens when it becomes recursive, why others' judgments matter, and through what mental representation they operate.

The second part shows how the system operates on two fronts: when we imagine being judged, and when we are the ones doing the judging—to the point where we tailor a judgment to our audience and then end up truly believing it ourselves. It concludes by explaining why a mechanism calibrated for small groups continues to react in a world of strangers.

The third chapter addresses change: why spontaneous remedies fail, under what conditions the system can recalibrate itself, what tools emerge from this, why both that recalibration and those tools eventually decay, and when it is advisable to stop self-observation.

Four appendices include references, what the model asserts, some questions to ask oneself when faced with a preformed judgment, and the terms used in a technical sense.

Two Fundamental Questions

The following pages ultimately converge on two fundamental questions.

- Can this person change anything in my life?
- What would I think if my judgment were not observable by anyone?

These are questions anyone can ask themselves in a matter of seconds, when faced with a judgment they've received or while forming one. This essay explains why it's worth asking them and what one can reasonably expect from the answer.

In addition to these two questions, there are others, compiled in the appendix "*Questions for Further Reflection*": these are not tools to keep at hand the moment a judgment arises, but aids to consult with a clear head, when it remains to be decided what to do with a judgment that has already been formed.

PART ONE — The Mechanism

1 To evaluate is to assign value

Every moment of our mental life is permeated by evaluations.

We walk into a room and perceive it as welcoming or hostile. We taste a food and find it pleasant or unpleasant. We read a news story and consider it interesting, irrelevant, or concerning. A sudden noise alarms us, music relaxes us, a face inspires trust or distrust.

These processes are so rapid and habitual that they go almost unnoticed. Yet they constitute one of the mind's fundamental functions.

Even before deciding how to act, we must determine what, for us, has value (positive or negative).

In this essay, we will refer to this basic process as "evaluation": the assignment of a positive or negative value, of a certain intensity, to a person, an object, an event, an idea, or a situation.

The definition is deliberately broad. It does not yet concern social judgment, reputation, or relationships. It concerns a general cognitive function, without which no goal-oriented behavior would be possible.

A mind incapable of evaluating would be unable to distinguish what furthers its goals from what hinders them, what deserves attention from what can be ignored, and what is worth pursuing from what is best avoided.

To evaluate means to attribute practical meaning to experience.

Not all evaluations are conscious. In fact, most occur automatically: we notice their results more easily than the process that generated them. We sense that we like or dislike something long before we can explain why.

A Level Already Studied, and a Broader One

Psychology has long studied this fundamental level.

In particular, *appraisal* theories, developed by authors such as Richard Lazarus and Klaus Scherer, describe how events are evaluated in relation to our goals, our needs, and our possibilities for action, giving rise to emotions.

This essay explicitly draws on that tradition. It does not intend to replace it, nor to propose an alternative theory of appraisal.

On the contrary, it assumes that such research has clarified an essential point: emotions do not depend directly on events, but on the meaning the mind attributes to them through a process of appraisal.

The level we are concerned with here, however, is broader.

Appraisal has been studied primarily as the process through which an evaluation produces an emotional response. In these pages, however, evaluation is considered a general cognitive function, of which emotional appraisal represents one of the most important—but not the only—instances.

We can evaluate an argument as convincing without experiencing an intense emotion. We can consider a source reliable, deem a project promising, or judge a procedure to be inefficient. In all these cases, the mind assigns value, even when the emotional component is minimal or barely discernible.

The basic level of the model therefore concerns every form of value attribution, regardless of whether it produces an obvious emotion.

This choice is important for another reason as well, which should be stated openly.

If we were to equate evaluation solely with emotional appraisal, the model proposed in these pages would coincide with an existing theory, and its contribution would be reduced to merely different terminology.

The starting point, however, is more general: the mind continuously evaluates the world based on what it considers relevant to its own purposes, and only a portion of these evaluations has been studied by the psychology of emotions under the term “appraisal.”

What Determines Value

As a first approximation, value is determined by our needs and desires.

These concepts will not be redefined here. They have been discussed at length in my essay **The Psychology of Needs**, to which this work refers. Suffice it to say that needs and desires constitute the criteria through which the mind assigns value to events and guides behavior.

The same situation can therefore be evaluated very differently by different people—not because one perceives reality correctly and the other does not, but because they assign value to different things.

For someone who is hungry, a sandwich has great value; for someone who has just finished a hearty lunch, the same sandwich may seem irrelevant or even repulsive.

For someone seeking peace and quiet, a noisy party represents a cost; for someone looking for opportunities to socialize or have fun, the same party represents an opportunity.

Evaluations change because the criteria by which value is assigned change.

Evaluating and Judging

As mentioned in the introduction, the two terms are not synonyms. It is time to clarify the distinction, which will be maintained throughout this essay.

By “*evaluation*,” we mean the attribution of value described so far: this thing, for me, at this moment, is worth a lot or very little. It is an open-ended process. It allows for degrees of value; it can be suspended or corrected. “I’m still evaluating” is a phrase that makes sense.

By “*judgment*,” we mean the evaluation that becomes a verdict and influences our relationship with what has been judged. These are two distinct elements, therefore, and it is best to consider them one at a time.

The first is the verdict. A judgment is not a point on a graduated scale: it is a qualification. It is good, it is unfair, it is mediocre, it is unreliable.

The difference lies in how we use the two words. We can evaluate a property at two hundred thousand euros, or a test at eight out of ten. We cannot judge eight out of ten.

And we can say “I’m still evaluating,” but we cannot say “I’m still judging.” A judgment is a point of arrival, a final decision, not a process.

The second element is the relational consequence, and this is what *essentially* distinguishes judgment from evaluation. Someone who judges a person does not merely record their worth; they adopt a more or less favorable attitude toward that person. The verdict influences what we will do—whether we will trust them, seek them out, avoid them, or choose them when we have the opportunity to choose. This is why feeling judged weighs more heavily than feeling evaluated. It’s not that it’s unpleasant for someone to have formed an opinion about us; it’s that that opinion will shape their behavior toward us.

And what if the object isn’t a person? Even judging a work of art, a law, or a decision has relational consequences, but these fall on the judge. Declaring a work of art mediocre means exposing oneself, because the judgment also says something about the person expressing it. We’ll see later just how much this matters.

The difference between the two terms, therefore, does not lie in the object—both can have things or people as their object—but in their respective consequences.

In everyday language, the two terms constantly overlap, and there is nothing to prevent one from saying that one is judging a wine or evaluating a colleague. Here, however, keeping them distinct will help separate the general operation from the specific case that interests us.

This distinction, however, does not yet explain why social evaluations behave so differently from others. The truly decisive characteristic is another, and the next chapter will introduce it.

With people, evaluation becomes a verdict

So far, evaluation has been described as an open-ended process: an assessment that can be revised, adjusted, or corrected. This is how it works with objects. With people, something different happens.

By observing someone’s behavior, we automatically infer stable traits—they are disorganized, generous, or unreliable—without having decided to do so, without an explicit purpose, and without realizing the process is taking place. This phenomenon is called spontaneous trait inference; the line of research dates back to James Uleman and his collaborators, and forty years of studies conducted using various methods have confirmed its existence.

This is a finding that deserves attention, because it describes the transition from estimation to judgment. An estimate of weight remains an estimate: it stays on a scale and is open to revision. An inference about a person becomes a predicate—it is so—and predicates have no degrees.

Two aspects make the phenomenon more widespread than it appears.

The first is that inferences are formed even on the basis of very few elements: it is not necessary to know someone; a single observed behavior is enough. This is why “I have no basis for judgment” is

almost always false. It does not describe an absent judgment, but a judgment formed on very little—and the fact that the person saying it is acting in good faith is part of the phenomenon, rather than refuting it.

The second concerns the outcome that appears to be an absence of judgment. Determining that someone is irrelevant, that they aren't worth paying attention to, that they can be ignored without consequences—this is not a failure to judge: it is a judgment, and one of the most consequential, because it establishes that the other person has no power over us. Even deciding that it isn't worth gathering information about a person presupposes an assessment of their expected value.

One cannot help but judge

From the foregoing follows a three-part argument.

The first: we continually evaluate things and people, and we cannot help but do so. This is the function described at the beginning of this chapter.

Second: when the object is a person, the evaluation tends not to remain open-ended. It becomes a dispositional attribution. This may happen deliberately, but it also happens unintentionally and unconsciously.

The third: since in society we are surrounded by people who perceive us—and for whom the same premises apply—it follows that

One cannot help but judge, nor can one avoid being judged.

This is a strong claim, and it is worth stating right away what would refute it. If inferences about people required intention, or if they remained situational and adjustable rather than solidifying into dispositions, it would be false. The available research indicates the opposite.

What this does not authorize

A conclusion of this kind can easily be interpreted in two erroneous and opposing ways.

The first is the moralism that it refutes. There is a tradition—moral rather than psychological—according to which *judging* is in itself a flaw—categorizing, labeling, condemning—and *not judging* is a virtue that can be attained through practice. If the argument just presented is correct, that virtue is unattainable: the verdict is formed before a decision is even possible, and resolving not to form it is like resolving not to recognize a familiar face.

The second error is the mirror image of the first, and it should be ruled out just as decisively. The fact that a judgment forms anyway does not mean it can be expressed freely. Forming a verdict and declaring it are two distinct acts: the first is automatic; the second is a move with its own consequences, and a later chapter will show that whoever makes it exposes themselves. To argue that one cannot help but judge is therefore not a license to judge aloud; and the ancient injunction against judging—interpreted as referring to speaking out and condemning rather than to thinking—remains perfectly sensible.

The essay describes a mechanism. It does not advise either going along with it or fighting it.

Toward the Next Chapter

So far, we have spoken only of the fundamental function of evaluation.

The introduction has already indicated that the true subject of this essay is a particular case of this process: one in which an evaluation is not limited to describing its object, but can itself become the object of further evaluations.

Social evaluation is not a faculty distinct from ordinary evaluation. It is the same cognitive operation situated within a social context.

When we assign value to something on our own, the evaluation ends with the object.

When we express it in the presence of others, or know that it may be observed, the situation changes. From that moment on, it is no longer just the object of our judgment that is evaluated: the very fact of having evaluated it also becomes information for others.

It is from this transformation of the structure of evaluation that the model developed in the following chapters begins.

2 When Evaluation Becomes Recursive

Let's imagine tasting a wine on our own, in the kitchen, at the end of the day. We find it good, or mediocre, or excellent. The evaluation arises, takes place, and ends there. No one will ever know about it.

Now let's imagine the same wine at the dinner table, with five people, one of whom chose it.

Something has changed—and it's not the wine.

What has changed is that our evaluation has become observable. And the moment it becomes observable, it ceases to concern only the wine: it, in turn, becomes something that others can judge. If we say it's excellent, someone will judge that we know our wines, or that we're being complacent toward the person who chose it. If we say it's mediocre, someone will judge that we have a refined palate, or that we're rude, or that we're putting on airs.

The point is not that we worry about these possibilities. Often we don't even notice them at all. The point is that the structure of the situation has changed: the evaluation has entered a cycle in which it itself is the object of evaluation—and often of judgment.

We call an evaluation that finds itself in this condition "recursive."

This is the property that the previous chapter left unresolved. What distinguishes social evaluations from others is not their object—we can evaluate people without anything like this happening, and evaluate objects in situations where it does happen—nor is it the type of criterion invoked. It is their systemic structure.

What Recursiveness Excludes

A defining property is useful only if it leaves something out. It is therefore worth stating right away which evaluations fall outside the scope of this essay.

It does not include evaluations that no one can know and on which nothing depends: the silent judgment of a landscape passed through by train, the never-stated preference between two shades of gray, the instinctive estimate of water temperature.

It does not include immediate reactions to physical stimuli—pain, hunger, alarm—that require no relational foresight.

Nor do they include, except marginally, evaluations made under conditions of true and definitive anonymity.

On this last point, however, a clarification is needed that will recur several times in this essay. The boundary is not clear-cut, because the mind does not seem to recognize anonymity well. In double-blind economic experiments, where no one can trace who decided what, participants' generosity plummets—which confirms just how much the possibility of being observed matters—but it does not drop to zero. Something continues to function as if someone were watching.

The scope of this essay is therefore narrower than that of evaluations in general, but its boundaries are blurred. This blurring is not a flaw in the description: it is a property of the phenomenon, and in the chapter devoted to the internalized observer, it will itself become the subject of analysis.

A Structure Already Described Elsewhere

Recursion is not a new idea. Its most famous formulation dates back to John Maynard Keynes, who in 1936 illustrated it with the image of a beauty contest organized by a newspaper. Readers must select the six most beautiful faces from among a hundred photographs, and the winner is whoever comes closest to the average choice of the participants. The shrewd contestant, Keynes observed, does not choose the faces he finds most beautiful: he chooses those he predicts others will choose. And since others reason in the same way, each person's choice depends on the predictions each makes about the predictions of others. Keynes was describing the stock market, but he was describing a much more general structure.

The same structure recurs elsewhere under different names. In sociology, Niklas Luhmann called it double contingency: each person adjusts their behavior based on the expectations they attribute to the other, who does the same. In psychiatry, Ronald Laing described its spirals—"I know that you know that I know"—as part of everyday clinical practice. In game theory, it corresponds to higher-order beliefs. And the tradition stretching from George Herbert Mead to Erving Goffman has built much of its understanding of social life upon it.

This essay therefore does not claim to have discovered recursion. As already stated in the introduction, the overall architecture described here exists in clinical form in the Clark and Wells model and in computational form in predictive processing models. What the essay proposes is to consider it ordinary functioning rather than a pathological mechanism, and to systematically extend it to the realm of judgment, which those traditions do not address.

How Deep Is Recursion?

Described in words, recursion may evoke the image of a dizzying spiral: I think about what you think about what I think about what you think, and so on endlessly.

The image is striking, but it is false. The beauty contest game was replicated in the laboratory in a simplified form: participants are asked to choose a number between zero and one hundred, and the winner is the one whose choice comes closest to a fraction of the average of all the choices. The logical solution to the game would require an indefinite number of iterations. Real people, on the other hand, almost always stop after one or two rounds. Rosemarie Nagel has shown that the observed behavior is well described by a multi-level model: most players think about what others will do, some about what others expect them to do, and almost no one thinks beyond that.

In short, the actual depth of recursion is shallow. It is not a spiral: it consists of one or two steps, taken quickly and almost never consciously.

From this finding arises a hypothesis proposed in the essay that has not yet been tested: that the depth of recursion is not the same for everyone, and that it increases as social anxiety rises. According to the model, those who are very afraid of others' judgment should go up more steps—not only wondering if they'll seem uncomfortable, but also if others will notice that they're trying not to seem that way.

The Main Consequence

If an evaluation can be evaluated, then expressing it is never a neutral act.

It follows that, in social situations, our evaluations do not depend solely on what we are evaluating. They also depend, to varying degrees, on what that evaluation will entail for us.

This statement should be carefully considered, because in its absolute form it would be false. It is not true that we never evaluate things in and of themselves. What is true is something more precise and verifiable: the distortion increases in proportion to the observability of the evaluation and the significance of its consequences. A judgment expressed in front of someone who can influence our career is more exposed than a judgment expressed among friends, which in turn is more exposed than a judgment confided to a diary.

And, as we have seen, it does not drop to zero even when no one is watching.

This graded formulation has an advantage over the absolute version: it specifies where to expect the effect and where not to, and can therefore be falsified. If, by experimentally manipulating the observability and consequences of a judgment, no systematic variations in its expression were observed, the model would be incorrect.

What's Still Missing

Recursion explains why social evaluations are conditioned. It does not yet explain why they are so often wrong.

To adapt an evaluation to how it will be judged, in fact, one must have an estimate of how it will be judged. And this estimate must come from somewhere.

It would be natural to assume that it comes from observation: we watch others, gather clues, and make corrections. This is where the model encounters its most counterintuitive result. Because that estimate, to a large extent, does not come from observing others.

It comes from us.

3 What Is the Purpose of Judgment?

So far, we have described a property of social evaluations: their recursiveness. An evaluation, once it becomes observable, can itself be evaluated. From this structure arises the possibility that our evaluations may be modified in anticipation of how they will be received.

A deeper question remains, however.

Why do we care so much about the judgment of others?

The most immediate answer would seem to be: because we desire approval. Or because we want to be appreciated, avoid rejection, gain prestige, or maintain a good reputation.

All these answers contain some truth, but they describe the phenomenon at a still superficial level. Approval, prestige, conformity, shame, and the pursuit of status are different manifestations of a more general problem. To understand them, we must ask ourselves what function judgment serves within human relationships.

Social Value

Every relationship involves, implicitly or explicitly, a choice.

We choose whom to collaborate with, whom to trust, whom to listen to, whom to spend time with, from whom to learn, with whom to work, whom to hire, whom to invite, whom to help, and whom to avoid.

Judgment helps guide these choices.

When we evaluate a person, we do not merely describe them: we generate information that can influence how others will relate to them. Being judged favorably increases the likelihood of being chosen; being judged unfavorably decreases it.

The judgment of others is therefore not an end in itself. It is a signal on which our social value depends: how desirable we are as interaction partners.

By “interaction partner,” we do not mean only a romantic partner. The same principle applies to a friend, a colleague, a collaborator, an employee, an employer, an ally, or a role model. The situations may change, but the fact remains that someone may prefer us over someone else.

It is the same process described in the first chapter, with the roles reversed. There, value was what something is worth to the evaluator, according to their needs and desires; here, the object being evaluated is us, and the criterion lies in the needs of the evaluator.

It follows that a person’s value is not a single quantity, nor is it a property of that person. It is relative to the evaluator and to what they need: one can be worth a great deal to one person and very little to another, without either of them being wrong, and without the person having changed.

The system, however, does not present it this way. The presumed judgment comes across as a single, sweeping verdict—“I’m worth little”—rather than as the assessment of a particular evaluator in light of what they need.

And we are not merely evaluated. We, in turn, evaluate others, and this, too, comes into play: the value of our judgments—whether we are people whose advice is worth seeking or from whom it is worth learning—is itself the subject of others’ evaluation. The evaluator is not outside the market in which they are evaluated—which is why the second part of this essay will also address the active side of the equation.

Being Chosen

Many phenomena that seem distinct can be traced back to this same mechanism. Cooperation, in fact, is choosing someone to do something together with; friendship is mutually choosing one another as privileged companions; and personnel selection is preferring one candidate over another.

Even more evident is the choice of a romantic partner, where reciprocity is explicit. In short, under different names, the same process always recurs: someone chooses, and someone is chosen.

Competition is no exception either. It is simply the case where multiple people seek to be chosen for a limited resource. We do not compete against someone just for the sake of competing: we compete to be preferred.

Being chosen is therefore the primary way in which judgment produces consequences. It is not the only one: those who judge us can also help or hinder us, reward or punish us, or speak about us to others. But it is the one that occurs most often, and it is around this that the system has plausibly formed.

Approval, prestige, and reputation thus take on a more precise meaning. They are not ultimate goals, but rather ways through which the likelihood that others will choose us increases or decreases.

What Judgment Makes Possible

Before moving on to the limitations, it is worth pausing to consider what this mechanism produces when it works—because it normally works, and this is the part we tend to forget.

Cooperation among unrelated individuals is rare in nature and striking in our species. It requires that each person be able to assess, before committing, how trustworthy the other is, and that each person know that they, in turn, are being assessed. Without a system capable of anticipating the social consequences of one’s actions, that assessment would not exist.

This same capacity underlies our trust in people we do not know, our adherence to rules when no one is watching, social learning—which requires recognizing whom to emulate—and reputation as an institution, that is, as information that circulates and outlives the person who produced it.

The self-censorship we will discuss as a problem is the very same capacity that prevents us from saying at a funeral what we really think of the deceased.

From this follows a consequence worth stating, because it runs counter to a widespread habit.

Condemning judgment across the board means condemning the mechanism that makes coexistence possible. This is not a moral defense of judgment: it is the observation that what is commonly deplored as “judging”—classifying people, predicting how they will behave, and adjusting one’s

own behavior accordingly—is the very same process that allows us to trust someone, to choose a collaborator, and to recognize who deserves to be heard.

The phenomena that the following chapters will address as problems are therefore not the functioning of this system. They are its miscalibrations.

How Much Does a Judgment Weigh?

This conclusion, however, would still be too general. If judgment served only to determine social value, any judgment should carry more or less the same weight—and experience shows that this is not the case: the same comment can leave us completely indifferent or worry us for days, depending on the person who makes it. The difference does not lie in whether we like or dislike that person, but in their power to influence something that concerns us.

A judgment expressed by a stranger we've met only once generally carries much less weight than one expressed by a colleague we work with every day, by a supervisor who decides on a promotion, or by someone on whom an important relationship depends.

What makes a judgment relevant, therefore, is not simply its existence, but the power that the person making it has over us: the extent to which what that person does—or does not do—can change something in our lives—choosing us or rejecting us, helping us or hindering us, rewarding us or punishing us, speaking positively or negatively about us to others.

That power does not create the effect of the judgment: it determines its intensity. A judgment produces an effect regardless, even when it comes from someone who cannot change anything in our lives—but how much it weighs on us depends on that power.

Our motivation to manage others' judgments grows in proportion to the growth of that person's power over us or over something that is valuable to us.

Why power over us explains more than the likelihood of meeting

In one of my previous works, **Presumed Judgment**, I had formulated this variable differently, attributing greater importance to the probability of future interactions.

The idea was intuitive: the more likely a relationship is to continue, the greater our interest in the other person's judgment.

That formulation, however, does not account for all cases.

Consider a deeply asymmetrical relationship, such as that between a master and a slave.

Future interactions are practically certain and very frequent. Yet the master, in general, has no reason to manage his reputation with the slave. Instead, he has an interest in maintaining it with other masters, on whom alliances, prestige, and cooperation depend.

The frequency of interaction is therefore not enough to explain the weight of judgment.

Power over us does.

A question to keep in mind

Before proceeding, it is worth noting a consequence of the above. This does not yet serve to modify behavior; rather, it serves to make a variable in the model observable.

If the weight of a judgment depends on the power that the person making it has over us, then when faced with a judgment that troubles us, there is a verifiable question:

Can this person change anything in my life?

This is not a consolatory question, nor does it ask us to downplay the issue. It calls for an assessment. In many situations, the answer is yes, and in those cases, the concern is justified. In many others—a comment from a stranger, the opinion of someone we’ll never see again, the view of someone who has no influence on our lives—the answer is no, yet the distress remains anyway, because the system produces it unconsciously and automatically.

Why the system fails to make this assessment is a question we can address only later.

Toward Presumed Judgment

We now have all the elements to address the next problem. Our evaluations are altered because the judgment of others can affect our chances of being chosen and/or of influencing our lives; but to adapt our behavior, we must first estimate what that judgment will be. Where does this estimate come from? The most intuitive hypothesis is that it is constructed by carefully observing others and gathering information about how they see us.

The following chapter will show that, in most cases, the process works differently. Our representation of others’ judgment does not coincide with either the actual judgment or our self-assessment: it constitutes a third variable—the presumed judgment—and it is this that governs much of our behavior.

4 Presumed Judgment

In the previous chapters, we saw why others' judgments can change our behavior.

An observable evaluation can itself be evaluated, and its weight depends on the power the observer has over us. If others' judgments can influence our lives, it becomes advantageous to try to predict them.

This is where the central concept of this essay comes into play.

To navigate our social relationships, we do not react directly to the actual judgment of others, which in most cases we are unaware of. We react to a mental representation of it.

We will call this representation "presumed judgment."

The concept is not new in this essay: I had introduced it in a previous essay, **Presumed Judgment**, defining it as the anticipated—generally implicit and unconscious—representation of the judgment that others might form about our current or future behavior. Those pages remain the definitive formulation; here, the concept is revisited, placed within a broader framework, and, in one respect, corrected.

A clarification of the original definition is necessary, because without it the model appears much more limited than it actually is. "Judgment" here does not mean a verbalized opinion. It means the entire expected social response: approval or disapproval, admiration or disappointment, acceptance or rejection, hostility or benevolence, willingness to cooperate or loss of trust. Most of the judgments made about us are never voiced, yet they have the same effect.

A Third Variable

It is easy to think that there are only two possibilities: the actual judgment—that is, what others actually think of us—and self-evaluation—that is, what we think of ourselves. In everyday behavior, however, a third factor comes into play.

We may consider ourselves competent and yet be convinced that others underestimate us, or have a low opinion of ourselves while at the same time believing we are highly regarded. We may be genuinely appreciated without realizing it, or feel admired while others judge us coldly.

None of these situations can be described by real judgment and self-evaluation alone. A third factor is needed: the way we believe we are evaluated. And it is this that guides most of our decisions.

The Judgment That Guides Behavior

Suppose we have to speak up during a meeting. If we were certain that others would appreciate what we have to say, we would speak with ease; if we were certain that they would judge us as incompetent, we would probably remain silent. In almost all cases, however, we have neither of these certainties: we have only a prediction. And it is that prediction—not the actual judgment—that determines whether we speak up or not.

The same applies when we decide how to dress, how competent to appear, whether to share a personal experience, whether to contradict someone, whether to ask for help, or whether to express an unpopular opinion.

Our behavior is governed not by the judgments of others, but by the judgments we attribute to others. The system reacts to our perception, not directly to reality.

Why Actual Judgment Isn't Enough

This may seem obvious, but it has an important consequence. In most cases, the actual judgment is simply unknown to us: others rarely tell us precisely what they think of us, and when they do, they may or may not be sincere, may know only part of who we are, or may change their minds over time. The mind cannot, therefore, wait for information that almost never arrives. To function, it must construct its own.

Presumed judgment arises precisely from this necessity. It is a response to the problem of uncertainty: without an estimate of how we are likely to be evaluated, every social decision would require information we do not possess.

Whether this is a good response, however, is another matter, and must be kept separate. A system can be indispensable and yet poorly calibrated. The next chapter deals precisely with this.

Presumed judgment does not coincide with actual judgment

Perceived judgment must be distinguished from actual judgment. Sometimes the two coincide; other times they diverge profoundly: one person may be highly regarded yet constantly feel inadequate, while another may provoke irritation without realizing it. Both base their behavior on perceived judgment, not on actual judgment.

It follows that changing others' judgments is not enough to change behavior. If the actual judgment changes but the perceived judgment remains unchanged, behavior tends to stay the same—a consequence that will become decisive later on.

An Observable Variable

Perceived judgment is not a theoretical construct introduced to explain just any phenomenon: it can be observed indirectly. We can ask a person how they think they are viewed, compare their response with the judgment expressed by others, and compare both assessments with the person's own self-assessment.

The three measures coincide only partially, and it is their divergence that makes the perceived judgment a distinct variable: if it always coincided with self-assessment, it would be pointless to distinguish it; if it always coincided with the actual judgment, it would suffice to measure the latter.

Research on metaperceptions has measured these three variables and shows that they overlap only partially. The results will be examined in the next chapter, because the most counterintuitive aspect of the model depends on the way in which they overlap.

The Criteria by Which We Judge Others

So far, we have considered the presumed judgment as what governs our behavior: what we say, what we show, and what we keep to ourselves.

But Chapter Two established that an evaluation, when it is observable, can itself be evaluated. And our evaluations are among the most observable behaviors we produce.

A consequence follows that the essay will revisit in the second part.

If we anticipate being judged even for the way we judge, then presumed judgment does not merely govern what we do: it helps shape the very criteria by which we evaluate things and people.

An example will make this idea clearer. Anyone who expects admiration for a certain author to be interpreted as a sign of naivety will tend to find that author less convincing than they would if that expectation were not present—not because they have decided to pretend, but because the expectation arises before the judgment is formed.

Presumed judgment thus operates in two directions: on how we behave in the presence of those who evaluate us, and on how we, in turn, evaluate others. It is not merely an adaptive mechanism that regulates what we display; it is also a constructive mechanism that intervenes in the formation of the evaluations themselves. The second part of this essay will examine this second function with the same attention as the first.

The Bridge to the Next Chapter

We have now identified the variable that mediates the relationship between the judgments of others and our behavior.

However, one crucial question remains. How is this presumed judgment constructed?

The most intuitive answer is that it derives primarily from our observation of others.

We observe their reactions, gather information, and gradually adjust our mental representation.

It's a plausible explanation, and it's the one that common sense spontaneously suggests.

The next chapter will show that things seem to work differently.

Our perception of how we believe we are judged stems, to a large extent, not from observing others, but from a much closer source: the way we see ourselves.

This is the most counterintuitive finding of the entire model.

5 Being Judged: Where Does the Presumed Judgment Come From?

In the previous chapter, we saw that behavior is not guided by others' actual judgment, but by presumed judgment: our perception of how we believe we are being evaluated.

However, one crucial question remains.

Where does this perception come from?

The most spontaneous intuition is that it derives primarily from observing others. We observe them, interpret their reactions, gather confirmations and refutations, and, little by little, construct a fairly accurate picture of how they see us.

Research, however, suggests a different answer.

The perceived judgment stems, to a large extent, not from the actual judgment of others, but from the image we have of ourselves.

An unexpected causal direction

For a long time, it was thought that the relationship worked the other way around: if others treat us with respect, we end up respecting ourselves; if they criticize us, we end up devaluing ourselves. This process exists and is well documented, especially in development and in meaningful relationships. But when an adult has to estimate how they will be judged in a specific situation, the prevailing pattern seems to be different: first, they evaluate themselves, then they use that evaluation to infer how others are likely to view them.

It's not that they stop observing others—they do observe them. But what they take away from those observations is largely predetermined, and an ambiguous expression will be interpreted in the way the person already expected.

The judgments of others are assumed rather than inferred

Research on metaperceptions compares three distinct dimensions: how a person evaluates themselves, how they think they are evaluated by others, and how others actually evaluate them.

In the review that shaped this field of study, David Kenny and Bella DePaulo (1993) reached a conclusion that seemed counterintuitive at the time: people infer how others see them not so much from the feedback they receive as from their own self-perception. Kenny reports an average correlation of approximately .87 between how a person sees themselves and how they believe they are seen—a very high value, indicating an almost complete overlap between the two representations.

In contrast, a comparison with the judgments actually expressed by others reveals much weaker correlations.

In other words, when we try to imagine how we are seen, we rely primarily on ourselves as a source of information.

This does not mean that we completely ignore what others do or say. It means that such information carries less weight than introspection would lead us to believe.

The General and the Particular

The same line of research has highlighted an asymmetry worthy of attention, because it explains something that everyday experience makes difficult to pinpoint.

Knowing how “the world in general” sees us—whether we’re perceived as extroverted, competent, or trustworthy—is something we’re fairly good at.

Knowing how a specific person sees us—and how their view of us differs from that of others—is much more difficult.

Generalized accuracy is therefore acceptable; dyadic accuracy is poor.

The system thus seems to work in the abstract and fail in concrete situations—that is, exactly where it’s needed most. The estimate we make of how a particular person will judge us, in a specific situation, is the least reliable of all—and it’s the one on which we base our decision whether to speak up, remain silent, or put ourselves out there.

The Liking Gap

A particularly instructive manifestation of this discrepancy was described by Erica Boothby and colleagues (2018), who called it the “liking gap.”

After a conversation with someone they had just met, participants tended to underestimate how much their conversation partner liked them. This was not a random error, but a systematic discrepancy in one direction only: others generally like us more than we imagine.

The phenomenon has proven to be robust and persistent, and it appears in various contexts.

If the presumed judgment were a reading of others’ judgments, such a discrepancy should not exist, or it should be randomly distributed in both directions. The fact that it is one-sided is consistent with the hypothesis that this representation is constructed elsewhere.

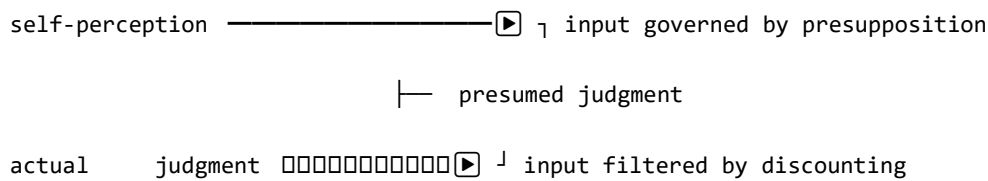
The Two Sources of Presumed Judgment

Studies on so-called meta-insight, conducted in particular by Erika Carlson and her colleagues, show that people are able to reliably distinguish between how they see themselves and how they are seen by others, and that they do not achieve a certain degree of accuracy simply by projecting others’ opinions onto themselves. They draw on additional and valid sources of information. They are also able to recognize, within certain limits, when their own assessment is more accurate and when it is less so—a capacity that is linked to the quality of their relationship with the person in question.

A recent systematic review, conducted by Harper and colleagues, shows, moreover, that meta-accuracy varies significantly depending on the relational context, the trait under consideration, and the measurement method used. Indiscriminate claims that “we don’t know how others see us” are therefore untenable.

The point, then, is not that the actual judgment lies outside the system. It is that it enters the system with a reduced and variable weight—and to understand just how reduced this weight is, we need to look at the sources of the information the system uses.

The presumed judgment does not stem from a single source. It receives information from two directions, and with very different weights.



The two arrows are not of equal weight: the external feedback exists, but it carries much less weight than self-perception. The two terms are defined a little further on.

At this point, a change in representation is in order. So far, presumed judgment has appeared as just one variable among others. From now on, it is more useful to consider it a node: it receives information from two sources, and the two classes of consequences described in the previous chapter—behavior and the criteria by which we judge—depend on it.

The two sources are self-perception and feedback from others, and the formation of the presumed judgment depends on the relative weight assigned to each. It follows that individual differences can have two distinct origins: a person may form the presumed judgment almost exclusively based on how they see themselves, or they may also use social feedback but assign it very little weight. The two cases produce similar behaviors but have different causes.

Two Parameters

To describe these two processes, we will use two parameters.

The first will be called presupposition: it indicates how much of the presumed judgment is already established before looking—that is, determined by self-perception rather than by the information available in the situation.

It does not concern whether one looks or not, but rather how much of what one believes one sees was already established. A high presupposition does not describe a distracted person: it describes a person who observes and finds what they expected to find.

The second will be called discount. It indicates the extent to which feedback from others is downplayed, reinterpreted, or simply ignored when it contradicts the presumed judgment.

The two parameters describe different aspects of the same system.

Presupposition concerns how the presumed judgment is constructed.

The discount depends on how it is maintained.

Why two parameters are needed

At first glance, a single parameter might seem sufficient.

This is not the case: there are two configurations that a single parameter cannot distinguish, and one of them forced us to correct the model. The following example illustrates this point.

In its initial formulation, the model predicted that the presumed judgment would be a function of self-perception. A very positive self-assessment should have led to an equally positive representation of others' judgment: same mechanism, opposite sign.

Research by Erika Carlson, Simine Vazire, and Thomas Oltmanns (2011) shows that this is not the case. People with high levels of narcissistic traits know that others view them less favorably than they view themselves; their metaperceptions are less distorted than their self-perceptions. In other words, they receive a relatively accurate signal—and reject it. According to the authors, one of the ways they maintain such a favorable self-image is by believing that others are incapable of recognizing their worth.

It should be noted that the literature is not unanimous on this point: alongside the position just described, there is a view that narcissists also overestimate the judgments of others, failing to adopt the perspective of their observers. The finding used here is the one with the strongest support, though not the only one available.

In social anxiety, the picture differs from both: the person tends both to construct their presumed judgment primarily based on self-perception and to devalue the positive information they receive.

A single parameter cannot describe two such different configurations. The single-variable model was incorrect and has been replaced.

Four Combinations

If we consider presupposition and discounting together, we obtain four theoretically possible combinations.

High presupposition combined with high discounting corresponds to the typical profile of social anxiety.

Low presupposition but high discount describes the functioning observed in narcissism.

The other two combinations are predictions of the model and should be interpreted as such.

A high presupposition combined with a low discount should produce an unstable self-image, easily altered by external feedback: one's sense of how one is perceived would fluctuate with every encounter.

A low presupposition and a small discount, on the other hand, represent the best-calibrated functioning, in which the presumed judgment is constructed using both self-perception and the feedback actually available. This is not a disorder: it is the case in which the system does what it seems designed to do.

These last two combinations have not yet been verified. They were formulated before consulting the literature, precisely to avoid tailoring them to the results, and remain predictions that could prove false. If presupposition and discounting were found to be closely correlated in the population—that

is, if the two mixed combinations were practically empty—then there would be only one parameter, and the distinction proposed in this chapter would have been an artifice.

.

A Practical Consequence

The distinction between presupposition and discount also changes the way we interpret change.

If the problem lies primarily in presupposition, we need to increase the quantity and quality of information coming from outside.

If, on the other hand, the problem concerns discounting, the information is already available: it is received and devalued. Adding more is useless, and in some cases makes the situation worse.

These are two different problems, requiring different approaches. We'll return to this point when we discuss recalibration.

A Bridge to the Next Chapter

So far, we have considered presumed judgment as the product of information coming from both internal and external sources.

However, there is another peculiarity of the system.

Even when no one is watching us, we often continue to behave as if someone might be judging us.

We feel shame even when we're alone.

We keep promises that no one could possibly verify.

We feel guilty even when no one will ever know what we've done.

To understand this phenomenon, we must introduce one final concept: the internalized observer.

6 The observer we carry within us

In the previous chapter, we saw that presumed judgment is constructed primarily from self-perception and only partly from feedback from others.

This conclusion, however, seems to conflict with a common experience.

Why do we continue to feel judged even when no one is watching us?

We feel shame for an action performed in complete solitude.

We keep a promise that no one could ever verify.

We feel guilty about behavior that will remain unknown forever.

If the judgment of others is the driving force behind the system, why does it continue to operate even in their absence?

The most intuitive answer is that, over the course of our lives, the external observer is gradually internalized.

An ancient idea—and what sets it apart here

The idea is not new, and it has illustrious precedents that should, however, be taken for what they are: partial, not equivalent, precedents.

Adam Smith's impartial spectator, in **The Theory of Moral Sentiments** (1759), is above all a normative device: an ideal observer against whose standard we morally evaluate ourselves.

George Herbert Mead's "generalized other" is a sociological construct: the internalization of the community's expectations.

The observer we are discussing here, on the other hand, is a predictive cognitive mechanism. The difference is not one of degree. The impartial spectator and the generalized other have a single, stable target—a moral ideal, the community—while the observer described here shifts targets: it simulates the partner, the superior, a stranger, a particular group, and moves from one to the next from one moment to the next. It is not the voice of a norm; it is the estimation of a reaction.

The crucial point, in any case, is not to determine whether an internalized observer exists. The evidence suggests that it does. The problem is understanding its role—and, above all, its significance.

A Real, but Incomplete, Observer

Behavior does not become completely selfish when no one can know what we are doing.

The most commonly used paradigm to test this is the double-blind dictator game, introduced by Elizabeth Hoffman, Kevin McCabe, and Vernon Smith in the 1990s: the procedure is designed so that not even the experimenter can trace the choice back to an individual participant. Under these

conditions, the share allocated to the other participant drops sharply compared to situations where anonymity is only apparent.

It decreases, however, without dropping to zero.

Some participants continue to give something. Outside the laboratory, many people continue to keep their promises, tell the truth, or help others even when no reputational benefit is possible.

There is, therefore, something that continues to exert pressure even in the absence of actual observers.

It is natural to interpret this as an internalized form of social judgment.

Why Talking About Conscience Isn't Enough

This conclusion, however, risks explaining too much.

Whenever a behavior continues to occur in the absence of observers, we could simply say that the internalized observer is at work.

But an explanation that can fit any result doesn't rule anything out. This is the mechanism by which some psychological theories of the past lost credibility: not because they were false, but because no outcome could contradict them.

To be useful, the concept must produce at least one prediction that could turn out to be false.

In the model proposed in this essay, the prediction is simple.

The internalized observer does not replace the external one. It represents a systematically attenuated version of it.

A Prediction of the Model

If this interpretation is correct, private behavior should never coincide perfectly with public behavior.

When it becomes impossible to accurately gauge the opinions of others, prosocial behavior should decrease, but not disappear entirely.

This is what is observed in the Dictator Game mentioned above, and what experiments on anonymity confirm with some regularity: cooperation, generosity, and honesty decrease, but none of the three disappears.

This prediction is important because it prevents the internalized observer from being used as a universal explanation.

If private behavior were indistinguishable from public behavior, the model would be falsified.

If, on the other hand, all cooperative behavior were to disappear under anonymity, it would be falsified just the same.

The model predicts an intermediate situation, and this is a prediction that can be tested.

A heuristic, not a substitute for the external observer

It remains to be determined what the nature of this attenuation is.

The system behaves as if reputation were at stake even when it is not. Not because it is deliberately calculating future consequences, but rather because it applies a shortcut—when in doubt, act as if someone were watching.

For most of their history, humans have lived in small, relatively stable groups lacking true anonymity. In such an environment, the shortcut was almost always correct, because almost always someone was actually watching, or would find out shortly afterward.

Situations of guaranteed anonymity, which are common today, are an anomaly for that system. The mental shortcut continues to work just the same.

What Is Natural and What Is Learned in the Internalization of Others' Judgments

At this point, it is worth distinguishing between two things that are often confused—and which this chapter would otherwise risk conflating.

The shortcut just described is a species-specific disposition: it does not require that a particular person have been judged in a particular way. In principle, it would operate even in someone whose social history were very different from our own.

The learned component, on the other hand, is an individual calibration: who has judged us, by what criteria, and with what severity helps determine what our internal observer notices and how demanding it is.

The two components coexist and are not in competition. The first explains why the mechanism exists in everyone; the second explains why it varies so much from person to person.

They are also, at least in principle, distinguishable: the species-specific component should show less variability across individuals and cultures than the learned component. If it showed no difference at all, the distinction would be meaningless.

Internalization does not explain everything

One significant fact remains.

The internalized observer does not exert the same force as actual judgment.

This helps explain why many people adhere more strictly to norms when they are being observed, and why reputation retains a regulatory power that no private conviction seems to match.

A useful distinction, supported by empirical evidence, is the one between shame and guilt studied by June Tangney. Shame centers on one's self-image in the eyes of others and aligns well with a reputational interpretation. Guilt centers on the harm caused, is fueled by empathy, and can be

intense even in complete secrecy. Not everything that restrains us when no one is watching can therefore be attributed to an internal observer: part of it belongs to a different realm.

The system does not function as if there were two equivalent observers.

There is a real observer and an internalized one. The latter derives from the former but does not replace it.

A connection to the previous chapter

The previous chapter showed that presumed judgment is constructed primarily from within.

This might suggest an almost entirely self-referential system.

This chapter clarifies that point.

That “inner” aspect did not arise out of nowhere. It is partly a disposition of the species, and partly the accumulation of a long history of real interactions.

The internalized observer is therefore not an alternative to society: it is, at least in its learned component, society continuing to act when others are not present.

The bridge to the next chapter

So far, we have examined the passive aspect of perceived judgment: the way we imagine ourselves being evaluated.

But the system is recursive.

We don’t just anticipate the judgments of others. We constantly judge people, ideas, works, and behaviors.

And those judgments, in turn, are observed and evaluated.

To understand the other half of the model, we must therefore shift our perspective.

No longer how we think we are judged, but how the way we judge becomes, in itself, a signal about us.

PART TWO — The Two Sides

7 To Judge Is to Expose Oneself

So far, we have examined social judgment from its passive side: the way we imagine ourselves being evaluated by others. From this point on, we must shift our perspective.

We are not introducing a new principle. It is the same mechanism described in the previous chapters that now reveals its other side.

We are not merely the objects of judgment.

We are also constantly the ones who judge.

We approve and condemn, praise and ridicule, admire and despise. It might seem like a completely different process. In reality, it is the continuation of the same system. Just as the behavior of others allows us to infer how they are likely to judge us, our own behavior also allows others to infer how we are likely to judge them.

Social judgment is recursive.

Every criterion also reveals something about the person using it

When we assert that a behavior is right, wrong, admirable, or shameful, it seems as though our judgment concerns only the object of evaluation.

In reality, it also reveals something about the person expressing it.

Every criterion made public suggests that we consider it sufficiently valid to use it to evaluate others.

Precisely for this reason, that criterion becomes potentially applicable to us as well.

This is the mechanism anticipated in the first chapter. A judgment does not remain without effect: it influences the behavior of the person who formulated it toward what they have judged and/or toward others who are judging the same subject. And precisely for this reason, it becomes evident—it need not be stated, because the way we treat someone leaves traces that others can interpret. Those who judge, explicitly or implicitly, therefore expose themselves to being judged for their own judgment.

Not because there is a moral rule that requires it.

But because the criterion itself provides insight into the person who proposes it.

Judgments are signals

This observation changes the way we interpret many forms of social behavior.

A public condemnation does not merely convey that a certain behavior is considered wrong.

It also conveys, at least in part, that the person making the condemnation wishes to be associated with the criterion expressed.

The self-presentational component of the judgment is not constant.

It increases when the judgment is public, when there are many observers or when the observers are influential, and when the issue concerns characteristics important to one's reputation.

It diminishes in private situations, but does not disappear entirely. As we saw in the previous chapter, we often continue to behave as if someone might be watching even when no one is present.

The passive and active aspects are therefore governed by the same principle.

When condemnation replaces giving

A series of studies by Jillian Jordan and David Rand (2019) illustrates this mechanism particularly clearly.

The experiments took place under anonymous, one-time conditions in which no reputation could be built. In some versions, participants had a direct way to demonstrate their trustworthiness and altruism—by giving a portion of their own money to a third party; in other versions, that option was not available.

Where it was not provided, the outrage toward those who behaved selfishly was stronger.

The two behaviors are interchangeable, and it is this interchangeability that reveals what they have in common: giving and condemning tell others the same thing about those who perform them. When one channel is closed, the other remains.

Condemnation, therefore, serves not only to sanction a behavior. It also serves to communicate something about the person who condemns—which is the thesis of this chapter, observed here in the laboratory rather than argued.

It should be added—so that the result does not promise more than it delivers—that the anonymity of these experiments is procedural rather than phenomenological: the experimenter remains a potential observer, and the extent to which participants truly believe they are not being observed cannot be directly ascertained.

Why We Detest Hypocrites

This interpretation receives particularly strong support from a study by Jillian Jordan, Roseanna Sommers, Paul Bloom, and David Rand (2017) on hypocrisy.

The intuitive idea is that hypocrites are despised because they are inconsistent.

The experiments tell a different story, and they do so through two distinct comparisons.

The first concerns the severity attributed to the behavior. Those who condemn a transgression that they then commit are judged more harshly than those who openly lie about their moral conduct, claiming not to do what they actually do. Direct lying is considered less serious than hypocritical condemnation.

The second comparison isolates the underlying mechanism. The experimenters constructed the figure of the “honest hypocrite”: someone who condemns a behavior while at the same time admitting to engaging in it. The inconsistency between what they say and what they do is identical to that of the typical hypocrite. Yet they are not judged negatively.

The difference between the two cases is not the inconsistency, which is the same.

It is the signal.

The hypocrite uses condemnation to gain the reputational benefit associated with a standard they do not uphold. The honest hypocrite, by acknowledging their own inconsistency, forgoes that benefit in the very act of condemning.

What is sanctioned is not the contradiction. It is the attempt to reap the benefits of a false signal.

It is difficult to imagine a more direct confirmation of this chapter’s thesis: if condemnation were not also a form of self-presentation, the honest hypocrite and the common hypocrite would have to be equally disagreeable.

Distinction and Belonging

Pierre Bourdieu has shown that tastes, preferences, and judgments do not merely classify the objects to which they refer.

They also classify those who express them.

Aesthetic, cultural, and moral preferences become indicators of belonging, competence, and social status.

The same applies to moral criteria.

When we judge, we do not merely describe the world. We also help to describe ourselves.

Why We Pretend to Make Judgments

If every public criterion has reputational consequences, a phenomenon studied by Timur Kuran becomes understandable.

People do not merely feign their opinions.

They may also feign their outrage, their approval, and even their silence.

A publicly expressed standard reveals something about the social identity of the person who puts it forward.

For this reason, the system creates constant pressure to align the public expression of one’s judgments with the expectations of one’s social environment, even when those judgments do not match the ones one actually holds.

The unspoken pact

There is a consequence of the foregoing that the model anticipates, and which everyone recognizes even though they have never articulated it.

If expressing a judgment exposes the person expressing it, then two people facing each other both have more to lose than to gain by judging each other openly. This leads to mutual reticence: I don't say what I think of you, and you don't say what you think of me. It is not an agreement—no one enters into it, no one remembers it, and calling it a pact is a metaphor. It is a balance: it holds because it suits both parties, and it breaks the moment it ceases to suit one of them.

Erving Goffman described this structure as “face-work.” In any interaction, each person maintains a certain self-image; others accept it, and in return, they ensure that their own self-image is accepted. Goffman calls this understanding a “working consensus,” and emphasizes a point that the model also highlights: it is not negotiated, and that is precisely why it is solid. Penelope Brown and Stephen Levinson later made this the basis of their theory of politeness, in which most of the linguistic forms we use are explained as mutual protection from exposure.

The model adds a prediction. If reticence depends on exposure, it should ease where exposure decreases: in anonymity, in highly asymmetrical relationships, and when the object of judgment is an absent third party. This is what we observe. People speak much more freely about those who are not present, and those in power voice judgments that those without power keep to themselves.

The criterion that is not adopted

The principle also holds in the negative, and the most visible case is that of shared customs—fashions, tastes, ways of speaking, and positions that are considered obligatory at a given moment. A custom, after all, is a criterion made visible through use.

If an expressed criterion says something about the person expressing it, then even a widespread criterion that one *does not* adopt says something. Those who do not follow what others follow—a fashion, a taste, a way of speaking, a stance that is considered a must at that moment—are not perceived as indifferent. They are perceived as having judged—and judged poorly—those who do follow it. Abstention is not perceived as an absence of judgment, consistent with what the first chapter established: deciding that something does not deserve attention is already a judgment.

This phenomenon has been measured. Julia Minson and Benoît Monin have shown that meat-eaters evaluate vegetarians more negatively the more the expectation of being morally judged by them is emphasized. This devaluation does not target the dietary choice itself; rather, it targets the judgment attributed to it. The authors call this “*do-gooder derogation*”—the devaluation of those who do the right thing—and identify its function: to preemptively defuse an expected rebuke.

This explains an otherwise curious phenomenon: those who adopt a more demanding standard than the prevailing one are sanctioned more than those who adopt a more lenient one, even though they have done nothing to anyone. Within the framework of the model, this is not curious at all. A standard declared to be valid also measures those who do not apply it, and those who feel measured react to a judgment that no one has expressed: the purest case of presumed judgment governing behavior.

It also follows that the pressure to conform does not arise solely from the desire to resemble someone else. It arises from the cost of not doing so—which is not the cost of being different, but that of being perceived as judgmental.

Moral Ostentation

There is a class of behaviors in which the communication of one's virtue seems to take precedence over the actual contribution to the cause being supported. In public debate, these behaviors are often referred to by the journalistic term "*virtue signaling*," which has been in circulation since the mid-2010s and is almost always used in an accusatory sense.

The corresponding concept in moral philosophy was developed and analyzed by Justin Tosi and Brandon Warmke under the name "*moral grandstanding*": the use of public moral discourse to impress others with one's moral qualities.

The phenomenon exists.

Within the framework of the model developed in this essay, however, it represents a special case of a much more general process.

Every public judgment also has a reputational component, the intensity of which varies with the context.

Moral grandstanding therefore does not introduce a new mechanism. It is the case in which that component becomes sufficiently evident to be noticed—and, for this reason, the case in which the signal ceases to function.

An Open Question

The research discussed in this chapter already shows that the sanction for hypocrisy depends on the falsification of the signal and not on mere inconsistency. On this point, we are not making a prediction: we are reporting a result.

However, there remains a consequence of the model that has not yet been verified.

If public judgment also functions as a reputational signal, it is reasonable to expect that the cost of violation will increase as the scope within which the criterion has been declared valid expands.

Those who propose a principle as valid for everyone should be judged more severely when they violate it than those who present it as a personal preference or as a norm limited to certain circumstances.

The Bridge to the Next Chapter

At this point, the reader might draw an unpleasant conclusion.

If our public judgments function as signals, it would seem that social life consists of an extensive performance, and that behind every stated criterion lies a calculation.

The next chapter will show that things are even stranger than that.

We aren't acting—or at least not primarily.

When we tailor a judgment to our audience, that judgment tends not to remain a mere pose. Convenience does not merely produce hypocrisy; it often also alters what we end up sincerely believing.

It is this shift that makes the model more unsettling—and at the same time less cynical—than this chapter might suggest.

8 Expressing a judgment changes us

The previous chapter might leave the reader with an unpleasant impression.

If our public judgments also function as signals of reputation, it would seem that much of social life is a conscious performance. We would say what is expedient to say, display the most advantageous opinions, and condemn whatever produces the best effect on others.

This impression is understandable—and partly justified. Human beings lie, feign, and pretend; hypocrisy exists, and the previous chapter described a particular form of it.

But if most of our judgments were the result of deliberate calculation, we would have to constantly remind ourselves which opinions are genuine and which are merely put on display. That does not seem to be the case.

In most cases, we are not playing a role.

We end up believing what we have said.

The Message Tailored to the Listener

This phenomenon has been studied in the laboratory using a simple paradigm.

A participant is given an ambiguous description of a person: some behaviors lend themselves to a favorable interpretation, others to an unfavorable one. The participant is then asked to describe that person to an interlocutor whose opinion they know. As expected, the participant adjusts their description: making it more favorable if the interlocutor is well-disposed, and more critical if the interlocutor is hostile.

So far, there is nothing surprising here. It is courtesy, or communicative adaptation.

The interesting result comes later.

When, at a later stage, the participant is asked what they really think of that person—and what they remember of the original description—their judgment and memory have shifted in the direction of the message they had formulated.

They didn't just speak differently—they began to think differently.

Communication, in other words, does not merely convey pre-existing beliefs. It helps to construct them. Every time we tailor a judgment to our listener, we also slightly alter the representation we ourselves will retain.

This effect has been called “*saying-is-believing*,” and was first described by Tory Higgins and colleagues in the late 1970s. Subsequent research has linked it to the need to construct a shared representation of reality with the interlocutor: modulation is not a superficial adaptation, but the way in which a common understanding is established—and what is established together then tends to be what is remembered and believed.

Three Related Theories

This phenomenon is sometimes confused with three other, better-known phenomena, from which it is important to distinguish it.

Leon Festinger's theory of cognitive dissonance describes the discomfort caused by the contradiction between what we do and what we think, and the tendency to reduce it by modifying one or the other.

Daryl Bem's self-perception theory proposes that in many circumstances we infer our attitudes by observing our own behavior, much as an outside observer would.

Fritz Heider's equilibrium theory describes the pressure to maintain consistent relationships within a triad: if I hold someone in high regard and discover that they evaluate something that concerns us both differently than I do, one of the two judgments tends to yield.

The three theories share a common feature: they describe a pressure that arises afterward, once a discrepancy has already emerged.

The effect described here occurs earlier, and involves different material. It does not concern behavior in general, but rather communication directed at someone: what changes one's conviction is not having acted, but having formulated a judgment for a recipient whose position was known. The discrepancy has not yet arisen—it is avoided.

This is the social case, and it is precisely the case addressed in this essay.

From Adaptation to Conviction

This mechanism helps us understand many seemingly spontaneous changes.

A person enters a new environment—a workplace, a group, a professional community.

At first, they adopt the prevailing language, criteria, and judgments primarily to facilitate communication: these are what the other person expects, and expressing them differently would require constant explanations and cause friction.

After a few years, they may feel as though they have always thought that way.

The change is not necessarily a successful pretense. In most cases, there was never any pretense to begin with: the gap between what was said and what was thought gradually narrowed, without anyone noticing.

This is also why a conversion of this kind is almost impossible to recognize from the inside. There is no single moment when one decided to change one's opinion.

The Case of Fashions

This process can be observed on a large scale in fashions—understood in the broadest sense: not only clothing, but musical tastes, reading habits, the positions one is expected to hold, the words one uses, and those one no longer uses.

Georg Simmel provided a description of this at the beginning of the 20th century that still holds true today: fashion thrives on the tension between two opposing forces—imitating those one wishes to emulate and distinguishing oneself from those one has left behind. It is a description of collective dynamics. The individual mechanism is the subject of this chapter.

Those who conform to a widespread custom do not remain in the position of those who are pretending. Adapting to the audience alters what one ends up believing, and the audience, in this case, is a widespread and anonymous public of whom only the average expectations are known. It follows that adherence to a fashion is almost never hypocrisy: those who find beautiful what is currently supposed to be beautiful truly do find it beautiful.

And this is also evident because, in hindsight, a change in fashion seems incomprehensible to those who had followed it. We haven't stopped pretending. We've stopped believing—along with the audience that was listening to us.

The Same Flaw on Both Sides

At this point, we can see the symmetry that underpins the entire second part of the essay.

In Chapter Five, we saw that we do not know precisely how others see us: the image we have of them is constructed largely from our own perspective.

Now we see that we do not even know precisely why we think what we think: part of our judgments was formed by adapting to those who were listening to us, and bears no trace of that origin.

Both sides share the same flaw, and they have it for the same reason.

In both cases, the mechanism that is supposed to detect this flaw is the very same one that is affected by it. There is no external vantage point within the system from which to verify either how we are perceived or why we believe what we believe.

This is the most precise formulation of the thesis stated in the introduction: a recursive process that operates on input that is not observed but inferred.

If this interpretation is correct, however, there is a further consequence. It no longer concerns the judgments we form, but those that do not even come to be formed.

What We Are Willing to Consider

So far we have spoken of formulated judgments: something we say, and which we then end up believing.

But if the system anticipates the relational consequences of what we express, there is no reason why it should be activated only at the moment of expression. It could intervene earlier, when an idea arises and must be examined.

Even before we ask ourselves, “Is it true?”, the mind may have already answered another question: “What would taking this idea seriously mean for my social life?”

Would it bring me closer to—or drive me further away from—the people who matter to me? How would it look, in the eyes of those who know me, that I’m even considering it?

This would explain why certain ideas seem unacceptable before they have even been examined—rejected not by an argument but by a resistance that precedes the argument. And why a thesis might be ignored not because it is deemed false, but because taking it seriously would alter the position of whoever does so.

We can also reinterpret Pierre Bourdieu’s concept of *habitus*—which we encountered in the previous chapter—in this light: embodied dispositions that reflect and reproduce the social position of those who possess them. There, it served to explain that our judgments classify us; here, it suggests something more, namely, that they may preemptively select what we come to consider. The difference from Bourdieu is that this would not merely be a deeply ingrained disposition, but an active, moment-by-moment predictive filter that operates on new ideas just as it does on behaviors.

It must be clearly stated that this extension has not been proven. It is a plausible consequence of the model, and as such may be false.

However, one clarification is in order, without which the hypothesis would become a weapon. If it is correct, it does not concern only the ideas of others. It also concerns these pages, and the way the reader is evaluating them at this very moment—and it concerns those who wrote them, who have no privileged vantage point from which to observe their own work.

Convergence and Divergence

From the foregoing follows something that concerns relationships between people, and not just each person’s judgments.

Two people who interact with some frequency adapt their judgments to one another, and as we have seen in this chapter, this adaptation is not merely a pretense: it becomes a conviction.

The convergence of judgments is therefore not merely a condition of cooperation. It is also a product of it, and continues to occur as long as the cooperation lasts.

It is not a side effect. Those who judge as we do are predictable, and predictability is what allows us to commit before knowing how things will turn out—the capacity that Chapter 3 attributed to judgment.

Divergence, on the other hand, is not interpreted as a difference. The previous chapter established that failing to adopt a widespread criterion is interpreted as a judgment on those who do adopt it, and that those who feel judged react to a judgment that no one has actually expressed. It follows that divergent judgments carry a relational cost even when there is nothing at stake: a conflict of interest is not necessary to generate hostility; divergence alone suffices.

And the relationship works in the opposite direction as well. Adopting an opponent’s criterion would mean validating it; if the extension proposed above is correct—that the relational cost comes into play before the idea is examined—taking what the other person argues seriously becomes costly regardless of how well-founded it is. Conflict produces divergence just as divergence produces conflict.

A testable prediction follows. Among those who cooperate, judgments should converge beyond what the shared information justifies; among those in conflict, they should diverge beyond what the information justifies. It is not convergence itself that is predicted—for that, it would suffice for two people to have the same data—but its excess relative to the available evidence.

This field has long been studied under terms such as group polarization and motivated reasoning, and these pages have not reviewed the literature on the subject. What precedes is a consequence of the model, not a contribution to those discussions, and it is possible that it is already covered there, in a better form.

Why the system works this way

At first glance, it seems like a design flaw. A system that modifies its own beliefs based on the interlocutor appears unreliable.

However, one can formulate a hypothesis as to why it is advantageous—provided it remains explicitly stated as a hypothesis.

Maintaining two parallel versions of every judgment—the private one and the one presented to others—carries a cognitive cost and a risk: the discrepancy may emerge, and when it does, the reputational damage is that of the hypocrite described in the previous chapter—that is, the most severe. Gradually reducing the gap between the two versions eliminates both the cost and the risk.

Once formed, a conviction requires no monitoring. Pretending does.

It must be noted that this explanation is plausible but unproven: the fact that a trait is advantageous today does not prove that it was selected for that advantage.

An important limitation

Nothing in the preceding implies that all beliefs arise from social adaptation.

Many stem from direct experience, reasoning, learning, or individual dispositions.

The model's claim is more limited: when a judgment is repeatedly expressed to interlocutors whose positions are known, the probability increases that it will shift permanently in that direction.

How strong this effect is, and under what conditions it prevails over other processes, remains an empirical question.

A Bridge to the Next Chapter

The two preceding sections have described a system finely tuned to a specific task: predicting the judgment of those who observe us and adjusting our behavior accordingly.

However, one difficulty remains.

That task made sense in an environment very different from the one in which we live today. The system evolved in small groups, where the observer was also the decision-maker, and where every judgment was shared in full.

Today we find ourselves exposed to observers who will never decide anything concerning us, and to judgments that will never reach anyone.

The next chapter examines what happens to a mechanism calibrated for a world that no longer exists.

9 **A Mechanism Calibrated for a World That No Longer Exists**

In the previous chapters, we have described a surprisingly sophisticated system. On the one hand, it reconstructs how others likely see us; on the other, it shapes our judgments based on the consequences they will have. Both aspects converge on the same task: predicting the social value of our actions and adjusting them accordingly.

If this system had been designed for the world we live in today, however, it would have to behave very differently.

Every day we meet people we'll never see again. We receive anonymous evaluations. We read comments from strangers. We're observed by individuals who have no power to make decisions that affect us. Under these conditions, the judgment of many observers should be almost irrelevant.

Yet it is not.

A single glance, a laugh, an online comment, or criticism from a stranger can be enough to provoke shame, embarrassment, or irritation. The system reacts as if that judgment had significant consequences, even when we know full well that, in all likelihood, it will have none.

The problem isn't explaining why judgment matters. The previous chapters have already done that.

The problem is explaining why it continues to matter when it shouldn't.

An Unresolved Question

This difficulty had already emerged, implicitly, in the chapter on the weight of judgment.

There, we established that the weight of a judgment depends on the power that the person making it has over us, and we derived a question from this: Can this person change anything in my life? We also observed that, in most cases, the answer is no—and that the distress remains nonetheless, because the system that produces it does not make that determination on its own.

The question of why remained unanswered.

It is time to answer, because the answer explains both the mechanism's inadequacy and the limitation of the practical tool that the essay will propose.

Judgment Is Not the Final Measure

The answer proposed by this essay is that judgment is not the true objective of the system.

The ultimate measure is relational consequences.

Being chosen or excluded. Gaining cooperation or facing rejection. Preserving or losing future opportunities.

Judgment is merely the signal that, in most situations in which the system has developed, has made it possible to predict such consequences.

The system therefore does not assign value to judgment in and of itself. It uses it as a quick indicator of the consequences that normally follow it.

Just as a thermometer is not a fever but a quick way to estimate it, so judgment is not social harm but the signal that typically anticipates it.

It is the role that, in a forecasting system, is called *a predictor*: a quick signal to detect, which stands in for an outcome that is slow to manifest or difficult to measure. The judgment of others, for this system, is a predictor of relational consequences.

For most of human history, this shortcut probably worked well.

In the small groups where humans lived for almost their entire past, almost everyone knew one another, almost everyone would meet again, and almost every judgment would have concrete consequences. Someone who criticized another person today could decide tomorrow whether to help them, avoid them, choose them as an ally, or exclude them from cooperation.

In such an environment, judgment and consequences almost always coincided. Using the former as an indicator of the latter was advantageous.

It must be noted that this reconstruction is a plausible inference about the conditions under which the mechanism arose, not an observed fact. The fact that a trait can be explained in these terms today does not prove that it was selected for this reason.

When the predictor continues to function on its own

The problem arises when the indicator survives the conditions that made it reliable.

Today, we are constantly evaluated by people who will never have the opportunity to influence our lives. The link between judgment and consequences has weakened significantly.

The mechanism, however, continues to treat judgment as if it were still a good predictor.

That's why we might blush in front of a tourist we'll never see again, worry about a comment from an anonymous user, or dwell for days on an embarrassing moment that happened in front of people whose faces we don't even remember.

From a rational standpoint, this behavior seems disproportionate. From the system's standpoint, it is not.

The system isn't evaluating the specific instance. It's reacting to a signal that, for most of its history, has been statistically reliable.

Emotions come before verification.

What about the power others have over us?

At this point, the attentive reader might notice a contradiction.

The chapter on social value argued that the weight of a judgment depends on the power that the person making it has over us. This chapter argues that we also react to judgments made by people who have no power over us. The two statements seem incompatible.

They are not, because they concern two different stages of the same process.

Real power determines how much a judgment truly matters: it determines the consequences, and over the long term, it effectively shapes behavior—we pay more attention to our image in front of those who can change something in our lives, and the system registers this.

The predictor, on the other hand, explains why the immediate reaction does not perform that assessment. At the moment the judgment is perceived, the system does not calculate the power of the person expressing it: it reacts to the signal.

Real power, therefore, describes what the system is calibrated to; the predictor explains why it does not check this on a case-by-case basis.

It is also the reason why the question posed in Chapter Three does not dispel the unease. Establishing that that person cannot decide anything about us corrects the assessment, not the reaction—which has already taken place.

A New World, an Ancient Mechanism

This interpretation helps us understand why the mechanism often seems excessive.

It is not excessive relative to the environment in which it was formed. It is excessive relative to the environment in which it continues to operate.

We live immersed in a multitude of observers that no human being could have encountered throughout nearly the entire history of the species. Social networks, media, large cities, and mobility have multiplied the number of gazes without increasing, to the same extent, the consequences those gazes produce.

The system, however, continues to treat many of us as if we still belonged to a small, stable community. It does not spontaneously distinguish between the judgment of a friend, that of a future employer, and that of a passerby encountered for just a few seconds.

Not because it makes a miscalculation: because it doesn't perform the calculation at all. The system applies a heuristic—it treats judgment as an indicator of social consequences, because for most of its history it has been—and heuristics work this way, replacing a difficult problem with a simpler one. Speed increases; accuracy decreases when the environment changes.

The distinction between a friend and a passerby can be reconstructed through reflection, but it comes after the automatic response, not before.

The Bridge to Part Three

So far, we have tried to understand how the system works. We now know its structure, strengths, and limitations.

One question remains, however.

If our presumed judgment is largely constructed from within, if our own judgments end up shaping what we believe, and if the system continues to react to signals that in today's world have often lost their original meaning, is it possible to modify its functioning, at least in part?

Part Three addresses this question. It does not propose eliminating the mechanism, but rather understanding which interventions truly succeed in recalibrating it and which, on the other hand, end up reinforcing it.

PART THREE — What to Do With It

10 Why Intuitive Remedies Make Things Worse

After describing how the system works, one question becomes inevitable.

If our perceived judgment stems largely from self-perception rather than from the observations of others, and if the system continues to react to signals that, in today's world, have often lost their original meaning, how can it be corrected?

The spontaneous answer seems obvious.

Observe what's happening more closely. Reflect more deeply. Monitor yourself more carefully. Ask others how they really see you.

These are all reasonable strategies.

And that is precisely the problem.

Almost all of them impair the functioning of the system they are meant to correct.

The Paradox of Self-Observation

When we want to understand a phenomenon, the natural solution is to observe it. If we want to know how an engine works, we take it apart. If we want to understand an animal's behavior, we observe it.

It's natural to think that the mind should work the same way. Observe ourselves more closely. Monitor our feelings more carefully. Pay closer attention to our own reactions.

The problem is that, in the case of presumed judgment, the observer is the very thing that is supposed to be observed.

There is no external point of view. Any attempt to control the system alters the system's very functioning.

This difficulty even has a name. Research on self-attention has highlighted what is called the self-absorption paradox: those who observe themselves more tend to experience both greater self-awareness and greater psychological distress. The two grow together, and it is not possible to achieve the former while avoiding the latter simply by observing oneself even more.

The more we focus on our own sensations during a social interaction, the less information we gather about the interaction itself.

This isn't just about people who suffer from anxiety

Before proceeding, it's worth clearing up a misunderstanding that the rest of this chapter might otherwise give rise to.

One might think that what follows applies only to particularly anxious people, and that those who aren't anxious can consider themselves exempt.

This is not the case, and research has shown this long before the phenomenon was studied in a clinical setting.

Shelley Duval and Robert Wicklund, with their theory of objective self-awareness (1972), demonstrated that when attention shifts to oneself as an object—in front of a mirror, a camera, or an audience—the gap between how one is and how one believes one ought to be becomes salient, and behavior shifts to reduce that gap: toward what that person believes is right, not toward what is actually right. Self-focus does not make people better; it makes them more consistent with themselves.

This effect does not apply to a specific category of people; it concerns a condition in which anyone can find themselves.

Allen Fenigstein and colleagues (1975) subsequently identified public self-consciousness as a stable trait: the disposition—which varies from individual to individual—to represent oneself as the object of others' gaze. Not a line separating the healthy from the disturbed, but a continuum along which each person is situated.

Taken together, these two studies shed light on the model presented in this essay: the situational conditions and individual differences that increase the activity of the presumed judgment system have already been observed in the laboratory, in ordinary populations.

What appears in an extreme form in social anxiety is therefore the same mechanism that operates, albeit more weakly, in those who are not anxious. This is the riskiest prediction of this essay: it argues that the presumed judgment system does not belong to social anxiety but to ordinary psychology, and that social anxiety is a particularly intense form of it. If it were discovered that non-anxious individuals use different processes rather than the same ones in a less intense form, this generalization would fall apart.

When Attention Shifts to the Wrong Place

In David Clark and Adrian Wells' (1995) model of social anxiety, this phenomenon is called self-focused attention.

Upon entering a social situation, the person gradually stops observing the environment and focuses their attention on themselves.

They monitor the tone of their voice. They monitor their blushing. They monitor their posture. They monitor their words. They even monitor whether they're monitoring too much.

The effect is paradoxical. The more they look for evidence of how others are evaluating them, the less data they gather about others' actual behavior.

Information from the outside world decreases. Information from within increases.

Physical sensations, anxiety, and the person's self-image thus end up replacing the observation of others and are taken as evidence of how they are appearing.

At this point, a connection emerges with the first part of the essay.

The “elevated presupposition” described in Chapter 5 and the “self-focused attention” described by Clark and Wells can be interpreted as two descriptions of the same phenomenon stemming from traditions that do not cite one another.

In the first case, the problem is described as a property of the way in which the presumed judgment is constructed. In the second, it is described as a shift in attention that prevents the person from gathering the information necessary to correct it.

The two descriptions are not mutually exclusive. They complement each other.

Two Different Ways of Failing

It is now useful to employ the distinction introduced in Chapter 5, as it allows us to see that intuitive remedies do not all fail in the same way.

The presumed judgment receives two inputs, and each has its own parameter: the presupposition, which determines the extent to which it is constructed based on self-perception rather than observation; and the discount, which determines the extent to which external evidence is devalued.

Some remedies fail because they act on the wrong parameter. Others fail because they actually worsen the very thing they are meant to correct.

The following sections examine the three most common remedies in light of this distinction.

Why Thinking More Isn’t Enough

Many people try to solve the problem by thinking about it even more.

They mentally replay a conversation. They analyze every sentence. They reconstruct every detail. They try to figure out where they went wrong.

This activity is called reflection, but it encompasses a wide variety of processes.

Anthony Grant and his colleagues have distinguished between two concepts that are often conflated in everyday language: reflection—that is, the tendency to examine one’s own thoughts and behaviors—and insight—that is, the clarity with which one understands them. The two measures correlate with each other much less than one might expect, and only the latter is associated with well-being, lower levels of anxiety, and better self-esteem. It is possible to reflect extensively without gaining any insight.

In the case of presumed judgment, this distinction is particularly important.

Continuing to think about the problem does not automatically increase the likelihood of correcting it, for two distinct reasons.

The first is that reflection works on internal material, and internal material is precisely what Chapter 5 showed to be the source of the problem. Rethinking a conversation does not introduce any new information about others: it reworks what the system had already produced. In terms of the two parameters, it does not lower the presupposition. It increases it.

The second reason is less obvious and stems from an observation I have developed elsewhere. The system does not communicate its outcomes in the form of thoughts, but in the form of emotions: an urge to act when the prediction is favorable, and resistance when it is unfavorable. What reaches consciousness is the result, not the calculation. If this is the case, reflection probes a level at which the response is not to be found—and the explanations it produces are plausible reconstructions in hindsight, not an account of what actually happened.

This does not make reflection useless. It makes it pointless to expect that reflection alone will bring about change.

Why reassurances don't work

A second intuitive strategy is to seek reassurance.

“Do you think I was convincing?” “Do you think I made a bad impression?” “How did I come across to you?”

The idea is simple: if the system has incorrect information, just provide it with the correct information.

The problem is that the system isn't lacking information. It devalues it.

Here, the factor at play isn't the presupposition but the discounting. Reassurance provides exactly what seems to be missing—an external, explicit piece of information from the right person—and fails to produce the expected effect because the information is received and downplayed.

“He's just being nice.” “He doesn't want to hurt me.” “He didn't see what really happened.”

The information arrives. It does not alter the system.

This is the most instructive of the three cases, because it shows that the problem isn't always the amount of information available. Adding more information, when the critical factor is discounting, is useless—and repeatedly seeking confirmation tends, for reasons we'll see, to reinforce the very habit of downplaying it.

Why Self-Monitoring Worsens Performance

Finally, there is an even more intuitive strategy: exercising greater self-control.

Prepare every sentence. Continuously monitor your own behavior. Eliminate every possible error.

In the short term, this may even seem effective. In the long term, it often produces the opposite effect.

Every check increases the amount of attention directed toward oneself and reduces the attention available for the environment.

The system loses precisely the data it needs to update itself. In terms of the two parameters: self-monitoring raises the assumption, because it deprives the system of the only input that could correct it.

It's a vicious cycle. The more you try to check how you're being evaluated, the less you observe how others actually react. The more you rely on your own internal sensations, the more the presumed judgment continues to be constructed from within.

It is worth noting that this is also why protective behaviors—preparing what to say, avoiding pauses, holding a glass to hide a tremor—are among the primary targets of social anxiety therapy. Not because they are ineffective in the moment, but because they prevent us from discovering what would have happened without them.

A Preliminary Conclusion

Intuitive remedies share a common characteristic.

They seek to improve the system by increasing monitoring. But monitoring, beyond a certain threshold, is precisely what sustains it.

Viewed through these two parameters, however, they do not all fail in the same way. Reflecting and self-control reinforce the presupposition: they deprive the system of external information and force it to rely even more on itself. Seeking reassurance does not affect the presupposition—external information does indeed provide it—but it runs up against the discount.

The distinction is not merely academic. It indicates that there are at least two useful interventions, and they are distinct from one another: one that restores real information to the system, and one that prevents its devaluation.

It is worth combining this result with that of the previous chapter, because the two limitations add up.

There, we saw that rationally ascertaining that a certain person cannot decide anything about us corrects the evaluation, but does not prevent the reaction: by the time the ascertainment occurs, the reaction has already taken place. Here, a second limit is added: many of the spontaneous attempts to correct that reaction end up fueling the mechanism that produces it.

The reason the next chapter is necessary lies in this combination. It is not enough to know how the system works, and it is not enough to strive to correct it using the means that come naturally.

This does not mean that the system is unchangeable. It means that the direction of change is not the intuitive one.

This is the subject of the next chapter.

11 The Principle of Recalibration

At the end of the previous chapter, we reached a conclusion that may seem discouraging.

Presumed judgment continues to influence us even when we know that others' judgments will have no real consequences. And our most spontaneous attempts to correct it often end up reinforcing it.

It might seem as though the system is simply unchangeable.

That is not the conclusion.

The system does not change through the reasoning we use to try to convince it. It changes when the predictions it uses change.

A Learning System

It's time to give a name to the system whose workings we've been following. In my previous works, I've called it *the social supervisor*; in these pages, I've preferred to describe it before naming it, because a label introduced too early takes the place of the description. From here on, the name is necessary, because Part Three deals with how to modify it.

Let's call it the system that has been described piece by piece in these pages: the one that constructs the presumed judgment by drawing primarily on self-perception, that continues to operate in a subdued form when no one is watching us, and that conveys its outcomes—not the calculation that produces them—to our consciousness in the form of emotional impulses and resistances. It is a forecasting system.

Its task is not to say whether something is right or wrong. It is to anticipate the relational consequences of our actions.

Like any predictive system, it must compare its expectations with what actually happens. If it did not do so, it could not adapt to a changing environment.

The social supervisor is therefore not rigid. It is flexible.

But its plasticity follows rules different from those of conscious reflection.

Words explain; experience brings things up to date

If someone tells us, "You shouldn't worry about what others think," we fully understand the meaning of the sentence. We may even agree with it.

Yet the system continues to react as before.

This isn't because it's irrational. It's because that statement isn't evidence. It's information.

The social supervisor revises its predictions especially when experience diverges significantly from what it expected.

It's not enough to know a conclusion. The prediction must be disproved.

The Prediction Error

Many contemporary models of learning, though diverse and not traceable to a single theory, share the idea that predictions are updated primarily when experience produces an unexpected result. In what follows, I will adopt this principle as an interpretive framework, without assuming that it fully describes every form of learning.

What follows does not, after all, depend on any particular formalization: it merely requires that the system update its expectations more readily in the face of a refutation than in the face of a confirmation.

A system formulates an expectation. Experience reveals a different outcome. The discrepancy triggers an update.

The principle is surprisingly simple. No system modifies its predictions when everything confirms what it already believes; it modifies them when reality produces a surprise.

For the social supervisor, this means that change requires situations in which the presumed judgment is actually put to the test.

Not imagined. Not discussed. Experienced.

It follows that any attempt to modify the social supervisor makes sense only if it creates situations in which one of its predictions can be compared with what actually happens. If the two coincide, the system has no reason to change; if they diverge, the information it needed becomes apparent.

Why Simply Exposing Oneself Is Not Enough

This idea helps clarify a common misunderstanding.

We often hear that it's enough to simply expose oneself to feared situations.

Exposure, however, does not in itself guarantee any change.

Consider a person convinced that, if they were to speak up in a meeting, others would find them incompetent. They speak, and for the entire duration of their remarks, they focus on their own voice, their own hands, their own hesitation. In the end, they remember above all their own discomfort; they do not remember the faces of those who were listening, and they conclude that it went badly.

They went through the feared situation without gathering a single new piece of information. They simply accumulated another experience consistent with their previous model.

For exposure to lead to learning, the social supervisor must be able to compare a sufficiently precise prediction with an observable outcome.

It is not simply enduring the situation that brings about change. It is the comparison between what was expected and what actually happened.

Two Necessary Conditions

Two conditions emerge from the above, and they correspond to the two parameters introduced in Chapter Five—presupposition and discounting.

The first condition concerns presupposition: the outcome must be observed. If attention remains inwardly focused, the system will continue to feed on its own sensations, and no experience will be able to disprove it because no experience will have actually been recorded.

The second condition concerns the discount: the prediction must be formulated in advance and in a recognizable form. If it remains vague, after the experience it will always be possible to reinterpret what happened—“it went well because it was an easy situation,” “they were kind”—and the refutation will be absorbed without leaving a trace.

This is why, in the treatment of social anxiety, behavioral experiments require participants to write down their prediction beforehand. This is not a mere formality: a prediction put in writing can no longer be rewritten after the fact.

We can now state the principle in its simplest form, before adding the clarifications it requires: the social supervisor does not change because we convince them, but because the experience refutes their predictions in a verifiable way.

Everything that follows serves to clarify this statement in three ways: what is required for a refutation to be recorded, where the most effective refutations occur, and in which direction the system updates itself most easily.

What Is Required for a Refutation to Be Recorded

There is, however, one aspect that the concept of prediction error overlooks, which I have addressed elsewhere in my discussion of attention.

For a prediction to be refuted, one must first realize that it is at work. And recognition, on its own, does not bring about change.

What recognition produces is a brief window: an interval in which the habitual response has not yet kicked in and a different response is still possible.

For that window to be used, one must have already decided beforehand what to do with it. Recognition without a ready alternative produces observers who are aware of their own automatisms, not people who modify them.

This is also why purely intellectual insight often changes nothing: one can know everything about one’s own mechanisms and still keep repeating them, because knowing is not the same as having a different response ready when it’s needed.

From this follows a distinction that concerns precisely the moment of recognition, and one that is worth making explicit, because two seemingly similar things pull in opposite directions.

The first is to keep the issue alive: to continue observing one's own behavior from different perspectives, preventing it from slipping into irrelevance. Here, variety is valuable, because attention fades in the face of identical repetition.

The second is recognizing the mechanism the moment it kicks in. Here, variety is harmful, and the opposite is needed: specific, consistent cues anchored to concrete circumstances that precede or accompany the reaction—a physical sensation, a recurring situation, or a type of thought.

Research on prospective memory—the ability to remember to do something at the right time—indicates that situational recognition works best when the association between cue and response is repeated exactly the same way, not when it is varied.

Confusing these two functions is why many resolutions fail. A reminder read in the morning arrives when the mechanism is not active; when it does activate, in the heat of the moment, the reminder will not be present.

In this context, consciousness does not update the system: it designs the conditions under which the update can occur, and decides in advance what to do with the window of opportunity that recognition opens.

Where the Most Effective Counterarguments Come In

So far, we've discussed deliberate interventions. However, it would be a mistake to conclude that they are the most effective.

They probably aren't.

The most profound recalibrations seem to occur through channels that no one has planned: a change in social status, entering an environment with different standards, aging, or a stable relationship that lasts long enough to produce thousands of small contradictions.

We also have empirical evidence supporting this last point. In Chapter 5, we saw that the ability to recognize when one's own assessment is accurate is linked to the quality of the relationship with the person in question.

The reason is plausible. In a close, long-term relationship, feedback does not come in a single, rewritable instance, but rather across a multitude of instances that would be difficult to reinterpret in the same way. The discount works well for a single episode. It works poorly over five years.

This has an inconvenient consequence for anyone seeking quick fixes: the factor with the strongest empirical support is not a technique, and it is largely beyond the control of those who would like to apply it.

In which direction does the system adapt most easily?

Finally, there is one aspect that discussions focused on change tend to leave out, and it must be mentioned because otherwise the picture would be incomplete.

The social supervisor does not recalibrate only for the better.

Public humiliation, betrayal, or a prolonged period of hostility in a workplace produce rapid and stable adjustments—in the opposite direction.

And they do so through the same mechanisms described in this chapter: these are experiences that disprove a prediction—the prediction of being safe, of being appreciated, of being able to count on someone—with a precision that no exercise can match.

It is worth noting the asymmetry. The disconfirmation in a negative direction usually comes in a single, intense, and unambiguous episode. The disconfirmation in a positive direction almost always comes in many small and easily reinterpretable instances.

The question remains as to why the two forms of refutation are so different, and the answer was already given two chapters ago. Since expressing a judgment exposes the person making it, a mutual reticence prevails between two people facing each other: a favorable judgment is almost never stated outright, but rather comes in the form of politeness. But courtesy is precisely what the system defends against—“he’s saying that out of kindness,” “he doesn’t want to hurt me”—so that positive information is already taken for granted, because social convention renders its source ambiguous.

Public humiliation, on the other hand, is a breach of that reticence: someone has paid the price of exposing themselves to say what they think, and for this reason the signal is clear and lends itself poorly to any other interpretation. In short: no one tells us clearly when things are going well, while we’re told clearly when they’re going badly.

It is therefore not true that the system is reluctant to adapt. It is reluctant to adapt for the better.

A criterion, not a method

The principle of recalibration can be summed up in a single sentence.

The social supervisor changes less through the explanations it receives than through the forecasting errors it experiences.

This statement does not yet constitute a method. It does, however, define the criterion by which to evaluate any proposal for change.

Any corrective intervention that does not produce a comparison between the system’s prediction and actual experience is unlikely to alter its functioning. Any corrective intervention that succeeds in producing that comparison becomes, at least in principle, a candidate for recalibration.

Translated into operational terms, the criterion breaks down into two questions that can be applied to any proposal for change, including one’s own.

These are not the two questions the essay poses to the reader—which are addressed in the next chapter: those questions are directed at oneself when making a judgment, while these are directed at a method that is being evaluated.

- Does the intervention increase the amount of external information reaching the social supervisor?
- Does the intervention reduce the supervisor’s tendency to devalue that information?

It is worth testing this on cases the reader already has at hand. The three remedies from the previous chapter are all interventions, and the grid below shows the verdicts already reached.

Reflecting further does not add external information—it reworks internal material—and does not reduce the discount: it answers “no” to both questions. Self-control makes things worse, because it subtracts the information that would otherwise have been available. Seeking reassurance, on the other hand, does indeed provide external information, and the process stops at the second question: the information is received and put into perspective.

This is the case that demonstrates how the two responses should be interpreted together, and it is the subject of the following paragraph.

If the answer to both is negative, it is unlikely that the method will produce a stable recalibration. If both are positive, there is a theoretical reason to expect an effect. If only the first is positive, the outcome depends on how high the discount is: where it is high, the information arrives but is not recorded, and this is precisely what happens with reassurances.

These two criteria do not guarantee that an intervention will work; they merely distinguish those interventions that are compatible with the model from those that, however plausible, are unlikely to alter its functioning.

Where the Method Lies

The fact that this chapter stops at a single criterion does not mean that the method does not exist.

In a separate work, **Recalibrating the Social Supervisor**, I proposed an operational approach built on these same premises: recognizing the prediction the moment it appears, making it explicit, reconstructing the elements that produced it, considering alternative scenarios, and comparing the prediction noted in advance with what actually happens. It includes a personal log, a catalog of the most frequent errors, and criteria for discontinuing the work.

I chose to leave it out of these pages for a matter of substance, not of space. An essay describing a mechanism and a self-therapy process answer different questions, are aimed at readers in different circumstances, and should be judged by different criteria: the former on its fidelity to the data, the latter on the effects it produces in those who apply it.

Confusing the two would have weakened both. The following chapter therefore provides two basic tools that anyone can use after reading these pages; those seeking a more extensive work will find it elsewhere.

The next chapter tests these criteria against the two questions this essay poses to the reader.

12 Two Fundamental Questions

From the very beginning, this essay has promised to focus on two fundamental questions.

It is time to formulate them and subject them to the criteria established in the previous chapter.

The outcome, it must be said right away, is not uniform. The two questions do not serve the same function, and one of them is not a tool for recalibration.

The questions collected in the appendix “*Questions for Further Reflection*” have a different status: they are aids to be consulted with a clear head, and they are useful later, when it remains to be decided what to do with a judgment that has already been formed.

Observations, Inferences, Predictions

Before formulating them, we need to make a distinction that both presuppose—and that we constantly lose sight of in daily life.

Events that can be perceived or documented are observable: a person utters a sentence, smiles, leaves quickly, or does not reply to a message. Describing them does not require attributing meaning to them.

Every time we attribute an intention, an emotion, or a judgment to someone, however, we are making an inference. And when we extend that inference to the future, we are making a prediction.

The difference is clearer when viewed in sequence:

“He greeted me briefly” is an observation.

“He was annoyed with me” is an inference.

“He won’t invite me again” is a prediction.

In everyday language, we shift from one to the other without realizing it, and we treat the third statement with the same certainty as the first. Inferences are indispensable—without them, social life would be impossible—and the problem isn’t making the inference; it’s forgetting that we’ve made it.

The same distinction, applied to the subject matter of this essay, separates three things. The expressed judgment is an observation: someone says something about us, and we’ve heard it. The actual judgment—what that person really thinks—is an inference: it is not observable, and it does not necessarily coincide with the expressed judgment, which may be reserved, polite, ironic, or false. The presumed judgment is a prediction, and it resides in another mind: not in that of the person judging us, but in our own.

Confusing a prediction with an observation means attributing the status of an established fact to what our mind has produced. Both of the following questions serve to prevent this.

The first question

Can this person change anything in my life?

This question stems from the chapter on social value. If the weight of a judgment depends on the power that the person making it has over us, then when faced with a judgment that troubles us, there is a possible way to assess it: determining whether that person can change anything in our lives.

In most cases, the answer is no. A comment from a stranger, the opinion of someone we'll never see again, the view of someone who has no influence on our lives.

This is not a consolatory question. It does not ask us to downplay the judgment or convince ourselves that it doesn't matter: it asks for an assessment, and the answer may very well be yes. When it is, our concern is proportionate, and the system is functioning as it should.

What the first question does—and what it doesn't do

The criterion from the previous chapter serves to determine whether an intervention can recalibrate the system. Applied to this question, the answer is no—and it's worth stating this before explaining what the question is for.

The chapter on the mechanism's maladjustment has already established why. The reaction precedes the verification. By the time the verification arrives, the system has already produced its prediction, and the organism has already begun to respond.

Verification corrects the assessment, not the reaction.

And with respect to the two operational criteria, it does not fare well. It does not increase external information: it introduces no data on how that person truly perceives us. It reduces the discount only partially, and indirectly: it makes it less automatic to assign a weight to the judgment that it does not have.

It would therefore be incorrect to present it as a technique capable of eliminating discomfort. It does not do so, and the model predicts that it cannot do so, at least in the immediate term.

Its function is different, and it's best to define it precisely rather than measure it against a criterion that doesn't apply to it. It doesn't serve to change the social supervisor: it serves to prevent the initial reaction from being amplified by subsequent processing—the imagined consequences, the scenarios, the second-guessing in the days that follow, which are almost always longer and more costly than the reaction itself. It is a tool for containment, not correction. It doesn't take away the first blow. It takes away what follows.

The Second Question

What would I think if my judgment were not observable by anyone?

This question concerns the other side of the coin: not how we are judged, but how we judge.

It stems from the chapter on recursion and the one on judgments as signals. If the way we evaluate something is itself observable, and if what we say tends to become what we think, then among our judgments there are some that we would not have formulated under different circumstances.

The question invites us to distinguish between what we evaluate and what it is in our interest to evaluate.

Why We Don't Answer by Looking Within

There is, however, a difficulty that the essay itself has created, and which it would be wrong to sidestep.

The chapter on adapting to the listener showed that expedient judgments do not remain recognizable as such. They become sincere. If this is the case, the question cannot be resolved through introspection: by looking within ourselves, we will find a judgment that seems authentic to us, and that is exactly what one would expect in both cases.

When posed as an examination of conscience, the question yields nothing.

It can yield something if posed as a comparison of situations—that is, as a prediction to be verified rather than as an introspective examination. What do I think of that book, that person, that decision, when I discuss it with someone who thinks differently from me? And when I discuss it with someone who thinks like me? Do the two answers coincide?

Formulated this way, the question satisfies the first criterion: it yields information that does not stem from self-perception. It is a comparison between two observable behaviors, not between two intuitions.

As for the second criterion, the effect is more uncertain. The question reduces the discount only if the discrepancy between the two situations is clear enough that it cannot be absorbed—and absorbing it is easy: “I expressed myself differently,” “I was more cautious out of politeness,” “it was a different context.” These explanations are almost always available, and almost always partly true. That is why the comparison should be made in cases where the difference is striking, not in cases of subtle nuances.

What the Second Question Does—and What It Doesn't Do

For this question, too, it's worth clarifying what it does not promise.

It does not establish which judgment is the “true” one. This is the most direct consequence of what we've seen regarding adaptation to the listener: if the adjusted judgment tends to become sincere, there is no original, intact judgment somewhere to be recovered. Discovering that one's opinion varies depending on the interlocutor does not mean that one version is authentic and the other is feigned. It means that the judgment is context-sensitive—which is a different, and more useful, piece of information.

It does not eliminate conformism, nor does it automatically reduce it. Knowing that an opinion has been shaped in part by one's environment does not make it false, nor does it oblige one to change it: many beliefs inherited from one's environment are excellent beliefs.

It does not guarantee authenticity, and it would be suspicious if it did. An essay that argues for the non-existence of a vantage point external to the system cannot then offer a tool for reaching it.

What it does is make observable something that normally leaves no trace: the influence of the social context on one's judgments. It does not say how great that influence is in general, but it demonstrates it in a concrete case, with two behaviors that can be compared.

It is less than the reader might wish for. It is also the only thing the model allows us to promise.

Why They Must Be Anchored

The more practical problem remains, and it is the one for which the previous chapter has already indicated the solution.

When considered with a clear head, both questions fall under what we had called "training": they keep the topic fresh in one's mind, and do not alter anything at the moment they would be needed. The moment they would be needed is a different one—when the unease begins, when one realizes one is forming a judgment in front of someone.

If they are to function as triggers, what we saw in the previous chapter points to one condition: that they be associated with a specific and consistent signal, and that the association be repeated exactly the same way rather than reinvented.

The signal should not be sought in thought, but where the system first manifests itself: in the essay on social self-governance, I argued that this mechanism does not communicate through mental contents but through impulses and resistances. It is these that are recognizable—the knot in the stomach, the impulse to respond, the hesitation before speaking, the sudden need to explain oneself.

And one final condition is necessary—the easiest to overlook: having decided in advance what to do with the window that recognition opens. Recognizing without a ready alternative produces observers who are aware of their own automatisms.

In the case of the first question, the alternative may be minimal: responding to the inquiry and stopping there, without proceeding to construct scenarios. In the case of the second, jotting down the judgment before expressing it—because a judgment put in writing can no longer be rewritten afterward, and it is the only way to realize how much one's perspective has shifted.

If a prediction keeps coming back

It may happen that a prediction keeps recurring in exactly the same form, in different situations, without being affected by verification. In that case, the two questions aren't enough, and it's helpful to know what else could be done.

A prediction does not arise out of thin air: it is constructed from something. You can try to reconstruct it—which observable facts gave rise to it, what inferences were added, how much stems from the present situation and how much from past experiences, and what available information was overlooked. This process does not determine whether the prediction is right or wrong: it makes visible the process that produced it, and a prediction whose construction is revealed is already less like a fact.

It's a more time-consuming task than simply asking a question, and it requires taking notes. Anyone wishing to approach it in a systematic way will find a framework designed for this purpose in the essay **Recalibrating the Social Supervisor**.

What They Really Promise

It's worth taking stock without sugarcoating the results.

Neither question prevents the system from reacting. Neither recalibrates it on its own. Compared to the approaches described in the previous chapter—long-term relationships, changes in environment, and time—these are minor tools, and moreover, they are subject to the decay we'll discuss later.

What they do is make observable a mechanism that operates beyond the reach of self-examination. This is no small thing: it is the prerequisite for anything else to happen. But it is not the same as solving the problem, and presenting them differently would betray the criteria this essay has set for itself.

And they do this by shifting attention outward, not inward: they serve to interrupt self-examination, not to refine it.

Those who use them with some consistency should expect two modest and verifiable effects: less time spent reconstructing scenarios after receiving feedback, and a few instances in which they realize they've changed their opinion without realizing it.

If nothing of the sort happens, the tools are not producing the effects the model predicts. It may be that they were not used as described here; it may be that the model is wrong. The second possibility must be kept open just as much as the first. If this were not the case, this essay would be asking for trust rather than verification.

13 The Decline of Recalibration

If the model proposed in this essay is correct, we must accept an unpleasant consequence.

No recalibration is permanent.

It would be reassuring to think that, once a social supervisor's prediction has been modified, it remains stable. But a system whose task is to adapt to the environment cannot function that way: it would continue to use predictions designed for a world that might no longer exist.

Decay, however, does not take a single form. There are two forms, which operate through different mechanisms and are cumulative.

The first form of decay: the system continues to learn

The social supervisor never stops comparing its expectations with experience.

Every new relationship, every change in the environment, and every shift in one's social position yields new information. Some confirm previous predictions. Others modify them.

Recalibration is therefore not a one-time event, but a continuous process.

The same mechanism that allows us to correct an erroneous prediction can, subsequently, construct another one that is just as erroneous.

Why It's Easy to Fall Back into Old Patterns

Many people describe the same experience.

For a certain period, they manage to speak more freely, to open up without feeling particularly uncomfortable, and to attach less importance to the judgment of others.

Then something changes. An unpleasant incident. A new environment. A particularly critical person.

And the discomfort seems to return.

This phenomenon doesn't prove that the previous change was illusory. It may simply indicate that the social supervisor has continued to do what it exists to do: update its predictions.

The problem isn't that it has started learning again. The problem is what it has learned.

It is worth recalling, in this regard, the asymmetry described in the chapter on the principle of recalibration. A negative disconfirmation usually occurs in a single, intense, and unambiguous episode; a positive one occurs in many small, easily reinterpretable instances. It follows that a single episode can quickly undo what many experiences had slowly built up—not because the system is fragile, but because the two directions do not affect it with the same effectiveness.

The Importance of the Environment

This observation restores importance to a theme we have already encountered.

In the chapter dedicated to the principle of recalibration, we saw that the most profound changes seem to occur within lasting relationships and relatively stable environments.

The reason for this is now clearer.

A stable environment not only shapes expectations but also continually confirms them. Each successive confirmation makes it less likely that a single episode will overturn them.

An unpredictable environment, on the other hand, forces the supervisor to make constant readjustments, and no prediction has time to take hold.

The second form of decay: we grow accustomed to the tools

There is, however, a second way in which recalibration wears down, and it directly concerns what this essay proposed to the reader in the previous chapter.

Voluntary attention is a limited resource, and it fades when the stimulus does not change. This is a general principle of mental functioning: a constant noise ceases to be perceived, the contact of clothing against the skin disappears from consciousness, a sign seen every day becomes invisible.

The same happens to the psychological mechanisms we set out to observe, and to the tools we use to observe them.

This gives rise to a paradox that I have addressed elsewhere in my work on attention. To change an automatic response, we must observe it; but precisely because it is repetitive, it tends to escape our attention. And the same is true of the question that is supposed to help us grasp it: after a few weeks, it ceases to be noticed.

Two clarifications make the phenomenon more insidious than it seems.

The first is that habituation does not concern only the individual stimulus, but its category. We do not merely get used to a repeated formula: we get used to “the question I ask myself when someone judges me,” even though it is phrased differently each time.

The second is that habituation is not passive. In response to excessive repetition, the mind does not merely shut down perception; it responds with boredom, which actively drives us to seek something else. This is why self-observation practices are not merely forgotten—they are abandoned. First, the tool loses its vividness; then it begins to repel us.

It is therefore best to state plainly what the reader should expect. The two questions from the previous chapter will cease to work. Not because they are wrong, but because they are tools of mindfulness, and no tool of mindfulness can withstand repetition.

Stating this is not a precaution: it is a prerequisite for ensuring that the decline, when it comes, is recognized for what it is rather than interpreted as a refutation of the model or a personal failure.

No Final Destination

From this perspective, speaking of definitive healing is misleading.

A predictive system does not reach a final state. It achieves, time and again, a balance between what it has learned and the environment in which it finds itself.

When the environment changes profoundly, that equilibrium changes as well.

This does not mean that all progress is precarious. It means that no amount of progress eliminates the need to keep learning.

A Different Mindset

This conclusion also changes the way we evaluate ourselves.

If discomfort resurfaces, it's natural to interpret it as a personal failure.

The model suggests a different interpretation. The social supervisor isn't necessarily regressing; it's trying to adapt to information it considers new.

The useful question, then, is not "Why did I relapse?" but "Which predictions is the system trying to update?"

The second question does not eliminate the problem. However, it allows us to view it in the same way we have viewed the rest of the process: as the functioning of a system that makes predictions and modifies them in light of experience.

The model also sheds light on the second decay. If one becomes accustomed to the category rather than to the specific formulation, then replacing a worn-out tool with a better-formulated one is of no use: what has worn out is not what the tool requires—which remains valid—but the attention's capacity to notice it—and that capacity does not return simply by changing the words, because it was not the words that had worn it out.

Recognizing this second decay means, above all, avoiding attributing to the tools a failure that actually belongs to the normal functioning of attention.

14 When to Stop the Observation

This essay has argued, with support from the clinical literature, that self-directed attention is one of the mechanisms that perpetuate social discomfort.

It is therefore bound to say something uncomfortable about oneself.

A text that describes at length the way we imagine being judged, and that proposes two questions to ask oneself when one feels observed, inevitably leads to an increase in self-attention.

For most readers, this is not a problem. For some, it may be.

The Mechanism to Be Wary Of

The chapter on intuitive remedies has already described the mechanism: those who shift their attention to themselves stop gathering information about how others are reacting, and end up using their own feelings as evidence of what others think. It is the mechanism that, in Clark and Wells' model, sustains social anxiety.

The previous chapter also showed that even attentional tools lose their effectiveness over time, because we get used to them. The risk discussed here is different: it does not concern their effectiveness—which wanes—but their use, which under certain conditions can be harmful.

This is the mechanism that Chapter 5 described as heightened presupposition.

A tool that requires one to observe one's own social functioning can, in principle, fuel it. There is nothing paradoxical about this: it is the direct consequence of what the model asserts.

Hence the most important—and simplest—point of the entire essay. The useful direction of attention is not inward but outward. The two questions serve to recognize that presumed judgment is largely a construct, not to better monitor how one is being evaluated.

The purpose of the method is not to increase the time spent observing the social supervisor. It is to reduce it.

What to Observe

Before even determining when to stop, it is worth clarifying a misunderstanding that makes observation harmful from the very beginning.

This essay encourages you to observe automatic behaviors, not to evaluate yourself.

These are two completely different mental activities.

The first goes like this: in this situation, the system produced this prediction.

The second goes like this: I'm messed up.

The second is not an observation: it is a judgment—and, moreover, a judgment about oneself formulated using the very mechanism the essay has just described as unreliable. If self-observation takes that form, it is not applying the model; it is providing an example of it.

When to Stop

Observation should produce at least one of two effects: a more precise understanding of how one's own system works, or a gradual decrease in the thought processes that follow a received judgment.

If, after a few weeks, it produces neither of these, and instead leads to more worry, more rumination, or more self-control, it is best to stop.

The criterion can be stated even more simply: if the distress increases and the insights do not—that is, if the discomfort grows but the number of times you catch the mechanism in action does not—you must stop.

This criterion allows for no middle ground: reducing the observation is not enough; it must be stopped.

It is a rough criterion, but it has the advantage of being observable: both conditions can be verified within a few weeks, without the need to judge oneself.

And stopping does not mean you have failed. It means you have applied what this essay argues: that an intervention is evaluated by its effects, not by its plausibility.

When Not to Use It at All

Certain conditions make intensified self-observation more risky than helpful. These are not diagnostic categories, but situations in which attention risks becoming part of the problem.

A strong tendency to ruminate. Intense anxiety. Clinically significant depression. Obsessive-type functioning, in which control is already the habit one wishes to break free from.

In all these cases, there is already a self-perpetuating mental process, and adding yet another focus for that process is not helpful.

The appropriate scope of what has been proposed is the ordinary automatisms of psychological life—not disorders. For disorders, there are treatments backed by decades of research, and a theoretical essay is no substitute for a professional—particularly for social anxiety, which appears in these pages as an extreme case of an ordinary mechanism, not as something to be addressed on one's own.

When to Seek Help

Finally, there are signs that no longer concern the use of a tool, but rather the condition of the person using it.

If using the tool consistently increases your distress instead of reducing it.

If it interferes with work.

If it interferes with relationships.

If it takes up an increasing portion of the day.

In these cases, the helpful thing is not to change methods, nor is it to read another essay: it is to talk about it with someone who deals with this professionally.

It's worth adding that the criteria outlined in this chapter also apply—and with even greater force—to the practical approach proposed in **Recalibrating the Social Supervisor**. A method that requires you to record your predictions daily involves a much greater degree of self-observation than answering just two questions, and the same contraindications apply even more strongly.

The Most General Limitation

There remains a limitation that no precaution can eliminate, and it is a consequence of the very principle of this essay.

There is no external vantage point within the system. Anyone who observes their own presumed judgment does so using the very same mechanism that produces it. Even the assessment “I’m observing myself too much” is, in turn, an assessment.

This is why the criteria proposed here are based on observable facts—distress increases, recognition does not, the day fills up—and not on a subjective impression. Subjective impressions, in this matter, are precisely what the essay advises us to be wary of.

And this is also why an outside person, when available, remains the best tool: not because they see things correctly, but because they see from a vantage point that the system cannot reach.

On this point, I can speak more directly, and from personal experience.

The most widely accepted cognitive model of panic is attributed to David Clark—the same author who, together with Adrian Wells, formulated the model of social anxiety cited several times in these pages. He describes the attack as a feedback loop: certain ordinary bodily sensations are interpreted as signs of an impending catastrophe; the interpretation increases arousal; the arousal confirms the interpretation. It follows that the loop is broken when the interpretation changes, and this is a documented outcome.

I have had panic attacks, and they ceased when I realized what fueled them: the fear—never articulated—that others would notice the judgments I held about them and make me pay for them. This is what the chapter on “judging as exposure” describes, seen from the perspective of someone who fears being discovered.

Two clarifications, however, without which this account would do more harm than good.

The first is that I didn’t figure this out on my own, nor through reading. It took about ten psychotherapy sessions. This is consistent with what I stated a few lines above: the cycle is fueled primarily from within, and what breaks it comes from the outside. I don’t know of a way to get there through introspection.

If a reader who experiences panic attacks does not see them subside by understanding this, they should not conclude that they have misunderstood, nor that what I have described is false: a single case proves nothing, and for the same reason, it is not disproved by another case. It may be that their panic stems from something else; and it may be that they have understood perfectly well, and that understanding, on its own, is not enough. It wasn't enough for me: that understanding came during the course of psychotherapy, not in place of psychotherapy.

Conclusion

Every model of the mind runs the same risk: turning a name into an explanation. One begins by trying to account for a phenomenon, and ends up with a label that seems to explain it simply by naming it.

The social supervisor can meet this fate, and in the chapter on the internalized observer I tried to prevent it by saying that it is not the voice of a norm but the assessment of a reaction. It was a promise, and it is worth verifying it now.

If, upon finishing this essay, the reader were to simply think, “That’s my social supervisor,” the model would have failed. Not because the name is wrong, but because no label adds knowledge unless it changes the way we observe the facts.

The purpose of these pages was not to introduce a new mental faculty. It was to propose a different way of interpreting phenomena we know well: embarrassment, self-censorship, the need for approval, the relief that follows confirmation, the weight we attribute to some people and not to others, and the tendency to adjust our opinions depending on who is listening.

The model can be summarized in a few statements.

Social behavior is guided primarily by the judgment we anticipate, not by the judgment we receive.

That prediction is the product of a system that operates recursively on input that is not observed but inferred, and constructed primarily based on how we see ourselves.

The weight we assign to a judgment does not depend on the judgment itself, but on the power that the person making it has over us—even if the system, at the moment it reacts, does not verify that power.

The same mechanism works both ways: it shapes what we show and the judgments we express. And the judgments we’ve adapted to our listeners end up seeming to us as if they’ve always been our own.

The system continues to update itself, but not symmetrically: some experiences alter its predictions much more easily than others, and no change is permanent.

If these statements are correct, social discomfort does not depend solely on what others think of us. It depends above all on what our system predicts they will think—and that prediction, in most cases, is not an accurate reading of others.

This distinction does not eliminate the problem. It changes its meaning.

When a prediction is recognized as a prediction, it ceases, at least in part, to present itself as an inevitable description of reality.

Whether this interpretation is correct cannot be decided by a scholar. It can only be determined by testing against the facts.

The most important criterion, therefore, is not how convincing the model appears, but how well it succeeds in producing new observations, verifiable predictions, and questions we previously did not know how to formulate.

If it succeeds in doing so, it will have been valuable. If it fails, it must be corrected or replaced.

This is the normal fate of any model that claims to describe the workings of the mind. And it is, ultimately, the same rule that this essay attributes to the social supervisor: to modify one's predictions when experience disproves them.

Appendix I — References

This list compiles the works on which the essay’s empirical claims are based, so that the reader may verify them—and, where they see fit, challenge them.

The order is thematic and follows the line of argument.

The Social Foundations of the Self

A. Smith, **The Theory of Moral Sentiments**, 1759 — the impartial spectator.

C. H. Cooley, **Human Nature and the Social Order**, 1902 — the looking-glass self.

G. H. Mead, **Mind, Self, and Society**, 1934 — the generalized other.

E. Goffman, *The Presentation of Self in Everyday Life*, 1956 — impression management.

P. Rochat, **Others in Mind: Social Origins of Self-Consciousness**, 2009.

Relational value, belonging, partner choice

M. R. Leary and R. F. Baumeister, “The nature and function of self-esteem: Sociometer theory,” in *Advances in Experimental Social Psychology*, 32, 2000, pp. 1–62.

M. A. Nowak and K. Sigmund, “Evolution of indirect reciprocity,” **Nature**, 437, 2005, pp. 1291–1298.

R. Bénabou and J. Tirole, “Incentives and Prosocial Behavior,” *American Economic Review*, 96, 5, 2006, pp. 1652–1678 — reputational utility.

Recursion and Higher-Order Beliefs

J. M. Keynes, *The General Theory of Employment, Interest, and Money*, 1936, Chapter 12 — the beauty contest.

R. D. Laing, H. Phillipson, and A. R. Lee, **Interpersonal Perception: A Theory and a Method of Research**, 1966.

N. Luhmann, *Social Systems: An Outline of a General Theory*, 1984 — double contingency.

R. Nagel, “Unraveling in guessing games: An experimental study,” *American Economic Review*, 85, 5, 1995, pp. 1313–1326 — the leveled model.

T. Mauersberger and R. Nagel, “Levels of Reasoning in Keynesian Beauty Contests: A Generative Framework,” 2018 — review.

Perception of People

J. S. Uleman et al., on spontaneous trait inferences: the automatic and unintentional attribution of stable characteristics based on the observation of behavior.

A. Bott, L. Brockmann, I. Denneberg, E. Henken, N. Kuper, F. Kruse, and J. Degner, “Spontaneous Trait Inferences From Behavior: A Systematic Meta-Analysis,” *Personality and Social Psychology Bulletin*, 2022 — 86 studies, 13,630 participants, average effect size $d_z = 0.59$.

J. Degner, “Revisiting Spontaneous Trait Inferences: Questions, Challenges, and Intradisciplinary Connections,” *Social and Personality Psychology Compass*, 2025 — critical review.

Metaperceptions and Accuracy

D. A. Kenny and B. M. DePaulo, “Do people know how others view them? An empirical and theoretical account,” *Psychological Bulletin*, 114, 1, 1993, pp. 145–161.

E. N. Carlson, R. M. Furr, and S. Vazire, on meta-insight: the ability to recognize when one’s own estimate of others’ judgments is more or less accurate.

E. J. Boothby, G. Cooney, G. M. Sandstrom, and M. S. Clark, “The Liking Gap in Conversations: Do People Like Us More Than We Think?”, *Psychological Science*, 29, 11, 2018, pp. 1742–1756.

Harper et al., “How Well Do We Know How Others See Us? A Systematic Review of Meta-Accuracy Across Relational Contexts, Research Foci, Attribute Domains, and Measurement Approaches,” *Journal of Personality*, doi 10.1111/jopy.70075.

Narcissism and Metaperception

E. N. Carlson, S. Vazire, and T. F. Oltmanns, “You probably think this paper’s about you: Narcissists’ perceptions of their personality and reputation,” *Journal of Personality and Social Psychology*, 101, 1, 2011, pp. 185–201.

It should be noted that the literature is not unanimous: there is a contrary view, according to which narcissistic traits are also associated with an overestimation of others’ judgments. The finding used in this essay is the one with the strongest support, though it is not the only one available.

Judgment as a Signal

P. Bourdieu, **La distinction: critique sociale du jugement**, 1979.

T. Kuran, **Private Truths, Public Lies: The Social Consequences of Preference Falsification**, 1995.

J. Tosi and B. Warmke, “Moral Grandstanding,” **Philosophy & Public Affairs**, 44, 3, 2016, pp. 197–217.

J. J. Jordan, R. Sommers, P. Bloom, and D. G. Rand, “Why Do We Hate Hypocrites? Evidence for a Theory of False Signaling,” *Psychological Science*, 28, 3, 2017, pp. 356–368.

J. J. Jordan and D. G. Rand, “Signaling when no one is watching: A reputation heuristics account of outrage and punishment in one-shot anonymous interactions,” *Journal of Personality and Social Psychology*, 117, 1, 2019, pp. 57–88.

E. Goffman, “On Face-Work: An Analysis of Ritual Elements in Social Interaction,” *Psychiatry*, 18, 1955, pp. 213–231 — the working consensus.

P. Brown and S. C. Levinson, **Politeness: Some Universals in Language Usage**, Cambridge University Press, 1987 — mutual face-saving.

J. A. Minson and B. Monin, “Do-Gooder Derogation: Disparaging Morally Motivated Minorities to Defuse Anticipated Reproach,” *Social Psychological and Personality Science*, 3, 2, 2012 — the disparagement of those who adopt a more demanding standard.

How Communication Shapes Beliefs

E. T. Higgins and W. S. Rholes, “‘Saying is believing’: Effects of message modification on memory and liking for the person described,” *Journal of Experimental Social Psychology*, 14, 1978, pp. 363–378.

G. Echterhoff, E. T. Higgins, and J. M. Levine, “Shared reality: Experiencing commonality with others’ inner states about the world,” *Perspectives on Psychological Science*, 4, 2009, pp. 496–521.

L. Festinger, *A Theory of Cognitive Dissonance*, 1957 — cited for distinction, not as a foundation.

D. J. Bem, “Self-perception theory,” in **Advances in Experimental Social Psychology**, 6, 1972 — idem.

F. Heider, “Attitudes and cognitive organization,” *Journal of Psychology*, 1946, and *The Psychology of Interpersonal Relations*, Wiley, New York, 1958 — the equilibrium theory.

G. Simmel, “Fashion,” *International Quarterly*, 10, 1904 — imitation and distinction.

Anonymity, observation, behavior

E. Hoffman, K. McCabe, and V. L. Smith, “Social distance and other-regarding behavior in dictator games,” *American Economic Review*, 86, 3, 1996, pp. 653–660.

M. Bateson, D. Nettle, and G. Roberts, “Cues of being watched enhance cooperation in a real-world setting,” *Biology Letters*, 2, 3, 2006, pp. 412–414—the so-called “watching-eyes effect,” according to which simple images of eyes are said to increase cooperative behavior. This would be the natural experiment for a model such as the one described in these pages, and the model does not rely on it: the meta-analyses below do not confirm it.

On the meta-analyses that do not confirm the effect: “Artificial surveillance cues do not increase generosity: two meta-analyses”; and K. Dear, K. Dutton, and E. Fox, “Do watching eyes influence antisocial behavior? A systematic review and meta-analysis,” *Evolution and Human Behavior*, 40, 3, 2019, pp. 269–280.

Self-awareness, attention, insight

S. Duval and R. A. Wicklund, **A Theory of Objective Self-Awareness**, Academic Press, New York, 1972.

A. Fenigstein, M. F. Scheier, and A. H. Buss, "Public and private self-consciousness: Assessment and theory," *Journal of Consulting and Clinical Psychology*, 43, 1975, pp. 522–527.

T. Gilovich and K. Savitsky, "The spotlight effect and the illusion of transparency: Egocentric assessments of how we are seen by others," *Current Directions in Psychological Science*, 8, 1999, pp. 165–168.

A. M. Grant, J. Franklin, and P. Langford, "The Self-Reflection and Insight Scale: A new measure of private self-consciousness," *Social Behavior and Personality*, 30, 8, 2002, pp. 821–836.

R. Harrington and D. A. Loffredo, "Insight, Rumination, and Self-Reflection as Predictors of Well-Being," *The Journal of Psychology*, 145, 1, 2011, pp. 39–57 — on the self-absorption paradox.

Clinical Aspects of Social Anxiety

D. M. Clark and A. Wells, "A Cognitive Model of Social Phobia," in R. G. Heimberg, M. R. Liebowitz, D. A. Hope, and F. R. Schneier (eds.), *Social Phobia: Diagnosis, Assessment, and Treatment*, Guilford Press, New York, 1995, pp. 69–93.

J. P. Tangney and R. L. Dearing, **Shame and Guilt**, Guilford Press, New York, 2002 — the distinction between shame and guilt.

D. M. Clark, "A Cognitive Approach to Panic," **Behavior Research and Therapy**, 24, 4, 1986, pp. 461–470.

Dimensional Structure of Social Anxiety

A. M. Ruscio, "The latent structure of social anxiety disorder: consequences of shifting to a dimensional diagnosis," *Journal of Abnormal Psychology*, 119, 4, 2010, pp. 662–671 — taxometric analyses of data from the National Comorbidity Survey Replication.

S. Crome, A. Baillie, T. Slade, and A. M. Ruscio, "Social phobia: further evidence of dimensional structure," *Australian and New Zealand Journal of Psychiatry*, 44, 11, 2010, pp. 1012–1020.

Predictive Processing

A. Clark, "Whatever next? Predictive brains, situated agents, and the future of cognitive science," *Behavioral and Brain Sciences*, 36, 3, 2013, pp. 181–204.

There is a recent and evolving body of literature on interpreting social anxiety and loneliness in terms of the accuracy attributed to social cues; this essay draws on it as an interpretive framework rather than as a foundation.

Depression and Accuracy

M. T. Moore and D. M. Fresco, "Depressive realism: A meta-analytic review," *Clinical Psychology Review*, 32, 6, 2012, pp. 496–509 — seventy-five studies, over seven thousand participants, negligible effect size.

“Sadder ≠ Wiser: Depressive Realism Is Not Robust to Replication,” *Collabra: Psychology*, 8, 1, 2022 — unsuccessful replication of the original study by L. B. Alloy and L. Y. Abramson (1979).

On the negative bias in metaperceptions associated with depressive symptoms and low self-esteem: studies on self-other agreement and the accuracy of metaperceptions in the traits of the five-factor model.

Cultural Variations

R. A. Kleinknecht, D. L. Dinnel, E. E. Kleinknecht, N. Hiruma, and N. Harada, “Cultural factors in social anxiety: a comparison of social phobia symptoms and Taijin kyofusho,” *Journal of Anxiety Disorders*, 11, 2, 1997, pp. 157–177.

B. Zhou, F. Lacroix, J. Sasaki, Y. Peng, X. Wang, and A. G. Ryder, “Unpacking Cultural Variations in Social Anxiety and the Offensive-Type of Taijin Kyofusho Through the Indirect Effects of Intolerance of Uncertainty and Self-Construals,” *Journal of Cross-Cultural Psychology*, 45, 10, 2014, pp. 1561–1578.

Social Evaluation in Early Childhood

J. K. Hamlin, K. Wynn, and P. Bloom, “Social evaluation by preverbal infants,” *Nature*, 450, 2007, pp. 557–559.

K. Lucca et al. (ManyBabies Consortium), “Infants’ Social Evaluation of Helpers and Hinderers: A Large-Scale, Multi-Lab, Coordinated Replication Study,” *Developmental Science*, 2025 — unsuccessful replication of the previous result.

N. Botto and P. Rochat, “Sensitivity to the Evaluation of Others Emerges by 24 Months,” *Developmental Psychology*, 54, 9, 2018, pp. 1723–1734.

Previous works by the author

The following works do not constitute independent sources supporting the model. They represent the stages of its development and are cited in the text to avoid repeating arguments already presented elsewhere.

The distinction matters: a model is not validated by citing its own earlier versions.

Psychology of Needs — for the concepts of need and desire, which are not redefined here.

The Presumed Judgment, 2026 — the first formulation of the construct, the social judgment supervisor, and the expected relational cost. It identified the “*watching-eyes*” effect as the model’s most characteristic prediction; subsequent meta-analyses did not support this, and these pages do not revisit it.

Social Self-Governance, 2026 — the system that guides through emotions, and the extension to what one is willing to take into consideration.

Self-Therapy as Attention Management, 2026 — the window of choice, the distinction between training and triggering, and habituation to attentional tools.

Recalibrating the Social Supervisor, 2026 — the operational pathway derived from this model: recognition of the prediction, reconstruction of the elements that produced it, alternative scenarios, behavioral experiments, personal log, most frequent errors, and criteria for interruption. This is the work to which the present essay refers for its practical section, and from which the distinction between observations, inferences, and predictions stems, along with its application to expressed, actual, and presumed judgment.

A Precedent Not Adopted

I. Kant, *Critique of Judgment* (*Kritik der Urteilskraft*), 1790.

When I first conceived of this essay, I wondered if I were reinventing a wheel that Kant had already built. I looked into it. The answer is no, and I mention this because the question is legitimate and others might ask it.

The *Critique of Judgment* concerns the faculty of judgment and deals with the beautiful, the sublime, and teleology in nature. It says nothing about the relational consequences of judging. There is an observation in it that still holds true: whoever says “this is beautiful” is not expressing a private preference but is seeking the assent of others, even though they cannot present anything to justify it. It is the most illustrious precedent for the demarcation criterion that this essay has considered and rejected.

I must, however, add my own judgment, which is harsh. The distinction between “I like it” and “it is beautiful” was new in 1790 and is today a truism—which is the ordinary fate of successful ideas, not a demerit. The rest struck me as an arbitrary and unnecessarily complicated analytical exercise, constructed to hold a system together rather than to understand how people judge. I gained nothing from those pages. This is a reader’s impression, not a thesis on the history of philosophy, and I state it as such.

Note on the completeness of this list

The research on which this essay is based was conducted using accessible academic sources in English and Italian. It does not constitute a systematic review, and the major specialized databases were not consulted.

Therefore, in all likelihood, there are relevant works that do not appear here—including works that might anticipate or contradict parts of the model.

Appendix II — What the Model Asserts

Several predictions have been made throughout the essay. Scattered across the chapters, however, they remain invisible as a whole: they are noticed one at a time and then forgotten.

This appendix brings them together in one place.

The predictions are ordered by decreasing risk: from the most critical—whose failure would compromise the entire framework—to those whose failure would require only local corrections.

Generalization from the Pathological to the Ordinary

The differences between high and low social anxiety are primarily quantitative—the weight of prior beliefs, the precision attributed to external information, the depth of recursion—rather than qualitative.

This is the essay's riskiest prediction. The architecture described already exists in the literature as a model of a disorder; what is argued here is that it describes ordinary functioning, of which social anxiety would be the extreme case.

Available support: taxometric studies converge on a dimensional structure of social anxiety that is continuous with normal functioning.

Limitation of the evidence: the dimensionality pertains to symptoms, not mechanisms. A continuous phenotype does not prove that individuals with low anxiety use the same processes, albeit in an attenuated form.

The two parameters are indeed distinct

Assumption and discounting—the extent to which presumed judgments are constructed from self-perception, and the extent to which external feedback is devalued—vary relatively independently within the population.

It follows that all four possible combinations should be represented: high and high in social anxiety, low and high in narcissism, and the two mixed combinations yet to be verified.

The two combinations not yet observed

High presupposition and low discounting should produce an unstable self-image: the sense of how one is perceived fluctuates with every piece of external feedback.

Low attribution and low discounting should correspond to the best-calibrated profile, in which the presumed judgment is constructed using both self-perception and available feedback. Not a disorder: the case in which the system does what it appears to be designed to do.

These two combinations were formulated before consulting the literature, precisely to avoid tailoring them to the results.

The Depth of Recursion

The number of iterated levels—what will they think of me, what will they think of the way I appear—increases as social anxiety increases.

This prediction arises from the intersection of two bodies of literature that do not cite one another: the hierarchical model in guessing games, where people stop after one or two steps, and clinical research on social anxiety.

It can be verified using existing tools.

The weight of a judgment depends on the power of the person making it

The weight attributed to a judgment depends on the power that the person making it has over us, and not on the probability of future interactions.

This is a correction to a previous formulation by the author, and the difference is verifiable: by experimentally manipulating the power an observer has over the subject while keeping the frequency of interaction constant, self-censorship, conformity, and emotional arousal should vary in a predictable manner.

The Distortion of Expressed Judgments

The self-presentational component of a judgment increases in proportion to its observability and the relevance of the listener, and does not drop to zero even under conditions of anonymity.

The cost of violation

Anyone who proposes a criterion as valid for everyone should be judged more severely when they violate it than someone who presents it as a personal preference or as a rule valid only under certain circumstances.

This must be distinguished from what has already been noted: that the penalty for hypocrisy depends on the falsification of the signal and not on mere inconsistency. That is an established finding. The proportionality between the stated scope of the criterion and the severity of the penalty is not.

Private and Public Behavior

The internalized observer is systematically weaker than the external one. It follows that behavior under conditions of true anonymity should be reduced compared to public behavior, though it should not disappear entirely.

It is this prediction that prevents the internalized observer from being used as a universal explanation.

The Expected Effects of the Two Questions

Those who consistently use the two questions proposed in Chapter 12 should observe two modest effects: less time spent reconstructing scenarios after receiving a judgment, and a few instances in which they realize they have changed their opinion without knowing it.

This is the only prediction that individual readers can verify for themselves, and it is also the most fragile, because it does not distinguish between a flawed model and a tool used incorrectly.

Depression Among the Four Combinations

The model posits that the negative interpretation of others' judgments is largely inferred rather than observed. A line of research might have refuted this—depressive realism, according to which, under certain depressive conditions, some evaluations would be more accurate rather than more distorted—but the effect did not hold up: a meta-analysis of seventy-five studies attributes a negligible effect size to it, and a recent replication study failed to reproduce it.

The relevant evidence actually points in the opposite direction: depressive symptoms and low self-esteem distort metaperceptions in a negative way—that is, those affected believe they are seen in a worse light than they actually are. This is what the model calls “high presupposition,” and depression thus occupies the same combination as social anxiety: it does not require the addition of a parameter. This does not prove that the mechanism is identical—it shows that the two conditions share the values of the two parameters.

The Mechanism Outside the West

The model was constructed based largely on Western literature, and the normative weight of others' judgments varies significantly between more individualistic and more collectivist societies. It was necessary to verify whether it was transferable.

The most studied case is *taijin kyofusho*, the Japanese form of social anxiety. The difference from the Western description does not concern the existence of the mechanism, but its content: the fear centers on offending others rather than on self-embarrassment. The anticipation of judgment is the same; what changes is what one fears the judgment will concern.

The variable mediating the association with culture is not the presence or absence of the mechanism, but the type of self-construction—-independent or interdependent—and the phenomenon has also been observed outside Japan, in other collectivist contexts.

This independently confirms a consequence of the model that had not previously been verified: the mechanism is general, and the criteria it applies are imported from the environment. This leads to a prediction that can already be formulated and is now corroborated—intercultural variability should be high regarding the content of the feared judgment and low regarding the structure of the process.

Convergence and Divergence of Judgments

Among those who cooperate, judgments should converge beyond what the shared information justifies; among those in conflict, they should diverge beyond what the information justifies.

The prediction does not concern convergence in and of itself—which could be explained by the sharing of data—but rather its excess relative to the available evidence.

Preventive screening of ideas

The expected social cost of being associated with an idea comes into play before the idea is evaluated and reduces the probability that it will be considered, regardless of its content.

This is the extension proposed in the chapter on adaptation to the listener, and it is the least supported among those put forward in these pages: it arises from the model by analogy with what happens to expressed judgments, not from an empirical finding.

An anomaly that the model does not explain

However, the analysis of cultures has yielded a result that the model does not predict.

Average social anxiety scores are higher in collectivist countries, while clinical prevalence is higher in individualistic ones. The two measures point in opposite directions.

The model described in these pages does not explain this. It suggests a possible interpretation—what varies is not only the intensity of the mechanism but also the threshold beyond which an environment considers it pathological—and this interpretation is consistent with the idea, developed in the chapter on maladjustment, that the system must be evaluated in relation to the environment in which it operates. But it is a plausible interpretation, not a consequence of the model: nothing in what precedes it predicted it.

It is included here because a dissociation of this kind is precisely the sort of data that a theory should account for rather than absorb.

Appendix III — Questions for Further Reflection

The two questions in Chapter 12 are triggers: they work if they are always linked to the same bodily signal—the knot in the stomach, the hesitation before speaking—at the moment the system is activated.

The questions collected here have a different status. They are deliberative aids: they are used with a clear head, when faced with a decision to make—whether to express a judgment or keep it to oneself, how to express it, to try to verify its validity, to change it, or to downplay its significance. They should not be anchored to anything, and they do not claim to influence the reaction: they come afterward, when the reaction has already occurred and it remains to be decided what to do with it.

It is worth clarifying, however, in what sense they provide deeper insight, because Chapter 10 established that reflecting more, on its own, produces no change: reflection reworks internal material, and in terms of the two parameters, it raises the presupposition rather than lowering it.

The kind of deeper exploration proposed here is of a different sort, found in Chapter 12, in the section titled *“If a Prediction Keeps Coming Back.”* It does not consist of revisiting a judgment until it becomes convincing, but rather in making its construction visible: which observable facts gave rise to it, which inferences were added to it, and what might disprove it. A judgment whose construction is glimpsed is already less like a fact. If, on the other hand, the use of these questions boils down to simply thinking longer about the same thing, they are producing the very effect that the essay advises against, and it is best to stop.

Nor should all of them be adopted. Ten questions used every day quickly cease to mean anything: the decay described in Chapter Thirteen applies to them as well—and all the more so. This appendix is a reference to be consulted when needed, not a tool to be kept in constant use.

Each question derives from a specific point in the essay, indicated in each case. This indication is not merely decorative: separated from the chapter that justifies it, each of these questions becomes a maxim of common sense, and the criterion for using it—with its limitations—lies there.

What You Have in Your Hands

Was the judgment in question actually expressed, or did I anticipate it?

From Chapter 12, which distinguishes between an expressed judgment, a real judgment, and a presumed judgment. Some of the decisions made “in response to a judgment” are actually based on a prediction: no one has said anything. Establishing this first changes the significance of all subsequent questions, because an expressed judgment can be reasoned about as if it were a fact, whereas a presumed judgment is being reasoned about as a product of one’s own system.

What did I observe, and what did I add?

Same origin: the distinction between observations, inferences, and predictions, applied this time to one’s own verdict. “He greeted me briefly” and “he won’t invite me again” seem the same thing at the moment; a decision based on the first and one based on the second do not share the same foundation. The question does not ask us to give up inferences—without them, social life would be impossible—but to know how many have been made.

Whether to express it, and how

If I express it, what does that say about me to the listener?

From Chapter Seven. Every criterion made public is a piece of information about the person proposing it, and from that moment on, it applies to them. The question does not suggest remaining silent: it asks us to factor in the cost of the signal, which is always present, even when the judgment is well-founded.

Does the person listening to me have power over me?

This is the essay's first question, turned to the active side. The cost of an expressed judgment does not depend on the size of the audience but on the power of those who make up the audience: the same judgment can be free in one setting and costly in another. Here, too, the answer may very well be yes—and in that case, reticence is not cowardice: it is the balance described in **The Pact That No One Makes**, seen from one's own perspective.

Will my silence be interpreted as a judgment?

From Chapter Seven, section "*The Criterion That Is Not Adopted*." Abstention is not perceived as an absence of judgment: those who do not follow what others follow, or who remain silent where everyone else speaks out, are interpreted as having passed judgment. The real choice, therefore, is not between speaking out and remaining silent, but between two different forms of communication—and it is wise to consider how the second will be interpreted before choosing it out of prudence.

Is what I am about to express the judgment I hold, or the one I ought to hold here?

From Chapter Eight. The question does not ask whether adaptation will occur: the chapter argues that it tends to happen anyway, and without leaving a trace. It asks for an assessment of the gap between the judgment one holds and the one one is about to express, because this is the only moment when the gap is still visible. If there is a gap, the decision changes in nature: it is no longer merely a question of whether to express oneself, but whether one accepts that the convenient, repeated version becomes one's own. If there is no gap, the question is resolved in an instant, as it should be.

Should we verify it, change it, or reduce its weight?

Would I say the same thing in front of someone who thinks differently?

This is the essay's second question, posed as a comparison of situations: not an inspection, but a comparison of two observable behaviors. Chapter Twelve also points out its limitation: the comparison is best made in cases where the difference is striking, because when it comes to nuances, the skeptic always has a ready explanation.

Can my prediction be disproved by anything?

By Chapter 11. If one wishes to verify the validity of a judgment, the key is not to think about it further: it is to formulate it first, in writing, in a form that experience can contradict. A vague

prediction is always reabsorbed; a written prediction can no longer be rewritten after the fact. If one is not willing to do this, the verification has not yet begun.

How much of my attachment to this judgment is attachment to what I have stated?

From the two central chapters of the second part, taken together. A stated judgment is a signal that has already been sent: changing it exposes one to inconsistency, and the system knows this before we do. In resisting a change of judgment, it is therefore advisable to distinguish the epistemic aspect—the reasons why it still seems valid—from the reputational aspect, that is, the cost of contradicting oneself. Only the former concerns validity.

Is downplaying its importance a realization or a consolation?

From Chapter 10. The reassurances that the system devalues can also be self-administered: “It doesn’t matter,” said to calm oneself down, is information that arrives and changes nothing. The alternative is verification—whether the power over us exists or not—with the function that Chapter 12 attributes to it: to contain subsequent processing, not to take the initial blow away. Downplaying the importance of a judgment is a conclusion one reaches; when it is a starting point, it is a consolation.

Finally, the warning from Chapter 14 applies to this appendix as well. To delve deeper here does not mean to revisit the issue: these questions serve to make a judgment observable and to decide what to do with it, and once the decision is made, they have fulfilled their purpose. If you find yourself revisiting them later, you are no longer delving deeper—you are simply ruminating with better material. The criteria for stopping are those already given, and they allow for no middle ground.

Appendix IV — Glossary

Some terms used in this essay have a narrower meaning than their common usage. Since the precision of a model depends on the stability of its vocabulary, it is advisable to define them.

The entries fall into three categories, and the distinction is made clear in each case: terms defined in this essay, terms taken from the scientific literature, and terms drawn from the author's previous works.

Training and Triggering

Coined. Two opposing functions of attention directed toward one's own automatism. Training keeps the subject alive by observing it from ever-changing perspectives and requires variety. Triggering serves to recognize the automatism the moment it kicks in and requires the opposite: a specific, constant signal, repeated identically every time.

Confusing the two is why many attempts at self-observation fail. These terms are drawn from the author's previous work on attention.

Being Chosen

Agreed. The primary way in which others' judgment produces consequences: the likelihood of being preferred as an ally, friend, partner, collaborator, employee, or role model. What is at stake is social value—that is, how desirable a person is as an interaction partner.

It is not the only channel—those who judge may also hinder, punish, or report to others—but it is the one around which the system has formed. Competition is no exception: it is competition to be chosen under conditions of scarcity.

Judgment

A formal judgment, in a stricter sense than in common usage. An evaluation that becomes a verdict and influences the relationship with the person being judged—or with other judges, if the object of judgment is a thing.

The verdict is reflected in the use of the word. Judgment consists of classifications, not measurements. And it is a point of arrival, not a process: "I am still evaluating" makes sense, "I am still judging" does not.

The relational consequence is what distinguishes it from evaluation. Toward a person being judged, one adopts a more or less favorable attitude, which leans toward inclusion or exclusion, trust or distrust, and—in moral cases—the desire to punish or reward. When the object is not a person, the consequence falls on the one who expressed the judgment.

The distinction does not concern the object—one can judge a work of art and evaluate a person—but the nature of the act. And it does not imply awareness: when it comes to people, the verdict forms automatically, as shown by the research on spontaneous trait inferences discussed in the first chapter.

Presumed Judgment

From the author's previous work. The anticipated, generally implicit and unconscious, representation of the judgment that others might form about our behavior.

“Judgment” is to be understood here in a broad sense: not the verbalized opinion, but the entire anticipated social response—approval or disapproval, acceptance or rejection, willingness to cooperate, or loss of trust.

It is a third variable, distinct from both actual judgment and self-evaluation, and it is this that guides behavior.

Internalized Observer

A term with a long history, used here in its proper sense. The mechanism that continues to estimate the judgment of others even in the absence of actual observers.

It must be distinguished from three predecessors with which it is related but does not coincide: Adam Smith's impartial spectator and Mead's generalized other have a single, stable target—a moral ideal, the community—while the observer described here changes targets depending on who is expected to judge. It is not the voice of a norm: it is the estimation of a reaction.

In the model, it is systematically weaker than that of the external observer.

Assumption

Stipulated. One of the two parameters governing presumed judgment—the other being discounting. It measures the extent to which our estimate of how others evaluate us is derived from how we evaluate ourselves, rather than from what is observable in the situation.

When it is high, it is assumed that others see us as we see ourselves—for better or for worse—and not by accident: one looks, and finds what one expected to find.

It concerns how the representation is constructed; discounting concerns how it is maintained.

Recalibration

Agreed. The modification of the social supervisor's predictions following experiences that demonstrably contradict them.

It does not consist of introducing new information, but of preventing it from being devalued. It is not a one-time event: the system continues to update itself, and may even recalibrate in a negative direction.

Recursion

Established, with antecedents in economics and sociology. The property of an evaluation that can itself become the object of evaluation.

It does not refer to a conscious reflection of the type “I think that you think that I think.”

Spontaneous recursion is limited: in most situations, it does not exceed one or two levels, and it is a rapid and largely unconscious process.

It is the criterion that defines the scope of this essay: non-recursive evaluations are not included.

Discount

Stipulated. One of the two parameters governing presumed judgment—the other being presupposition. It measures the extent to which feedback from others is downplayed, reinterpreted, or ignored when it contradicts the judgment.

It concerns how the representation is maintained. It is the parameter against which reassurances break: information arrives, yet it does not alter the system.

Social Supervisor

From the author’s previous works. The system that estimates the relational consequences of possible behaviors and guides their selection.

It does not control behavior directly: it guides it through emotions, as an impulse to act when the prediction is favorable and as resistance when it is unfavorable. What reaches consciousness is the result, not the calculation.

The name is an abbreviation, and carries a risk explicitly acknowledged in its introduction: that the label might take the place of the explanation.

Social Value

Agreed. The extent to which a person is desirable as an interaction partner: ally, friend, collaborator, employee, employer, role model.

It is neither a property of the person nor a single quantity: it depends on the needs of the evaluator, and the same person may be highly valued by one person and not valued at all by another, without either being wrong.

It concerns both sides. What matters is what one is to others and the value of one’s own judgments, because the way one judges is also evaluated.

Evaluation

Agreed. The attribution of a positive or negative value, of a certain intensity, based on one’s own needs and desires.

The criterion is internal to the evaluator and does not claim to apply to others. It is the basic operation, and it also applies outside the social sphere.

The concepts of need and desire are not redefined in this essay; they refer to a previous work by the author.