MIRA

Measurement through Interaction, Revision, and Analysis

A Field Protocol for Belief Updating and Decision Revision Through Physical Behavioral Elicitation

Randy Villalba | ua28villalba@gmail.com | August 2026

Abstract

Understanding how people update beliefs and revise decisions when verified information conflicts with prior perceptions is a central problem in behavioral economics, public policy, and human–data interaction. We introduce **MIRA**, a low-cost field protocol for physical behavioral elicitation that combines resource-constrained choice, verified information, observed decision revision, and open-ended explanation. MIRA records an initial behavioral choice, exposes participants to a verified data point, and observes whether and how their subsequent decision changes. The protocol distinguishes three analytically separate constructs — belief, stated priority, and observed choice — and proposes two candidate measures, **Cognitive Delta** and the **Perceptual Correction Index** (PCI), whose measurement validity remains to be established. We report a pilot consisting of five field deployments with 76 participants across three Colombian cities. The pilot demonstrates the operational feasibility of the protocol in informal market settings, reveals substantial contextual heterogeneity in observed substitution patterns, and documents a systematic divergence in one deployment between participants' reported price perceptions and official price data. These findings are presented as exploratory evidence and hypothesis generation rather than causal or population-level inference. MIRA contributes a field-oriented approach to studying belief updating and decision revision through data physicalization and embodied interaction, together with a validation agenda for behavioral research, public policy, and future applications in decentralized decision-making.

Keywords: Revealed Preferences; Behavioral Economics; Decision-Making Under Scarcity; Cognitive Delta; Elicitation Methods; Belief Updating; Behavioral Elicitation; Decision Revision; Data Physicalization; Human–Data Interaction; Behavioral Decision-Making; Verified Information; Field Research; Mixed Methods.

Contents

1 Introduction and Scope **4**

1.1 The Structural Failure of Traditional Instruments in High-Friction Environments 4

1.2 What MIRA Is: General Overview 5

1.2.1 Phase 1: Preference Without Information 6

1.2.2 Phase 2: The Informational Shock 6

1.2.3 Phase 3: Reordering 6

1.2.4 Phase 4: The Conversation 7

1.3 Why It Works: Foundations in Behavioral Economics 7

1.4 Strategic Advantages of MIRA 8

1.5 Conceptual Framework: From Constraint to Decision Revision 9

1.6 Research Questions and Hypotheses for Future Validation 10

2 Physical Artifact Design and Operating Principles **11**

2.1 Design Principles and the Physical Artifact 11

2.2 How the Evaluated Elements Are Selected: The Weighting Criterion 11

2.3 A Dual Function, and Why the Format Varies Across Deployments 12

2.4 What the Literature Says About Data Physicalization 12

3 How It Is Deployed in the Field **12**

3.1 Control in Uncontrolled Environments, and the Field Team 13

3.2 How Participants Are Selected, and Social Desirability Bias 13

3.3 Documentary Record: Consented Photography and Video, Without Interview Structure . 14

4 How the Data Is Processed **14**

4.1 Cross-Referencing Cases 14

4.2 How the Conversations Are Coded 15

4.3 Candidate Measures: Cognitive Delta and Its Components 15

5 An Extra, Low-Cost Look: Measuring Learning **17**

5.1 A Candidate Index: The Perceptual Correction Index 18

5.2 Toward Causal Attribution: A Proposed Validation, Not an Obtained Result 19

5.3 Why It Is a Recursive, Low-Cost Tool 21

6 Limits of the Method and How to Extend It **21**

6.1 One Mechanism, Many Contexts: From the Household Basket to Community Governance 21

6.2 Two Routes Toward Representativeness 22

6.3 Future Applications: Governance, Environmental Assets, Organizational Culture, and Digital Product 23

7 Application Case: Empirical Origin and Multi-Case Pilot Deployment **25**

7.1 Deployment Structure 25

7.2 What Varies by Territory, and Why That Is the Finding, Not the Noise 26

7.3 The Priority Hierarchy: Quantitative Evidence by Intervention 27

7.4 What Repeats Across the Three Cities, Despite the Diversity 28

7.5 Cognitive Delta Observed in the Field: How to Read It and What Explains It 29

7.6 Three Notions of Price "Truth": Official, Locally Experienced, and Transactional 32

7.7 Territorial Heterogeneity and the Limits of Generalization 32

7.8 Sampling, Selection Mechanism, and External Validity 33

8 Conclusions **34**

1. Introduction and Scope

1.1 The Structural Failure of Traditional Instruments in High-Friction Environments

Conventional data-collection instruments —grounded in the one-directional logic of the surveyor and the form— operate under the premise of a controlled environment. That premise collapses in the face of the friction inherent to real transactional settings: high subject turnover, scarce time, and a declarative reactivity that distorts the response before it reaches the analyst. Three systemic failures explain this collapse.

The first is temporal friction: a person rarely has more than a couple of minutes available, and traditional instruments tend to demand considerably more, condemning conventional surveys to critical drop-off rates or to automatic responses with no real deliberation.

The second is the instrumental literacy barrier: Likert-type scales (for example, 1 to 5, from "strongly agree" to "strongly disagree") assume the subject is familiar with that language. That familiarity cannot be assumed in contexts where handling numerical or statistical scales is not an everyday competency.

The third, and the hardest to resolve, is reactivity. The moment a subject perceives they are "being measured" (a recorder, a visible form, someone taking notes), they adjust their response to sound more reasonable or more acceptable to whoever is asking. This phenomenon is documented in the literature as social desirability bias: the tendency to respond with what is believed to be socially correct rather than what is actually thought (Fisher, 1993), one of the most consistently documented threats to the validity of any instrument that depends on a response declared out loud.

These three frictions are not the end of the problem. The same temporal and scale friction that renders the traditional survey unworkable in the field is, to a large extent, the ground on which an even more problematic alternative has thrived: mass digital collection. Much of contemporary digital infrastructure captures human behavior continuously and opaquely, often without genuine informed consent, in what the literature has documented as data extractivism (Chagnon and Hagolani-Albov, 2023) and as "surveillance capitalism" (Zuboff, 2019): the conversion of human experience into raw material for a behavioral prediction market over which the subject has no visibility or control.

MIRA situates itself deliberately at the opposite end of that logic, and not merely as a statement of principle: a physical, anonymous artifact, with no persistent digital record, leaves no substrate on which an algorithmic traceability system could operate. The session ends when the interaction ends, and no accumulable, resellable, or cross-linkable profile remains. Section 1.2 details how this design responds, point by point, to the four frictions described in this section.

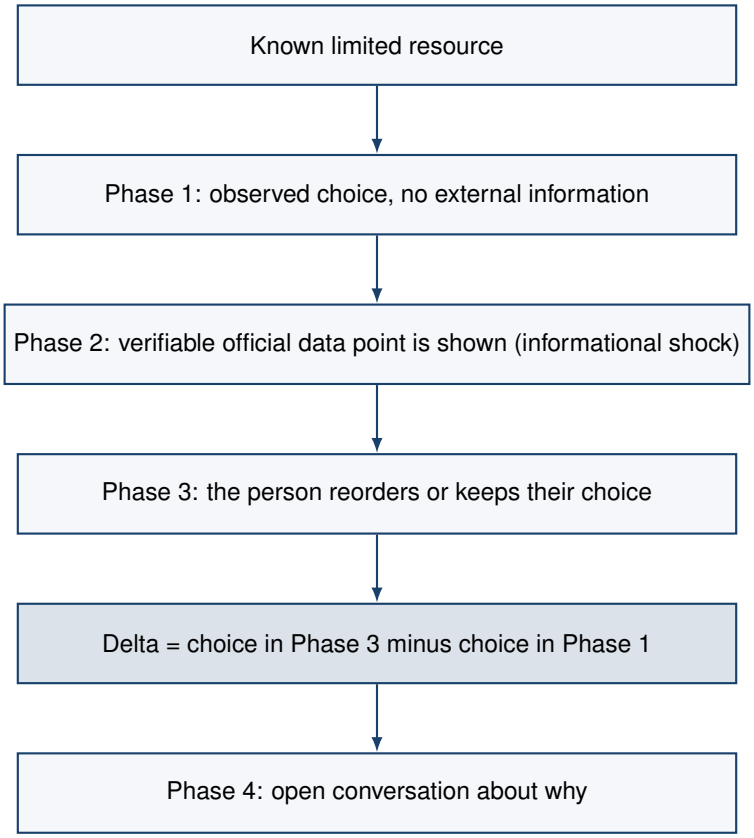
Before continuing, it is worth being explicit about the scope of this contribution, because it is easy for a new protocol to generate expectations that the available evidence does not yet support. This article introduces a field protocol (MIRA) and documents its operational implementation in a pilot study; it proposes a conceptual framework that distinguishes belief, stated priority, and behavioral choice, and situates MIRA within the literature on belief updating and corrective information; it proposes two candidate measures —Cognitive Delta and the Perceptual Correction Index— and explicitly lays out the psychometric validation agenda still required to treat them as established instruments; and it reports pilot evidence of operational feasibility, territorial heterogeneity, and observational patterns that generate hypotheses for future research.

This article, by contrast, does *not* establish that MIRA causes decision changes, does not establish the psychometric validity of its proposed measures, does not offer representative evidence of any population, and does not demonstrate a universal psychological mechanism. Each of these claims would require a design the current pilot does not have: a comparison group, random assignment, probabilistic sampling, and independent construct validation. Section 5.2 explicitly proposes the design needed to take those steps.

1.2 What MIRA Is: General Overview

MIRA stands for **M**easurement through **I**nteraction, **R**evision, and **A**nalysis. The name does not map, word for word, onto each of the four protocol phases described below; instead, it captures the spirit of the whole process: the subject interacts with a physical object and reveals a genuine preference, that preference is revised against verified evidence, and the subsequent analysis of the full sequence extracts the decision mechanism. The intermediate phase, a controlled informational shock, remains implicit in the interaction, because that is where the subject stops interacting only with the object and begins interacting with the data point itself. (The protocol originated under the Spanish name *Método de Interacción, Revisión y Análisis*; this document uses the English acronym expansion throughout.)

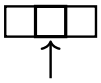
The protocol always runs through the same sequence of four phases, regardless of the application domain, and that order is non-negotiable: the validity of the captured data depends on it. Before going into the detail of each phase, it is worth seeing the complete process at a glance.



This scheme works regardless of how many elements or categories are at play, and without depending on which domain is being studied. Each phase is described below in greater detail, along with the

standardized pictogram that identifies it in any field material, independent of the physical artifact chosen for a given deployment.

1.2.1 Phase 1: Preference Without Information



The person interacts with the artifact without having seen any external data yet. This is where their priority or initial perception is revealed: what they rank first, what they discard, what they classify as important. Consider a simple example to fix the idea: a person is shown four tokens corresponding to four products from their usual basket, and is asked to rank them from most to least important, with no additional data.

1.2.2 Phase 2: The Informational Shock



Following the example above, the person is shown the data point corresponding to the elements they chose in Phase 1. This data point cannot come from just any external source: it must be a quantitative, macro-level figure, published by a verified official statistical authority (a national statistics institute, a central bank, or an equivalent reporting body depending on the application domain). This requirement is not incidental: the informational shock works precisely because it contrasts the subject's subjective experience against a figure whose authority and credibility are, in principle, unquestionable, which makes it possible to read with precision not only whether the person adjusts their choice, but how much credit they grant an official source relative to their own lived experience. Data from an informal or unverifiable source degrades this reading, because it introduces a source-trust variable that gets confounded with the variable of interest (sensitivity to the information itself).

It is essential that this phase occur after Phase 1, never before, because if the data point is shown first, it contaminates the original perception and it becomes impossible to know what the person thought before seeing the evidence (this is known as anchoring bias).

1.2.3 Phase 3: Reordering



The person is asked whether, with the new data point in hand, they keep their original choice or change it. This is where what we call the **Substitution Unit** is measured: the pair formed by what the person discards and what they choose in its place, along with what they decide to keep despite the new data (which is also valuable information: it indicates that element resists change).

1.2.4 Phase 4: The Conversation



Finally, in a brief, open conversation, the person is asked why they kept or changed their ranking. This phase is what makes it possible to distinguish between two very different types of reason, something developed in more detail in the data-processing section: some people do not change because they do not fully trust the new data, and some do not change because, even believing the data, something else matters more to them (a tradition, a bond, an identity).

The difference between what the person chose before and after seeing the data point is what this document calls **Cognitive Delta**: a measure of how well-calibrated the person's perception was relative to the official evidence. A small delta indicates the person already had an accurate sense of reality; a large delta indicates a gap between perception and evidence, and the Phase 4 conversation is what explains that gap: whether it is because the person lacked the correct information, or because, even having it, they decided not to act on it for other reasons, such as a tradition, an identity, or a bond. Section 4 formalizes this metric in greater mathematical detail.

This sequence directly addresses the problems described in the previous section, though it is worth being precise about the scope of each improvement rather than presenting them as complete solutions. Being a physical interaction lasting only a few minutes, it reduces temporal friction. By not depending on numerical scales, it reduces the literacy barrier. By centering the person's attention on the physical object rather than on the fact of being evaluated, MIRA seeks to reduce reactivity, though this pilot has no comparative evidence confirming by how much, relative to a conventional instrument. And by recording no identity and generating no persistent digital trace tied to a person, MIRA departs, by design, from the mass, opaque collection model described at the close of the previous section.

1.3 Why It Works: Foundations in Behavioral Economics

There is a structural reason, not merely a practical one, why this four-phase sequence captures something a closed-ended questionnaire does not, and it comes from contemporary behavioral economics.

The starting point is the idea of bounded rationality: people do not evaluate every option perfectly, but instead apply reasonable shortcuts given the time and information available to them (Simon, 1955), an idea that gave rise to an entire line of research on heuristics and biases in decision-making under uncertainty (Kahneman and Tversky, 1979). Resource scarcity does not merely restrict a person's options; it consumes their cognitive bandwidth and activates what Mullainathan and Shafir (2013) document as survival heuristics: under economic pressure, the mind automatically prioritizes the urgent over the optimal, and a consumption decision stops being a simple price calculation and becomes a negotiation between identity, tradition, and immediate need. Building on that same foundation, Thaler and Sunstein (2008) show that decisions do not occur in a vacuum but within an architecture: the order in which information is presented, who delivers it, and at what moment, determines whether an objective data point manages to move behavior or has no effect. This is precisely why Phase 2 must occur after Phase 1 and never before, as explained in the previous section.

This has a direct implication for reading the cognitive delta defined in the previous section. If a person does not adjust their choice in Phase 3 despite seeing the correct data, the superficial reading ("this

person is stubborn," "they don't care about data") wastes the method's most valuable signal. The COM-B model formalizes a more precise reading in three components: it could be a problem of Capability (they did not process the data), Opportunity (the context did not allow them to act), or Motivation (they processed the data and, even so, decided not to act because something else outweighed it). Distinguishing between these three scenarios is exactly what Phase 4 is designed to capture, and it is why MIRA does not settle for measuring the final outcome (changed or did not change), but documents the full mechanism.

This same logic is what supports why a small number of cases worked through in depth can contribute something that no increase in sample size can achieve on its own: isolating which of the model's three components is where the decision is getting stuck.

The method places its trust as much in what the person *does* (chooses, reorders, discards) as in what they *say*; Section 1.5 specifies the limits of this theoretical connection. In economics, the revealed-preference theory tradition holds that a real choice, made under a concrete constraint, says more about what someone actually prefers than any response declared in an interview (Chambers and Echenique, 2016; Ben-Akiva et al., 2019). This idea is not merely theoretical: it has even been tested in primates. In a study with rhesus monkeys, the animals repeatedly chose between different combinations of two juices, and their choices followed patterns consistent enough to be described with the same mathematical tools used to model human consumer behavior (Pastor-Bernier, Plott, and Schultz, 2017). MIRA draws on that same logic —trusting observed choice more than self-report— without automatically assuming it meets the formal consistency and stability conditions the theory requires; later work further shows how choices of this kind can be analyzed even when the person is not perfectly rational, but instead decides using the heuristics and shortcuts described at the start of this section (Halevy, Persitz, and Zrill, 2018; Tipoe, Adams, and Crawford, 2021).

1.4 Strategic Advantages of MIRA

Against the traditional survey, MIRA does not compete on sample volume: it competes on mechanism depth, context specificity, deployment speed, cost, and a set of positive externalities that no purely quantitative instrument generates by design. Each of these advantages deserves precision, not just enthusiasm.

- **Mechanism richness and context specificity:** That MIRA, in its base form, is not designed to produce a figure representative of an entire population is a design decision, not a limitation to apologize for. Its strength is isolating why and how a decision moves, and characterizing the particular context in which it occurs —territorial, cultural, situational— with a level of detail that a standardized instrument, applied identically everywhere, cannot capture without losing precisely that specificity.
- **Ecological validity:** The person decides in their real decision-making environment, not in a laboratory or on a phone call, which reduces the distance between what is measured and what actually happens when no one is measuring.
- **Deployment speed:** This architecture rests on a concept extensively studied in qualitative research: theoretical saturation, that is, the point at which continuing to collect information stops

contributing new categories and the pattern is already sufficiently understood to make an operational decision (Guest, Bunce, and Johnson, 2006; Guest, Namey, and Chen, 2020). The literature reviewed solidly supports two of these three assets —mechanism richness and ecological validity— as a direct consequence of working with saturation rather than representativeness. The third, deployment speed, is a reasonable inference rather than an already-established conclusion: the available evidence supports it only indirectly, through adaptive monitoring methods that allow deciding, during collection itself, when to stop, rather than fixing an arbitrarily large sample size in advance (Guest, Namey, and Chen, 2020; Lim, 2025; Rahimi and Khatooni, 2024; Ahmed, 2025).

- **Low cost:** The complete deployment, documented in Section 7, was executed with a fraction of the budget an equivalent representative survey would require, without sacrificing depth of findings.
- **Pedagogical externality and interactive civic education:** Beyond the savings, the interaction itself has a secondary effect documented in the literature on participatory methods: when the collection process is dialogic and not purely extractive, it benefits the participant as well, not only the researcher. Löhr, Weinhardt, and Sieber (2020) document this directly for a comparable method, the "World Café": its participatory nature facilitates dialogue and mutual learning among participants, generating genuine motivation to participate that a closed-ended questionnaire does not produce. In MIRA, that same dynamic operates in every deployment: the Phase 2 informational shock is not just a methodological treatment variable; it is also, for the participant, a moment of literacy about their own economic or decision-making context. The person leaves the interaction with verified information they did not have upon arrival.
- **Attractive design and organic narratives:** The tactile, gamified nature of the physical artifact, described in detail in Section 2, generates a level of engagement that a printed form does not achieve, and Phase 4 produces spontaneous narratives, in each participant's own words, rather than responses boxed into categories predefined by the researcher.

1.5 Conceptual Framework: From Constraint to Decision Revision

MIRA does not take the person or the household as its unit of analysis, but rather the **moment of decision**: the instant in which someone, under a known limited resource, reveals a priority, receives new data, and adjusts (or does not adjust) that priority. This shift in focus is deliberate: what matters is not so much who decides, but how the decision is made. To study that moment with precision, MIRA relies on a simple conceptual sequence that nonetheless requires carefully distinguishing several constructs that the literature and everyday language tend to treat as interchangeable: *constraint* $\rightarrow$ *initial belief* $\rightarrow$ *verified information* $\rightarrow$ *belief updating* $\rightarrow$ *decision revision*. The central question organizing this article is: when does verifiable information correct a belief, and when does that correction translate —or not— into a revised decision under real constraints?

Answering that question requires separating three levels that MIRA captures at different moments in its sequence, and that this article deliberately treats as distinct constructs, not synonyms:

- **Belief:** what the person believes about an externally verifiable fact; for example, how much they believe a product's price increased. This is what Phase 1 reveals, before any informational shock.

- **Stated priority (*preference* / *priority*):** the order or relative importance the person assigns to different options under their limited resource; for example, a ranking or a selection of products within a physical artifact.
- **Behavioral choice (*behavior* / *choice*):** what the person actually does after receiving information; for example, whether they keep, substitute, or drop an option in Phase 3.

Treating these three levels as though they were the same construct —assuming that a corrected belief automatically implies a revised priority, or that a revised priority automatically implies a future behavior change— is exactly the kind of conceptual collapse this article seeks to avoid. MIRA observes all three levels within a single interaction, which makes it possible, in principle, to distinguish them empirically; but distinguishing them in the design is not the same as having validated them as separable constructs, a task that falls to the validation agenda proposed in Section 5.2.

On the terminology of preference: this document deliberately avoids calling every choice MIRA observes a *revealed preference* in the formal sense of economic theory. Revealed-preference theory (Chambers and Echenique, 2016; Ben-Akiva et al., 2019) requires conditions of consistency, transitivity, and stability that a single-interaction field protocol is not positioned to verify. For that reason, this article prefers the more neutral term *observed choice* or, when referring specifically to the change between Phase 1 and Phase 3, *choice revision*, reserving *revealed preference* solely for the theoretical connection that motivates the design (Section 1.3), not as an automatic equivalence with what the protocol measures.

1.6 Research Questions and Hypotheses for Future Validation

Building on the framework above, this article proposes a set of hypotheses that the pilot described in Section 7 did not confirmatorily test, but that this same pilot helped generate, and that are explicitly formulated here for future validation (Section 5.2):

- **H1.** Exposure to verified official information reduces the discrepancy between perceived belief and objective condition (a reduction in Belief Error, defined in Section 4.3).
- **H2.** A greater reduction in that discrepancy is associated with a higher probability of behavioral choice revision.
- **H3.** The effect of information on choice revision varies according to the type of constraint, the level of trust in the source, and the identity weight of the product or domain being evaluated.
- **H4.** Physical behavioral elicitation captures aspects of decision revision —in particular, the distinction between not adjusting due to lack of information and not adjusting due to motivational resistance— that a conventional self-report does not capture with the same resolution.

These hypotheses were not preregistered nor subjected to formal statistical testing in the current pilot; they are presented as hypotheses proposed for future validation, and the pilot is interpreted solely as the feasibility evidence and observational patterns that motivate formulating them in these terms.

2. Physical Artifact Design and Operating Principles

2.1 Design Principles and the Physical Artifact

MIRA rests on four simple design rules: short duration, a maximum of ten minutes per interaction, so as not to lose participation in high-turnover locations; no screens, because it requires no digital device or connection, which lowers the barrier to entry and does not alter the natural environment in which the person decides; no personal data, because no name or any identifying data about the person is recorded, which also helps reduce the feeling of "being evaluated"; and an unconditioned symbolic reward, given at the end of the interaction regardless of the response, so as not to contaminate the choice with an incentive that pushes in a particular direction.

On that basis, instead of a closed-ended question, MIRA proposes a manipulable interaction. There are three typical ways of building this artifact: a **classification board**, where the person places elements into simple categories (for example: up, down, unchanged); a **variable-size token system**, where the physical size of the token the person chooses represents how large they believe a change was, with no need to be able to read numbers or percentages; and a **fixed-capacity container**, where the person must choose a subset of elements within a fixed limit, forcing them to prioritize explicitly, just as they would with any constrained resource in real life.

2.2 How the Evaluated Elements Are Selected: The Weighting Criterion

No MIRA deployment presents the participant with the full universe of possible products or categories: doing so would saturate the interaction and dilute the physical artifact's capacity to capture a clear priority. The selection of which elements to evaluate in each deployment was not, however, an ad hoc decision left to the subjective judgment of whoever facilitated the exercise. It was made under an explicit technical criterion: the Consumer Price Index (CPI) weight corresponding to each item, which reflects the relative importance of household spending on that specific category within the official basket. Products with a higher weight within the domain of interest were the natural candidates for evaluation, under the assumption that a price shock in those items has a higher probability of generating an observable substitution decision.

That quantitative criterion was further calibrated with a qualitative criterion of territorial context: typical dishes and food-territoriality patterns documented in prior research that underlies the dietary guidelines of the Colombian Family Welfare Institute (ICBF). Cross-referencing the official spending weight with available territorial knowledge made it possible to adjust the list of products evaluated in each city to what actually forms part of the local diet and culinary identity, rather than imposing a generic list of nationally "representative" products. This design is not a minor methodological detail: it implies that MIRA, as implemented in the pilot, is not an instrument that can be deployed without a prior study of the intervention locality and, implicitly, of its sociodemographic and consumption characteristics. That prior study —light, but necessary— is an entry condition of the protocol, not an optional step.

2.3 A Dual Function, and Why the Format Varies Across Deployments

The artifact serves two functions at once. On one hand, it captures the data point. On the other, the mere act of classifying or discarding elements generates in the person an awareness of their own decision pattern. This does not affect the validity of the data, as long as the correct order is respected: the preference is captured first, with no information (this is what we call Phase 1), and only afterward is any reflexive effect activated.

MIRA, moreover, is not always applied with the same physical artifact: the sequence of phases is always the same, but the object changes from one deployment to another. This serves a design purpose: if the same decision pattern appears again and again with physically different artifacts, it is less likely that the pattern is merely an instrument effect. It is worth being precise, however, about what kind of evidence this produces. In the pilot deployments described in Section 7, the artifact was not the only variable that changed across interventions: the city, the population, the product evaluated, and the season also changed. When several conditions change at once, the observed consistency does not allow the artifact's effect to be isolated from these other factors, and for that reason this article does not call this evidence triangulation in the strict causal sense of the term, but rather **heterogeneous replication**: the observation of the same qualitative pattern under conditions that vary along multiple dimensions simultaneously, a form of robustness evidence that is different —and less conclusive— than a triangulation in which only the measurement instrument varies while the rest of the conditions remain constant.

2.4 What the Literature Says About Data Physicalization

The field of data physicalization studies physical, manipulable representations of information, as opposed to digital graphics on a screen. Studies in this field show that a tangible object is not merely a cheap version of a digital interface: it lowers the barrier to entry for people with little familiarity with digital or statistical tools, allows several people to manipulate the data at the same time (which facilitates conversation around it), and creates an emotional connection with the information that an abstract chart does not achieve in the same way (Jansen et al., 2015; Dragicevic et al., 2021; Bae et al., 2022). One of the proposed mechanisms for explaining this effect is *embodiment*: the physical object does not merely represent the data, but the person experiences it bodily, which reinforces retention and comprehension relative to a purely optical visualization (Zhao and Vande Moere, 2008). More recent research further shows that this type of artifact draws on the way the body and the senses participate in understanding complex information, not just sight (Djavaherpour et al., 2021; Hornecker et al., 2023; Engert et al., 2025). This foundation, from human-data interaction, is what explains why the choice of physical artifact (Section 2.1) is not an aesthetic detail, but a decision that directly affects the quality of preference revelation.

3. How It Is Deployed in the Field

3.1 Control in Uncontrolled Environments, and the Field Team

Each interaction records the level of environmental interference (noise, interruptions, foot traffic) as contextual information, not as something to be eliminated. That environment is, precisely, part of what gives the exercise its validity: the person is deciding in their real environment, not in a laboratory. What does stay fixed, even when the physical artifact changes across deployments, is the opening and closing verbal script for each interaction, so that variation comes from context rather than from whoever facilitates the exercise. The order of the sequence never varies either: perception is always captured before the data point, because reversing that order invalidates Phase 1 as a measure of uninformed preference, and makes it impossible to calculate the cognitive delta described in Section 1.2.

Each deployment was also carried out with a field team of five people on average, not a single facilitator. Of that team, at least two people carried out the systematic recording of responses independently and cross-checked them against each other at the close of each interaction, as a quality control on the transcription of what was observed on the physical artifact. This cross-checked recording was further complemented by the audiovisual backup described in Section 3.3: a photograph or short video of the artifact's final state that makes it possible to corroborate, after the deployment, that the systematized data corresponds to what the person actually did. This combination —double human recording plus audiovisual verification— is what makes it possible to treat this pilot's data as auditable case by case, not merely as the report of a single observer.

3.2 How Participants Are Selected, and Social Desirability Bias

In its base version, MIRA is applied to whoever approaches or is available at the location (known as convenience sampling), with no randomization or population quotas. This limitation must always be honestly declared in any report that uses this data. Later, in the section on limits and extensions, we show how to convert this sampling into a probabilistic one with relatively small design adjustments.

On social desirability bias, there is, moreover, a point that tends to be overestimated in the behavioral research market: not recording the person's name, and giving the dynamic a participatory rather than evaluative tone, helps reduce it, but does not eliminate it entirely, and no serious methodology on the market eliminates it completely.

The reason is simple: even without a visible form, there is still a person facilitating the exercise, face to face. And the mere presence of a human facilitator is, by itself, a well-known source of response bias (Tourangeau and Yan, 2007). Anonymity in the record does not erase that courtesy effect in front of someone who is watching. This problem is not exclusive to markets or plazas: it is documented in fields as different as well-being surveys (Caputo, 2017), substance use (Latkin et al., 2017), social identity (Brenner and Delamater, 2016), and environmental behavior (Zhu, Greene, and Dolnicar, 2024), which suggests it is a structural limitation of any self-report, not a defect particular to this method (Durmaz, Dursun, and Kabadayı, 2019).

For that reason, MIRA incorporates two practical adjustments to bound this problem, without claiming to fully resolve it: first, rotating the person facilitating the exercise across different sessions, so that no response pattern is tied to a single facilitator; second, letting Phase 1 (the initial choice) happen silently, meaning the person classifies or chooses with their hands before being asked to speak their choice out loud, thereby reducing the social pressure generated by verbalizing a response in front of another

person.

3.3 Documentary Record: Consented Photography and Video, Without Interview Structure

There is a third source of reactivity, distinct from the two above, that deserves separate treatment: the audiovisual recording of the exercise itself. Field experience showed that video-recording a participant under an interview structure (camera in front, directed questions, recorded response) reintroduces much of the reactivity that MIRA's physical design seeks to avoid: the person stops reacting to the artifact and starts reacting to the camera.

The solution is not to eliminate the recording, but to change its function. MIRA distinguishes between two purposes that tend to get conflated in the design of field instruments: recording to *interview* (which reintroduces reactivity) and recording to *verify* (which does not, if done well). A photograph of the board with the final arrangement of tokens, or a short, consented video of the moment of interaction with no camera-directed question structure, serve a different function than a recorded interview: they make it possible to audit, after the deployment, that the data actually entered corresponds to what the person did, rather than to a memory reconstruction or a facilitator's transcription error. Without this minimum of documentary evidence, there is no way to distinguish a real ranking from an invented one at the moment of systematizing the data.

The operating criterion is therefore as follows: every deployment must leave at least a photographic record of the artifact in its final state, with explicit consent and without identifying the person; video recording, when used, is limited to documenting the physical interaction with the artifact, not to producing a structured interview piece. This distinction preserves the dual function described in Section 2.3 (measuring and prompting reflection) without sacrificing the ability to later verify that the systematized data has documentary backing.

This low-cost operating criterion does not close the door to future technological extensions. The same principle —recording to verify, not to surveil— could be scaled up with specialized computer vision or IoT (Internet of Things) sensors explicitly designed to preserve privacy: for example, cameras that process video locally and extract only the final arrangement of the tokens, without storing or transmitting any images of people, or proximity and weight sensors built into the physical artifact itself that automatically record the choice pattern with no need for any visual record. These extensions would increase the speed and scale of deployment auditing without compromising the anonymity principle described in Section 1.1, provided the processing occurs on the device itself (*edge computing*) and no identifiable data leaves it.

4. How the Data Is Processed

4.1 Cross-Referencing Cases

Before cross-referencing anything, it is worth specifying exactly what a **deployment** (also called an **intervention** in this document) is: a complete execution of the four-phase protocol, at a specific place and time, with a given physical artifact and a particular group of participants. Each deployment is,

in itself, a complete, self-contained case: it has its own territorial context, its own market or decision dynamic, and its own findings.

Multi-case analysis occurs when several deployments, carried out with the same methodological design and the same research objective but in different settings, are examined first separately and then together. Each deployment is first analyzed in isolation, looking for patterns specific to that case. The deployments are then compared with one another, looking for coincidences and differences, without assuming they are directly comparable in a statistical sense. That the same pattern appears across different artifacts and contexts is, in this approach, the main criterion of robustness: not statistical significance, but the theoretical saturation described in Section 1.4.

4.2 How the Conversations Are Coded

Phase 4 transcripts are read and coded openly: fragments representing an idea (codes) are identified, and then grouped into broader categories. How many times each code appears is used as an exploratory indicator of relevance, not as a figure that can be projected onto an entire population.

This frequency count is only the first level of analysis, and not the most important one. Reducing coding to counting repeated keywords risks losing precisely what MIRA is designed to capture: the reasoning behind the pattern. For that reason, each identified code is also reviewed qualitatively, reading the full fragment in context —not the isolated word— to understand the language, tone, and cultural associations with which the person justified their choice. Two participants may use the same keyword for opposite reasons (one out of tradition, another out of price calculation), and that distinction, invisible in a frequency count, is exactly what enriches the pattern with a real justification rather than an empty label.

4.3 Candidate Measures: Cognitive Delta and Its Components

Before defining any indicator, it is necessary to be explicit about the measurement scales involved, because mixing them without justification is one of the most common measurement errors in this type of protocol. In MIRA, three types of information coexist within a single interaction, and they are not directly comparable without an explicit transformation: the official data point (O_i) is typically a ratio variable (a percentage change with a meaningful absolute zero); the belief stated in Phase 1 and the choice in Phase 3 ($P_i^{(1)}$, $P_i^{(3)}$) are, depending on the artifact used, typically ordinal (a priority ranking, with no interpretable numerical distance between positions) or, at best, approximately interval, when the artifact uses magnitude tokens, under the unverified assumption that the person interprets token size in an approximately linear way.

Directly subtracting an ordinal variable from a ratio variable, as though both lived on the same scale, is a mathematically questionable operation unless the transformation that makes them comparable is documented. This article defines that transformation explicitly, rather than taking it for granted, and openly states that the specific transformation has not yet been psychometrically validated.

Formally, for each element i evaluated in a deployment: O_i is the verified official variation revealed in Phase 2; $P_i^{(1)}$ is the position or ranking assigned to the element in Phase 1; $P_i^{(3)}$ is the position assigned to the same element in Phase 3. When the physical artifact allows an approximate magnitude of belief to be captured, a transformation function $f(\cdot)$ is defined that converts O_i to the artifact's scale:

$E_i^{measured} = f(O_i)$. This function is specific to each artifact (for example, the mapping between percentage change and token size), is not independently validated in this pilot, and its choice directly affects the magnitude of the resulting indicator. The observed change between phases is defined as $E_i^{perceived} = P_i^{(3)} - P_i^{(1)}$, and the candidate indicator this article calls **Cognitive Delta** is defined as:

$$\Delta_i^{cog} = E_i^{measured} - E_i^{perceived}$$

Because Δ_i^{cog} combines a ratio-scale difference with a transformed ordinal-scale difference, this article presents Δ_i^{cog} as a construct under development, not as a validated index. Its exact quantitative interpretation —for example, whether a value of $\Delta_i^{cog} = 10$ carries the same substantive meaning across two different physical artifacts— depends on $f(\cdot)$ being correctly specified, something none of the pilot's five deployments (Section 7) independently tested. Its design logic, by contrast, has already operated across the five deployments (Section 7), and it is on that operational basis that the formal psychometric validation agenda —reliability, convergent validity, stability of $f(\cdot)$ across artifacts— described in Section 5.2 is built.

A More Conservative Reformulation: Belief Error and Correction

To avoid the scale-mixing problem, this article also proposes an alternative, more conservative formulation that explicitly separates belief (a Phase 1 construct) from behavioral choice (a Phase 3 construct), rather than treating them as the same quantity expressed on different scales.

When the artifact allows a numerical estimate of belief, **Belief Error** at moment t is defined as the absolute distance between stated belief and the official data point:

$$BE_{i,t} = |\text{Belief}_{i,t} - O_i|$$

and **Belief Correction** as the reduction in that error between the moment before and the moment after the informational shock:

$$BC_i = BE_{i,\text{pre}} - BE_{i,\text{post}}$$

Separately, and without assuming any fixed mathematical relationship with BC_i, **Choice Revision** is defined as a distance measure between the post-shock choice and the pre-shock choice:

$$CR_i = \text{distance}\left(\text{Choice}_{i,\text{post}}, \text{Choice}_{i,\text{pre}}\right)$$

and **Choice Persistence** as the binary indicator that the choice did not change:

$$CP_i = \mathbb{1}\left(\text{Choice}_{i,\text{post}} = \text{Choice}_{i,\text{pre}}\right)$$

This second formulation is, strictly speaking, more defensible than the original Cognitive Delta, because it does not combine a ratio scale with an ordinal scale within the same subtraction: BE and BC operate entirely within the space of belief (when it was captured numerically), and CR and CP operate entirely within the space of observed choice, without forcing an equivalence between the two. Whether BC_i predicts CR_i —that is, whether greater belief correction predicts greater choice revision— is precisely Hypothesis H2 (Section 1.6), and is an open empirical question, not an assumption built into the defini-

tion of the measures. This article presents BE, BC, CR, and CP as conceptual and methodological proposals, illustrated retrospectively with pilot data (Section 7.5) where the artifact used allows it, not as indicators with already-established evidence of reliability or independent construct validity.

What Distinguishes a Discrepancy From an Error

A high value of Δ_i^{cog} or of BE_i should not automatically be read as meaning the person "was wrong," nor as meaning the official data point is the individual truth that applies to that person. Section 7.6 develops this distinction in detail under the concept of *benchmark of truth*: the official data point is an aggregate reference value, not necessarily the price the person paid in their last transaction, and a discrepancy may reflect geographic heterogeneity, temporal mismatch, product differentiation, negotiation, or simple memory error, not only an information deficit.

With that caveat in place, the distinction that is genuinely useful, when Phase 4 is recorded in detail, is between two different explanations for the lack of choice revision despite an informational shock: that the person did not process the information (closer to what the COM-B model calls Capability), or that they processed it and, even so, decided not to act on it because something else outweighed it in the decision (closer to what that same model calls Motivation). There is evidence, outside this pilot, that people sometimes reject an objective data point not because they do not understand it, but because accepting it would conflict with something they value: this is known as identity-protective resistance to evidence (Kahan, 2013). MIRA does not measure these dimensions directly nor operate them as psychometric scales: it generates behavioral and qualitative evidence that, in subsequent analysis, can be coded using a COM-B-informed framework (Michie, van Stralen, and West, 2011), without this implying that the protocol identifies Capability, Opportunity, and Motivation as independently measured and validated constructs.

Cognitive Delta as a Calibration Signal: A Metaphor, Not an Established Capability

It is tempting to describe Cognitive Delta as a kind of "behavioral oracle" that automatically indicates where the available information flow is failing. This analogy is useful as design intuition —in the same way a data oracle feeds a system with external information, the indicator flags which elements show observed choice diverging from what the official data point would suggest— but it does not describe an already-demonstrated capability. There is no evidence in the current pilot that the indicator accurately predicts future decisions, nor that it reliably distinguishes an information problem from a motivational one without the direct support of the Phase 4 narrative, read case by case. Treating it as an automated, reliable signal in the literal sense would be exactly the kind of overclaiming this article seeks to avoid.

5. An Extra, Low-Cost Look: Measuring Learning

This section does not change MIRA's objective as described so far. It simply makes additional use of the same data the method already collects.

5.1 A Candidate Index: The Perceptual Correction Index

The sequence of Phase 1 (baseline), Phase 2 (informational stimulus), and Phase 3 (post-stimulus measurement) is, with no additional change to the design, what research calls a single-group, pre-post design: this makes it possible to estimate, with no need for a separate instrument, whether the exercise corrected the person's perception in some way. Building on that, it is possible, in principle, to calculate how much the gap between what the person believed (Phase 1) and the official data point (Phase 2) narrowed, using a post-shock re-estimate, when the artifact's format allows it. This article calls that calculation the **Perceptual Correction Index** (PCI), and presents it explicitly as a candidate index, not as a validated instrument: its formal definition, given below, immediately reveals several measurement problems that do not yet have a documented solution in this pilot.

For each element i, in addition to O_i (the official data point) and $P_i^{(1)}$ (the initial belief, in Phase 1), let $P_i^{(3')}$ be the re-estimate the person makes of the magnitude of the change, immediately after learning the official data point. The initial gap and the post-shock gap are:

$$B_i^{(1)} = \left| O_i - P_i^{(1)} \right| \qquad B_i^{(3')} = \left| O_i - P_i^{(3')} \right|$$

and the candidate index is defined as the proportion of that gap that closed following the informational shock:

$$PCI_i = \frac{B_i^{(1)} - B_i^{(3')}}{B_i^{(1)}}$$

This definition has at least four measurement problems that must be made explicit, rather than assuming the index behaves well in every case:

- **Indeterminacy when $B_i^{(1)} = 0$:** if the person's initial belief already exactly matched the official data point, the index is undefined due to division by zero. This case is not infrequent —it can occur precisely among the best-calibrated people— and a naive implementation of the index would exclude them or treat them as an error, biasing the sample toward those who started with a less accurate belief.
- **Interpreting negative values:** a negative PCI_i means the person moved away from the official data point after learning it. This could reflect active motivational resistance, but it could also reflect measurement noise in the re-estimate, fatigue, or a careless second estimate; this pilot has no way of distinguishing between these explanations.
- **Ceiling and floor effects:** someone with a very small initial gap has, by construction, little room to show a large correction (ceiling effect), while someone with an extreme initial gap can show a large proportional correction without that implying a meaningful absolute correction (floor effect). The index, as defined, does not distinguish between these scenarios.
- **Non-unified scale of B across artifacts:** $B_i^{(1)}$ and $B_i^{(3')}$ inherit the scale of $P_i^{(1)}$ and $P_i^{(3')}$, which, as discussed in Section 4.3, is not necessarily interval-level nor comparable across different physical artifacts. That two PCI values calculated with two different artifacts represent the same substantive magnitude of correction is an assumption, not an established fact.

None of these four problems invalidates the PCI as a conceptual proposal, but they do prevent it, given

current evidence, from being treated as an instrument with known psychometric properties. This article has no evidence of reliability (for example, test-retest stability) nor of convergent or predictive validity for the PCI, and presents it, accordingly, as a candidate index subject to future validation, not as an already-operational metric.

Unlike Cognitive Delta, which can be approximated with the data MIRA collects by default (Phase 1 and Phase 3), the PCI requires the additional numerical re-estimation step described above; without that step, the phenomenon can only be approximated qualitatively, based on whether the person adjusted their behavior in Phase 3 or not, as illustrated in Section 7.5.

As an optional extension, and at no additional instrument cost, a transfer test can be applied: showing the person an element that was not part of the original informational shock, and seeing whether the adjustment generalizes to that new element as well. This logic is not an invention of this document: it comes from the learning-transfer literature, which distinguishes between near transfer (applying what was learned to a situation very similar to the original) and far transfer (applying it to a situation that differs in context, format, or domain) (Barnett and Ceci, 2002). If there is correction on the original element and also transfer to the new element, that is a stronger signal that real learning of the underlying mechanism occurred, not just point memorization of the data point shown; if there is correction but no transfer, the adjustment was more likely specific to that data point. This test, like the PCI, is presented as a proposed procedure, not as a result already obtained in the current pilot.

5.2 Toward Causal Attribution: A Proposed Validation, Not an Obtained Result

None of the pilot's five interventions (Section 7) included a comparison group, random assignment, or a longitudinal structure with sufficient pre-shock observations. For that reason, no result reported in this article should be read as causal evidence: what the pilot documents is that a choice *changed after* exposure to official information, not that the information *caused* that change. Ruling out alternative explanations —learning within the interaction itself, demand characteristics induced by the facilitator, fatigue, or social desirability in response to the second question— requires comparing the observed outcome against what would have happened without the informational shock, something this pilot did not include. This section accordingly proposes how to incorporate that comparison into future deployments, without giving up the protocol's low-cost constraint, and without presenting any of these proposals as though they had already been tested.

Intra-deployment comparison groups, with random assignment (primary proposal): the most direct, lowest-cost extension is to introduce, within the same field session, a subgroup of participants randomly assigned (for example, alternating each participant, or by coin flip) to not receive the official data point in Phase 2, or to receive it with a deliberate delay, and to compare their subsequent choice with that of the group that did receive it. Because assignment to the subgroup would be random, this design —if implemented— would approximate the logic of a small-scale randomized controlled trial (RCT; Duflo, Glennerster, and Kremer, 2007), without requiring any instrument beyond those MIRA already uses. This design was not implemented in the pilot reported in this article; it is proposed here as the most economical and methodologically most defensible step toward future causal validation.

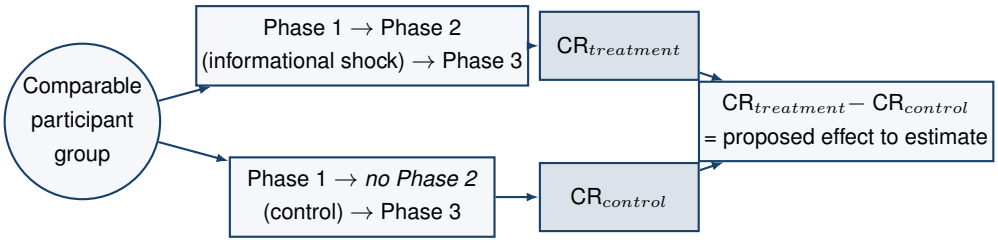


Figure 2. Proposed causal validation design (not implemented in the current pilot). CR = Choice Revision, defined in Section 4.3.

This comparison logic is the same one that, in the econometrics literature, underlies placebo-style tests. When random assignment within a session is not possible, and one resorts instead to comparing different deployments already carried out separately, it is worth being precise about which comparison is valid and which is not. It is not methodologically valid to compare the Phase 1 of one deployment with the Phase 3 of a different deployment: that would mix different people, in different contexts, with no way to isolate the effect of the informational shock from any other difference between the groups. The comparison, when it exists, must always happen within: Phase 1 versus Phase 3, within the same group of people or the same deployment.

Synthetic controls: a future methodological extension, not a solution already applicable to this pilot's data: when splitting a deployment's participants is not possible, the econometrics literature offers the synthetic control method (Abadie, 2021), which constructs a weighted combination of untreated units ("donors") that reproduces, before treatment, the behavior of the unit of interest, and uses permutation tests when the number of treated units is very small. This method requires, however, conditions the current pilot does not satisfy: a set of comparable donor units, a panel structure with repeated observations over time, and sufficient pre-treatment observations to fit the synthetic combination's weights. None of this pilot's five interventions generated a longitudinal structure of that kind. For that reason, this article does **not apply** synthetic controls to the existing data, and relegates this methodology to a future extension, conditional on a later deployment being explicitly designed with multiple units and repeated measurements over time.

Single-case designs: a future possibility with specific conditions, not an automatic property of repeated measurement: a third alternative, well established in clinical disciplines that face populations too small or atypical for a randomized trial, is the single-case experimental design (SCED). It is worth being precise about what this design requires, because having several measurements of the same person is not, by itself, enough to obtain causal inference: a valid SCED requires a stable baseline before the intervention, sufficiently dense repeated measurements, a controlled introduction of the intervention —ideally with multiple baselines or reversal designs where the phenomenon allows it— and explicit control of threats such as history effects, maturation, and carryover between repeated measurements (Krasny-Pacini and Evans, 2018). The current pilot implemented none of these elements: each participant was measured only once, in a single session. This article accordingly presents the SCED solely as a possible design for future validation, conditional on building the baseline and repeated measurements the method requires, and not as a result already achieved.

Structural limit common to these three routes: none of these extensions turns MIRA, as implemented in the pilot, into a design with already-established causal inference. MIRA starts from a convenience sample, in populations predefined by their location or context, not from population-level random assignment; this is not a defect that gets resolved by any of the extensions above, but rather a structural

feature of the current pilot, which Section 7.8 discusses in more detail under the problem of external validity. What this pilot does offer, and this is distinct from causality, is evidence that the proposed mechanism —revising a choice after learning an official data point— is observable and varies heterogeneously across contexts (Section 7), which is precisely what motivates the validation design proposed here.

5.3 Why It Is a Recursive, Low-Cost Tool

The advantage of this reading is that it requires no additional instrument: it is obtained with the same data MIRA already collects in its standard sequence. This allows it to be applied repeatedly, in every deployment, accumulating evidence without increasing the cost of collection. It is useful, for example, when someone needs to show, economically, that an exercise of this kind generated a measurable correction in people's perception, and not merely an interaction with no observable effect. When the objective requires, in addition to this mechanism-level reading, a figure defensible in terms of population representativeness, the next section describes how to adjust the sampling design without losing this same low-cost logic.

6. Limits of the Method and How to Extend It

6.1 One Mechanism, Many Contexts: From the Household Basket to Community Governance

MIRA's skeleton does not depend on the limited resource being money, nor on the decision being a purchase. This is one of this article's central ideas, and it is worth revisiting here, before discussing the method's limits, because those limits differ depending on the domain in which it is applied.

The deep structure is this:

Known limited resource → observed choice with no external information → a verifiable data point is shown → the person adjusts or does not adjust their choice → they are asked why

That same skeleton serves, for example, to observe how a community that collectively manages shared resources (including digitally organized communities) allocates its decision-making capacity when shown verifiable evidence about the actual state of those resources. Later in this same section we develop this application in more detail, including how it connects to verifiable digital-record infrastructure.

For this extension to be valid, and not a hasty generalization, three conditions must be met:

(a) There must be an official quantitative data point, from a verified statistical authority, against which to contrast the person's perception.

(b) The decision must be one of those that recur frequently and carry low or medium risk, not a single high-risk decision (such as a surgery or an irreversible life decision), because the ten-minute design is not meant to capture lengthy deliberations.

(c) The limited resource must be divisible into units the person can physically manipulate, such as tokens, cards, or a fixed-capacity container.

6.2 Two Routes Toward Representativeness

MIRA deliberately prioritizes understanding the mechanism over estimating a representative figure. Any use of its findings to speak about a broader population must honestly declare this limitation, and must avoid automatically extending a pattern seen in a small, non-representative group to the general reference population. This is precisely where the gap between the exploratory and the representative can be closed for real, not merely declared. There are two paths, and they are not mutually exclusive.

Statistical Route: Adjusting Sampling From the Design Stage

Instead of approaching only whoever is available at the location (convenience sampling), one can adopt a technique already developed in the social sciences for populations that gather at specific places and times: Time-Location Sampling (Magnani, Sabin, Saidel, and Heckathorn, 2005). This technique was created precisely for mobile populations that are difficult to fully enumerate, such as those who recurrently visit a public location.

It works as follows: first, every combination of day and time slot at which the population of interest tends to be present at the location is mapped. Some of those combinations are then randomly selected, giving higher selection probability to those that concentrate more foot traffic. Within each selected combination, every person meeting a systematic rule is approached —for example, every third person passing the point. Finally, results are adjusted according to how often each person visits the location, so as not to overrepresent those who go more frequently.

This last step is not new to this method: it is exactly the same logic already used to estimate how many people visit a location as unique visitors from the total visit flow, distinguishing those who go several times a week from those who go once a month. By incorporating this logic from the sampling design stage, and not merely as a post-hoc calculation, MIRA can become a probabilistic sample with known selection probabilities, which makes it possible to calculate real margins of error and confidence intervals, without having to change the instrument itself. The adjustment is one of sampling design, not of method.

Qualitative Route: Trading Number for Depth

The second route does not seek more people, but more depth per person. This involves extending Phase 4 with longer, semi-structured conversations, with a subgroup of participants, or even reconnecting with the same people on later visits, until reaching the point of theoretical saturation mentioned earlier. In practice, and for relatively homogeneous topics, that point is usually reached between twelve and twenty well-conducted conversations (Guest, Bunce, and Johnson, 2006).

Under this logic, a small number of cases, worked with sufficient depth, does not aim to statistically represent a population, but rather to produce what qualitative research calls transferability: the possibility that a pattern found in a specific context is recognizable and applicable in another similar context, even though it cannot be numerically projected (Lincoln and Guba, 1985; Drisko, 2025). It is, in other

words, a different form of rigor: not statistical representativeness, but contextual truth, sustained by the quality and depth of the narrative rather than by sample size. This is the same logic behind the choice architecture described in the introduction: understanding the "how" of a mechanism, in sufficient detail, leaves a recognizable trace in other similar contexts, even if it does not allow calculating how many people in the world share exactly that pattern.

Which of the two routes is preferable depends on the objective: if a defensible figure is needed for an audience that demands statistical representativeness, the time-location sampling route is the one indicated. If the objective is to thoroughly understand the reasoning behind an already-detected pattern, the qualitative-depth route delivers more value per unit of effort invested.

6.3 Future Applications: Governance, Environmental Assets, Organizational Culture, and Digital Product

This subsection deliberately and speculatively explores application domains that are not part of this article's empirical contribution. Nothing that follows was implemented or tested in the pilot described in Section 7; it is presented as an applied research agenda, conditional on the validation agenda in Section 5.2 first being completed in the original domain (consumption decisions under resource constraints) before extending to these other contexts.

As stated at the start of this section, MIRA's same skeleton (revealing a choice under a limited resource, showing verifiable data, observing the adjustment) could, in principle, be applied to decisions about time allocation, installed capacity, quotas, or attention, provided the three conditions already stated are met: an objective data point exists to contrast against, the decision is one that recurs frequently and carries low or medium risk, and the resource can be divided into manipulable units.

Community Governance and Collective Decisions

Under the same three conditions, MIRA's skeleton could, in principle, be applied to the design of decision mechanisms in communities that collectively manage shared resources, including those organized through verifiable digital records. In that hypothetical domain, the limited resource would become decision power or voting capacity that each person distributes across competing proposals, satisfying the divisible-manipulable-units condition; the informational shock would correspond to the verifiable state of the resources the community manages, satisfying the verifiable external data condition; and the recurring nature of collective decision cycles would satisfy the third condition.

Under this speculative application, MIRA would not measure only a person's stated voting intent, but something closer to what Section 1.5 calls choice revision in the face of auditable evidence: what part of that intent remains firm against new information about the actual state of the managed resources, and what part changes. This could be, in theory, an input for calibrating incentive mechanisms or decision rules sensitive to the quality of available information, although no current deployment has tested this.

A particularly concrete, though equally untested, application domain is that of conservation markets and digital environmental assets: carbon funds, payment-for-ecosystem-services schemes, or tokenized conservation certificates. These markets often face an accountability problem well known in the environmental-governance literature: "hard" impact metrics (hectares conserved, tons of carbon captured) are reported top-down, frequently with no validation of whether the local community per-

ceives that reported impact as matching what is happening on their land. MIRA could, in principle, serve that validation function: it would allow contrasting the official metric (Phase 2) against community perception (Phases 1 and 3), and Phase 4 could reveal whether the gap, when it exists, stems from a measurement problem, a communication problem, or legitimate distrust toward whoever administers the fund. This use, however, has not been implemented or piloted; it is presented here as an application hypothesis for decentralized autonomous organizations (DAOs) that govern this type of asset, not as a demonstrated capability.

A second speculative example, in an entirely different domain, is organizational culture. When a company introduces a new tool, policy, or way of working, the official metric (adoption rates, usage figures) rarely explains why one team adopted it enthusiastically and another resisted it. Applied to this context, MIRA could, in theory, help leadership anticipate the impact of an internal decision on its teams before rolling it out broadly, contributing a reading in both directions of the hierarchy rather than a purely top-down audit. None of these organizational applications has been tested; they are presented as possible extensions, not as results.

A third speculative example, increasingly relevant for digital product teams, is user experience (UX/UI) design and digital analytics. Conventional product metrics —click-through rates, screen time, conversion funnels— document what a user does, but rarely document whether that behavior reflects a genuine preference or an interface limitation the user does not know how to overcome. Applied to this domain, MIRA's skeleton could, in principle, be adapted to a controlled digital environment: revealing a user's priority among a set of design options, exposing them to verifiable data on each option's actual performance (load times, error rates, plan cost), and observing whether their choice adjusts. The difference relative to a conventional A/B test would be the same one that separates MIRA from a closed-ended survey: an A/B test measures the aggregate outcome of an already-made decision, while this hypothetical extension would additionally capture the narrative of why a user keeps an interface preference despite evidence that another option is objectively more efficient —a reading of product friction that conventional quantitative analytics is not designed to produce on its own. As with the other applications in this subsection, this extension has not been implemented or piloted.

Each MIRA phase would, moreover, have a hypothetical functional equivalent in a verifiable digital-record infrastructure, without this implying commitment to any specific technology or vendor:

- **Phase 1** would correspond to recording a decision intent, prior to any proposal under discussion being resolved.
- **Phase 2** would correspond to updating an externally verified data source, available after the initial intent has already been recorded, always respecting the order described earlier: perception first, then the data.
- **Phase 3** would correspond to a second decision record, after the verified data point is known.
- **Cognitive Delta**, if backed by future validation, could become an adjustable parameter within the decision mechanism itself —for example, a threshold that triggers additional review— although this depends entirely on the validation agenda in Section 5.2 being completed first.

The concrete technical implementation of this application (which technology to use, how to verify each record, how to sign each decision) falls outside the scope of this article, and should be developed, if the base protocol's empirical validation supports it, in a second-tier technical specification.

7. Application Case: Empirical Origin and Multi-Case Pilot Deployment

The protocol described in the previous sections is the formalization, by abstraction, of a field deployment already carried out: five interventions conducted between March 2024 and September 2025, in three Colombian cities, on consumption decisions under real price shocks. It is important to be precise about how this process unfolded, because that is what gives this section its evidentiary value. The interventions were not a spontaneous exercise from which MIRA "emerged" by chance: they were systematically designed, with impact-evaluation criteria established in advance (informational-shock sequence, before-and-after preference capture, deliberate space for open narrative), precisely so that, experimentally, those criteria could be refined into an extrapolable methodology for other contexts and make the most of the potential of these data-physicalization instruments. MIRA is the result of systematizing, after the deployment, the common principles that deliberate design made it possible to observe.

7.1 Deployment Structure

The five interventions were distributed as follows: three in traditional marketplaces in Bogotá, one in Cartagena, and one in Cali, with a total of 76 participants. Each intervention followed the same methodological sequence described in Section 1.2 (baseline preference revelation, informational shock with official price-variation data, reordering, and open conversation), but with a different physical artifact and research question in each city, following the heterogeneous-replication logic described in Section 2.3. Table 1 documents, intervention by intervention, the minimum operational information that allows auditing how each data point was generated: date, location, sample size, artifact used as treatment, the outcome measured, and the analytical rule applied to code it.

Intervention	Market / City	N	Artifact (treatment)	Outcome measured	Analytical rule
1 – Mar 19, 2024	Perseverancia, Bogotá	10	Binary board of 6 products (4 real, 2 ultra-processed); perceived up/down/unchanged, then official DANE data	Keep vs. substitute the purchase; narrative justification	Open coding; count of mentions by substitution type
2 – Jun 25, 2024	Las Ferias, Bogotá	17	Forced ranking of 20 products into 4 categories (High/Medium/Low), then official data	Priority reordering after the shock	Frequency table by priority tier; products with zero selection
3 – Dec 18, 2024	Perseverancia, Bogotá	22	Choice among 3 Christmas menus (traditional, mixed, ultra-processed); then aggregate variation for each menu	Menu fidelity vs. component substitution	Binary fidelity/substitution classification; coding of cultural vs. economic motive
4 – Apr 1, 2025	Bazurto, Cartagena	11	Selection of 3 of 8 regional products; perceived-magnitude token, then official data	Delta between perceived magnitude and official magnitude	Direct perception vs. official-data comparison, by product (Section 7.5)
5 – Sep 11, 2025	La Alameda, Cali	16	Selection of 5 of 20 "indispensable" ingredients in a fixed-capacity physical container; then removal after the shock	Frequency of inclusion and removal by product	Inclusion/exclusion count; coding of identity-based justification

Table 1. Operational structure of the five interventions that gave rise to MIRA. Each intervention was carried out in a single field session, with individual interactions of at most ten minutes per participant (Section 2.1).

The following photographic sequence illustrates the exact moment of physical interaction with the artifact: on the left, a participant places their token during Phase 1, having not yet seen any data; on the right, the same artifact reveals the verifiable official data point of Phase 2 and invites the Phase 3 reordering.

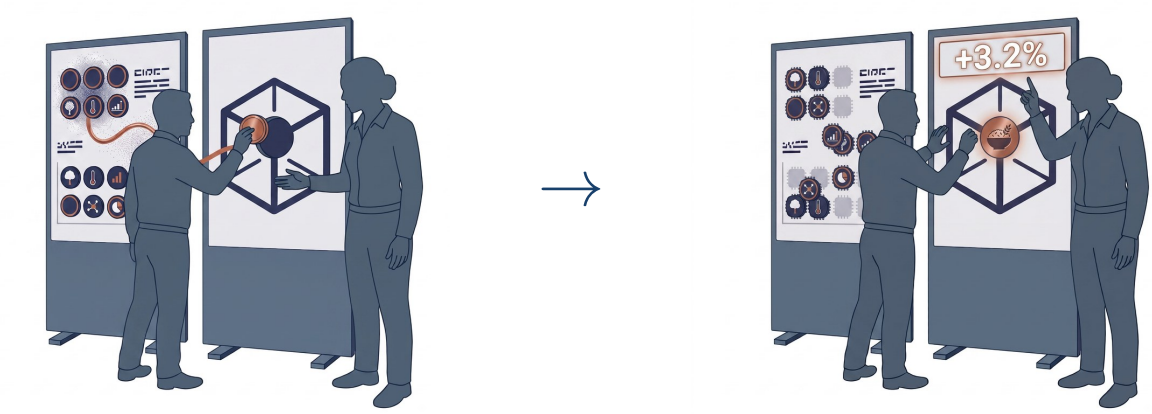


Figure 1. Phase 1, on the left: the participant reveals their preference with no external information. Phases 2–3, on the right: the verifiable official data point is revealed and the participant reorders their choice.

This diversity of format was not a limitation of the deployment: together with territorial diversity, it was the main source of the finding's robustness. A pattern that appears consistently across three cities, with three different physical artifacts and three different market dynamics, has a kind of robustness that no single-format survey, applied only once, can offer, even if applied to a thousand people.

On the handling of participants: no intervention recorded personally identifiable information. Participation was free and spontaneous, based on proximity to the intervention point, with the general purpose of the exercise explained verbally before each interaction and the option to withdraw at any time. Compensation was always given regardless of the responses provided, and no photographic or video record captured participants' faces, following the documentary-record protocol described in Section 3.3.

7.2 What Varies by Territory, and Why That Is the Finding, Not the Noise

A standardized quantitative instrument would have forced the same product list across the three cities, and would have missed exactly what this deployment captured: that the map of what is indispensable and what is substitutable is not the same in every territory, and that this difference follows an identity-based logic, not only an economic one.

Bogotá, the rice-egg-onion tripod, with potato as the adjustment variable: in Perseverancia, a traditionally working-class neighborhood with strong tourist appeal and a mixed socioeconomic profile, the ten participants in the first intervention revealed a consistent hierarchy of sacrifice: four of them explicitly mentioned reducing their consumption of meat or fish as their first response to rising prices,

two mentioned reducing fruits and vegetables, and one preferred to absorb the extra cost out of pocket rather than sacrifice quality; that last case was the owner of a small restaurant in the area, who insisted she could not remove protein or vegetables from her dishes, and instead chose to reduce a side dish. Onion and rice appeared in almost every list of indispensable products. In the second intervention, at the Las Ferias market —with a mixed socioeconomic profile, buyers who were mostly homemakers, and owners of neighborhood produce stands and restaurants looking for wholesale prices— this hierarchy was confirmed with a larger sample and a forced-ranking exercise across twenty products, detailed in Section 7.3: eggs, oil, and rice led the high-priority rankings, while salt, despite being a non-negotiable staple, appeared mostly in the low-priority tier, possibly because its relatively low cost keeps it off households' budget-worry radar.

Caribbean coast, plantain as the sovereign carbohydrate, fish as an identity frontier: in Bazurto, Cartagena's largest wholesale market, a road hub connecting the historic center with the industrial zone and functioning as the cradle of Champeta culture, with a clientele drawn mostly from strata 1 to 3, the same adjustment-carbohydrate role that potato plays in Bogotá was played by plantain, but with an inverse pattern: on the coast, plantain proved *less* substitutable than potato, not more. Participants described an explicit substitution hierarchy (meat is swapped for grains or eggs, fish is not swapped for anything, plantain is swapped for potato only if necessary), and several maintained that not even a price increase would make them reconsider fish, explicitly invoking their children's diet as the reason not to substitute it. Rising prices, moreover, appeared in some cases as a trigger for a healthier decision: one participant described having reduced their consumption of ultra-processed foods and increased that of fish, citing health reasons rather than tradition.

Cali, the sancocho pot as a cultural unit, with corn flour, beans, and chontaduro as an identity anchor: at La Alameda, a market that has become an icon of gastronomic tourism in the Valle del Cauca, with strong leadership by women in traditional cooking and herbal medicine, and that serves both working-class and upper-income residents of the city, the adjustment base was different once again: corn flour and red beans took the place occupied by rice and eggs in Bogotá. Chontaduro, a fruit with a strong regional identity, repeatedly appeared as a craving several participants refused to sacrifice, even while openly acknowledging that, in strictly nutritional terms, it was not indispensable.

No closed-ended questionnaire, applied identically across the three cities, would have registered this divergence without, at the same time, forcing participants to classify within categories that did not match their actual consumption reality. Territorial richness was not sacrificed in the pursuit of comparability: it was preserved, and the finding emerged from that preservation.

7.3 The Priority Hierarchy: Quantitative Evidence by Intervention

Beyond the narratives, three of the five interventions produced frequency data that make it possible to illustrate, with concrete figures and not only qualitatively, how food priority is distributed under budget constraint.

At Las Ferias (Bogotá, $N = 17$), the forced-ranking exercise across twenty products produced the following distribution of stated priority for the ten most frequently selected products:

Product	High	Medium	Low	Total
Eggs	6	3	3	12
Oil	5	1	6	12
Rice	6	3	2	11
Salt	2	1	8	11
Chicken breast	2	6	2	10
Onion	2	5	2	9
Plantain	3	2	2	7
Beef	4	0	1	5
Potatoes	1	2	2	5
Tomato	1	2	2	5

Table 2. Distribution of stated priority by product, Las Ferias intervention (Bogotá, $N = 17$). A given participant could prioritize up to ten of the twenty products evaluated.

The same intervention recorded three products —candy and confections, chocolate spread, and fried snacks— that no participant included among their ten priorities: a total absence, not merely a low priority.

At La Alameda (Cali, $N = 16$), the forced-selection exercise within a fixed-capacity container produced a complementary pattern: corn flour was chosen as indispensable by 6 of the 16 participants; red beans and chicken, by 5 each; and beef, potato, and carrot, by 4 each. Among the products initially included and then removed upon learning their price variation, fresh fish was the most frequent (removed by 5 of those who had selected it), followed by red beans (removed by 3, despite also being among the most chosen: a revealing duality, desired but not untouchable).

At Perseverancia (March 2024, $N = 10$), although the artifact's binary format did not produce a comparable frequency table, coding of the ten conversations recorded that 4 of the 10 participants spontaneously mentioned reducing their consumption of meat or fish as their first response to rising prices, and 2 of the 10 mentioned reducing fruits and vegetables, while no participant spontaneously mentioned reducing their consumption of rice or onion.

7.4 What Repeats Across the Three Cities, Despite the Diversity

At the same time —and this is what turns a collection of isolated case studies into a multi-case finding— certain patterns cut across territorial diversity without exception. Eggs, with 12 of 17 mentions at Las Ferias, emerged in all three cities as the universal protein substitute in the face of rising meat and fish prices. Ultra-processed products, without exception, were the first category to leave the basket under any price shock, regardless of the artifact used or the city: zero selections at Las Ferias, systematic removal in Cali, explicit mention of discarding them in Cartagena. In Cali, fresh fish was initially chosen by several participants as an indispensable ingredient, but 5 of them removed it after learning its price variation, while fish was almost never removed in the Cartagena intervention, illustrating the same territorial divergence documented in Section 7.2.

And in all three cities, Phase 4 (the conversation) revealed the same dual pattern that was later retrospectively coded in the vocabulary Section 4.3 proposes under a COM-B-informed framework: there were participants whose choice was revised as soon as they learned the official data point (closer to a Capability problem), and participants who, knowing the same data point, refused to adjust their choice because the product in question carried an identity weight that no inflation figure was going to

move (closer to a Motivation problem). This reading is a retrospective qualitative coding of the recorded narratives, not a psychometric measurement of the COM-B dimensions.

This last point deserves separate mention, because it was observed in the deployment before the technical vocabulary existed to name it. In the December intervention, centered on Christmas-dinner decisions, 8 of the 22 participants explicitly stated they would not change their traditional menu regardless of price ("I'm not changing my basket," "it's traditional, price doesn't matter"), compared with 5 who showed high price sensitivity and reordered their components immediately. That resistance to substituting traditional products was substantially greater than in the weekly-routine interventions, even though the price data shown was equally verifiable in both cases. That contrast, captured only because the protocol left room for open conversation, is precisely the kind of finding a purely quantitative instrument is not designed to produce: not a different figure, but an explanation of why the same information produces different effects depending on the symbolic context of the decision.

7.5 Cognitive Delta Observed in the Field: How to Read It and What Explains It

Unlike a form that would ask the person to mentally calculate a percentage, MIRA captures perception as a fluid conversation: the product is shown, the person is asked whether they believe it went up, down, or stayed the same, and by how much, with no exact figure required. This is not a limitation of the method relative to mathematical rigor; it is a design decision. Cognitive delta, as documented in this section, is not always an arithmetic subtraction between two numbers —though it can be, when the artifact captures a magnitude—; in its most common form, it is the contrast between a perceived direction and intensity (up, down, unchanged; a lot or a little) and the corresponding official data point. That contrast, obtained in seconds and with no cognitive friction, is exactly what makes it possible to weigh how a person ranks and decides under constraint, without turning the interaction into an arithmetic test.

One of the interventions, in Bazurto (Cartagena), incorporated this contrast exercise systematically. Table 3 aggregates responses from several participants for the products for which the corresponding official data, published by the national statistical agency, is available.

Product	Official variation	Stated perception (n)	Delta reading
Coconut	$+57.6\%$	"unchanged" (1/1)	Severe underestimation
Plantain	$+8.99\%$	"went down" (2/3), "unchanged" (1/3)	Underestimation, without exception
Cheese	-1.83%	"went up" (2/2)	Perceived direction reversed

Table 3. Contrast between stated perception and official data, aggregated by product, Bazurto intervention (Cartagena, $N = 11$).

None of these three patterns is better explained by assuming the person "was wrong." Participants themselves, in the Phase 4 conversation, offered explanations pointing to concrete, verifiable mechanisms, not a generic cognitive deficit.

The first mechanism is access to a trusted commercial network that cushions the list price relative to the price actually paid. The owner of a restaurant in the area explained it directly: "I have a restaurant here and I lived in Bogotá for a while, prices there in the capital are more stable, they don't vary as much as

here. The advantage here is that I know a lot of people, so I manage to buy everything because they give me good prices, and if things go up, I raise the price on the dishes." For someone with that kind of commercial relationship, the official plantain data point (+8.99%) simply does not translate into an equivalent purchasing experience: their plantain, in practice, did not go up, or went up much less than the city average. This same pattern —price protected by the relationship with the vendor— recurred among other participants with stable commercial ties at the market.

The second mechanism is selective attention according to a product's relative weight in the household budget. Coconut is, for most participants, a secondary ingredient relative to proteins such as fish or ribs, which concentrate weekly budget attention. It is reasonable that its price variation —even one as large as 57.6%— would go unnoticed when the person's attention is on the products that actually determine their week's balance. Coconut's severe underestimation is not, then, a random error: it is the direct consequence of that product not competing for the same attention as protein.

The third mechanism is exposure lag: someone who does not shop regularly, or who has not recently bought a specific product, has no way of having updated their belief, regardless of how well-informed they are in general. Several Bazurto participants explained their substitution decisions based on who in the household consumes each product and how often, which suggests the perception gap is distributed unevenly within a single household, not only between households.

A Second Example: Bogotá, Christmas Dinner

The December intervention at Perseverancia (Bogotá) offers a different contrast, because it revealed three official variations with very different signs and magnitudes —a marginal increase, a drastic increase, and a decrease— within the same conversation about the Christmas menu.

Product	Official variation	Stated reaction	Delta reading
Chicken	$+0.02\%$	"barely went up"	Precise calibration
Avocado	$+17\%$	"yes, it's very expensive"	Correct direction
Potato	$+86\%$	"I'd keep it, everything has potato"	Motivational persistence

Table 4. Contrast between stated perception and official data, Christmas-dinner intervention, Perseverancia (Bogotá, $N = 22$).

Chicken is the most interesting case of correct calibration in the entire pilot: before learning the official data point, one participant's spontaneous reaction already anticipated it was "a good option because it's barely gone up this season," a reading the official data point (+0.02%) confirmed almost exactly. Avocado shows correct-direction calibration, though not necessarily of magnitude: participants agreed it "is very expensive" before learning the exact 17% figure. Potato is the most revealing of the three cases, because it illustrates not a perception problem but a motivation one: several participants agreed they would keep it in their dinner "even though it went up a lot," and another was explicit about the reason —"everything has potato"— meaning they knew perfectly well the magnitude of the increase (86%) and decided not to substitute it anyway, precisely the kind of identity-based resistance to evidence that Section 4.3 formalizes as the motivational component, distinct from a purely informational one.

A Third Example: Cali, Three Types of Gap in a Single Intervention

The La Alameda (Cali) intervention did not capture official price variations by product, but its inclusion-and-removal exercise within the "pot" makes it possible to illustrate three distinct types of gap between an external criterion and the revealed decision, using three products from the same intervention.

Product	External criterion	Delta reading
Chontaduro	Dispensable in strictly nutritional terms	Purely motivational gap: "chontaduro is essential because it contains protein"
Fresh fish	Removed by 5 of 10 participants upon learning its price variation	Purely informational gap: the adjustment occurs as soon as the data is revealed
Potato	Perceived locally as one of the products that had fallen the most	Territorial gap: diverges from the national trend (+86% in Bogotá, same season)

Table 5. Three types of gap illustrated with products from the La Alameda intervention (Cali, $N = 10$).

Chontaduro is the motivational gap in its purest form: no price data is at stake, and participants nonetheless defend it as indispensable for nutritional reasons that do not hold up under strictly technical analysis, which can only be explained by its identity weight in Cali cuisine. Fresh fish, by contrast, is the informational gap in its purest form: removal occurs exactly when the data is revealed, with no associated narrative resistance. Potato illustrates a third type of gap that does not depend on the person at all, but on the territory: while in Bogotá potato rose 86% in the same season, several La Alameda vendors perceived it as one of the products that had fallen the most over the past year, a divergence that reveals not an individual perception error but a real difference in the Valle del Cauca's regional supply dynamics, consistent with the geographic heterogeneity documented in Section 7.6.

A Numerical Exercise: How the PCI Would Read on This Data

None of the pilot's five deployments captured the additional re-estimation step the formal PCI calculation requires (Section 5.1): asking the person, after revealing the official data point, how much they now believe the price changed. To illustrate the index's mechanics —not to report it as a real result— consider a hypothetical numerical exercise on the coconut case in Bazurto: if, before learning the data point, the person had estimated the variation at 0% ($P^{(1)} = 0$), and the official data point is $O = 57.6\%$, the initial gap is $B^{(1)} = |57.6 - 0| = 57.6$. If, after learning the data point, their re-estimate had moved to 40% ($P^{(3')} = 40$), the post-shock gap would be $B^{(3')} = |57.6 - 40| = 17.6$, and the resulting index would be $PCI = (57.6 - 17.6)/57.6 \approx 0.69$: a correction of approximately 69% of the initial gap, without reaching full calibration. This exercise does not describe a pilot finding; it describes how a future deployment, designed from the outset to capture re-estimation, would turn exactly this kind of conversation into a number comparable across people and products.

What these five examples —Bazurto, Bogotá, and the three Cali products— show together is that cognitive delta is not an indicator of individual error, but a tool for weighing, product by product, how sensitive a decision is to available information relative to factors no official figure captures on its own: the

quality of a person's commercial network, the relative attention a product receives within the household budget, and the cultural identity a food represents beyond its cost.

7.6 Three Notions of Price "Truth": Official, Locally Experienced, and Transactional

Before interpreting Table 3 as evidence of a perceptual "error," it is worth clarifying a conceptual point this article treats as a methodological strength, not as a limitation to apologize for: the official data point is not automatically the individual truth for each participant. It is worth distinguishing at least three different reference points: the **official reference**, the statistical figure published by the corresponding authority, aggregated at the city level or for a representative basket; the **locally experienced reference**, what the person reports having perceived in their own purchasing experience, at their usual stall or vendor; and the **transactional reference**, the price actually paid in the most recent concrete purchase, when that data is available, which may differ from both the official reference and the stated perception.

A divergence between stated perception and the official reference can stem from several distinct causes, and this article avoids automatically classifying it as cognitive error: geographic heterogeneity within the same city, temporal mismatch between the last purchase and the official measurement, product differentiation (variety, quality, packaging) not captured by the aggregate category, negotiated or loyalty-based prices not reflected in the list price, and the ordinary memory error of any self-report. In Bazurto, several participants explicitly described trust relationships with specific vendors that gave them access to prices different from the list price, which is consistent with a transactional reference systematically different from the official reference, not necessarily with a mistaken belief. Treating all of this divergence as an error attributable to the person would itself be a form of measurement bias: it would assume the only valid source of truth is the aggregate statistic, when the literature on price perception documents that these three references fail to coincide for legitimate reasons, not only due to cognitive failures. That field finding is, strictly speaking, the direct empirical precedent for the construct Section 4.3 formalizes mathematically, and the reason that same section warns against interpreting Δ_i^{cog} as an "error" in the evaluative sense.

The complete deployment, five interventions and 76 people, was also carried out with a fraction of the budget an equivalent representative survey would require, following the cost logic described in Section 1.4. The field team observed, anecdotally and not systematically, that revealing the official price data point against the participant's perception seemed to leave them with information they did not have upon arrival. This article has no before-and-after knowledge measurement that would allow quantifying that impression, and presents it, accordingly, as an informational-benefit hypothesis, not as a demonstrated pedagogical externality. Future validation of this hypothesis would require explicitly measuring the difference between post- and pre-treatment knowledge of the treated element and, where relevant, assessing whether that knowledge transfers to untreated products, in the sense of the transfer test described in Section 5.1.

7.7 Territorial Heterogeneity and the Limits of Generalization

What this deployment documents is not a population figure: it is a set of substitution and choice-revision patterns, observed consistently across three Colombian informal economies with different market logics. The psychology of scarcity described in Section 1.3 (Mullainathan and Shafir, 2013) is documented

in populations very different from this pilot's, which makes it plausible that the general mechanism is not exclusive to these three contexts; but this article has no direct evidence of that generalization, and accordingly avoids describing it as a universal mechanism. The formulation this article adopts is that the proposed pattern —choice revision following an informational shock, conditioned by structural and identity-based constraints— appears consistently across these heterogeneous Colombian contexts, not that its validity has been demonstrated for populations outside Colombia or outside the domain of food consumption.

The technical precision that must be stated here concerns magnitude, not the existence of the pattern: the convenience-sampling conditions described in Section 3.2 apply to this deployment, and no territorial proportion should be read as a figure projectable onto the general population. That distinction —a pattern replicated across heterogeneous contexts, a non-generalizable magnitude— is what separates this pilot from a representative survey, and is the entry point to the statistical-scaling route described in Section 6, for when the objective requires, beyond the qualitative pattern, the population magnitude.

7.8 Sampling, Selection Mechanism, and External Validity

This pilot's representativeness problem is not adequately summarized as "384 observations were needed and 48 were reached": that comparison, while informative, suggests the problem is purely one of sample size, and that increasing the number of participants under the same recruitment mechanism would resolve representativeness. That is not the case. The structural problem is the sampling mechanism: a convenience sample, recruited through spontaneous proximity to the intervention point, does not become representative of the general population simply by increasing its size, because those who voluntarily approach an intervention point may differ systematically —in time availability, willingness to interact with strangers, or interest in the topic— from those who do not. That systematic difference, not merely sample size, is the main threat to this pilot's external validity.

That said, this article also documents the population-sizing exercise that makes it possible to calibrate the magnitude of the gap, as a technical reference, not as a substitute for probabilistic sampling: using the aggregate monthly visit flow to Bogotá's district marketplaces (978,920 visits, Secretaría Distrital de Desarrollo Económico, 2019) and a visit frequency weighted between weekday and weekend shoppers, the population of unique visitors is estimated at approximately 147,650 people, a figure consistent with the 2.1% reported by an independent district study for a subset of three markets (IPES, 2018); a probabilistic sample of that population, at 95% confidence, would have required 384 observations, versus the 48 reached in Bogotá. This figure illustrates the magnitude of the gap, but is not, by itself, the full explanation for the representativeness problem: even with 384 observations recruited under the same convenience mechanism, the pilot would still not be inferable to the general population, for the selection reasons described above.

This pilot, accordingly, is explicitly declared not designed for population inference. The statistical-scaling route described in Section 6.2 —time-location sampling, which does incorporate a selection mechanism with known probabilities— is the proposed path for closing this gap in a future deployment, not something this pilot has already resolved.

That is the application case on which the MIRA protocol was subsequently formalized, and the empirical basis from which this article makes the case for the validation agenda proposed in Section 5.2.

8. Conclusions

MIRA is not proposed as a closed instrument, but as a field architecture useful for observing how people revise real decisions when evidence comes into tension with their beliefs, in short, low-cost interactions with high ecological validity. Its greatest strength lies in combining, within a single protocol, an initial choice, exposure to verifiable data, observable revision of the decision, and open narrative about the adjustment —something rarely found together in standard field instruments.

That design also responds to a paradigm shift the social and behavioral sciences have long been calling for: moving from statistical generalization as the sole standard of validity toward practical, actionable intelligence that complements it. Much of conventional impact evaluation starts from a vague prior diagnosis, without community validation, that statistically homogenizes heterogeneous populations and ignores the real decision-making mechanisms and social interactions that determine what actually matters to a targeted population. MIRA offers an alternative that complements, rather than replaces, that standard: data collected ethically and anonymously that gives whoever is designing an intervention an early, actionable reading of how the population they will work with actually decides, before —or alongside— committing resources to a larger-scale traditional impact evaluation.

Rather than producing only data, MIRA generates actionable learning: it makes it possible to detect decision patterns in context, refine interventions as they are deployed, and offer participants a moment of verifiable information that also carries pedagogical and civic value. The five-deployment, three-city pilot shows that this logic is already operationally viable, and that it can capture territorial heterogeneity and divergences between perception and official data that tend to get lost in more declarative instruments.

Given the current evidence, this usefulness has a defined scope that deserves to be stated with the same clarity as the claims above: the pilot included no comparison group or random assignment, and therefore does not establish causality; Cognitive Delta and the PCI remain, respectively, a construct under development and a candidate index, subject to the psychometric validation agenda in Section 5.2; convenience sampling precludes population inference, regardless of sample size (Section 7.8); and the observed pattern is documented across three heterogeneous Colombian contexts, not as a universal psychological mechanism (Section 7.7). None of these boundaries invalidates what the pilot does demonstrate: that the mechanism is observable, heterogeneous across contexts, and operationally cheap to capture.

The public-policy implications compatible with this evidence are, for now, methodological: MIRA can complement existing monitoring or evaluation instruments, offering an additional way to distinguish informational barriers from motivational ones, and a fast, low-cost diagnostic phase before committing resources to a larger-scale impact evaluation. This document invites two audiences to iterate on what is proposed here: those designing governance mechanisms for decentralized autonomous organizations, to consider the PCI and Cognitive Delta as possible inputs within their own decision flows (Section 6.3); and those doing econometric analytics and impact evaluation, to consider the factorial design proposed in Section 5.2 as a starting point for combining experimental data with the subjectivity of their own study groups as a design variable, not as noise to discard.

Taken together, MIRA offers an agile, replicable, and useful way of studying constrained decisions under real field conditions. Its value lies not only in measuring whether a decision changes, but in helping to understand when the change depends on information, when it depends on trust or identity, and how

that distinction can translate into better interventions, better impact measurement, and better incentive design across diverse social and community contexts. For that reason, more than a one-off experiment, MIRA is emerging as a promising methodological foundation for calibrating on-the-ground action, learning quickly, and extrapolating the protocol to new use cases where what matters is observing live decisions, not just declared responses —engaging, along the way, with other behavioral measurement protocols and with the interoperability standards that applied research continues to develop.

References

Abadie, A. (2021). Using synthetic controls: Feasibility, data requirements, and methodological aspects. *Journal of Economic Literature*, 59(2), 391–425.

Ahmed, S. K. (2025). Sample size for saturation in qualitative research: Debates, definitions, and strategies. *Journal of Medicine, Surgery, and Public Health*, 5, 100171.

Bae, S., Zheng, C., West, M. E., Do, E. Y.-L., Huron, S., & Szafir, D. (2022). Making data tangible: A cross-disciplinary design space for data physicalization. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (pp. 1–18). ACM.

Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn? A taxonomy for far transfer. *Psychological Bulletin*, 128(4), 612–637.

Ben-Akiva, M., McFadden, D., & Train, K. (2019). Foundations of stated preference elicitation: Consumer behavior and choice-based conjoint analysis. *Foundations and Trends in Econometrics*, 10(1-2), 1–144.

Brenner, P. S., & Delamater, J. (2016). Lies, damned lies, and survey self-reports? Identity as a cause of measurement bias. *Social Psychology Quarterly*, 79(4), 333–354.

Caputo, A. (2017). Social desirability bias in self-reported well-being measures: Evidence from an online survey. *Universitas Psychologica*, 16(2), 1–13.

Chagnon, C., & Hagolani-Albov, S. (2023). Data extractivism. https://doi.org/10.4324/9781003450160-14

Chambers, C. P., & Echenique, F. (2016). *Revealed preference theory*. Econometric Society Monographs. Cambridge University Press.

Djavaherpour, H., Samavati, F., Mahdavi-Amiri, A., Yazdanbakhsh, F., Huron, S., Levy, R. M., Jansen, Y., & Oehlberg, L. (2021). Data to physicalization: A survey of the physical rendering process. *Computer Graphics Forum*, 40(3), 569–598.

Dragicevic, P., Jansen, Y., Vande Moere, A., & Isenberg, P. (2021). Data physicalization. In *The Encyclopedia of Human-Computer Interaction* (2nd ed.). Interaction Design Foundation.

Drisko, J. W. (2025). Transferability and generalization in qualitative research. *Research on Social Work Practice*, 35(1), 102–110.

Duflo, E., Glennerster, R., & Kremer, M. (2007). Using randomization in development economics research: A toolkit. In *Handbook of Development Economics*, Vol. 4 (pp. 3895–3962). Elsevier.

Durmaz, A., Dursun, İ., & Kabadayı, E. T. (2019). Mitigating the effects of social desirability bias in self-report surveys. In *Handbook of Research on Social Desirability Bias* (pp. 146–185). IGI Global.

Engert, S., Peetz, A., Klamka, K., Surer, P., Isenberg, T., & Dachselt, R. (2025). Augmented dynamic data physicalization: Blending shape-changing data sculptures with virtual content for interactive visualization. *IEEE Transactions on Visualization and Computer Graphics*, 31(1), 7580–7597.

Fisher, R. J. (1993). Social desirability bias and the validity of indirect questioning. *Journal of Consumer Research*, 20(2), 303–315.

Gigerenzer, G. (2026). From bounded rationality to ecological rationality. *Industrial and Corporate Change*.

Gigerenzer, G., & Todd, P. M. (1999). *Simple heuristics that make us smart*. Oxford University Press.

Guest, G., Bunce, A., & Johnson, L. (2006). How many interviews are enough? An experiment with data saturation and variability. *Field Methods*, 18(1), 59–82.

Guest, G., Namey, E., & Chen, M. (2020). A simple method to assess and report thematic saturation in qualitative research. *PLOS ONE*, 15(5), e0232076.

Halevy, Y., Persitz, D., & Zrill, L. (2018). Parametric recoverability of preferences. *Journal of Political Economy*, 126(4), 1558–1593.

Hasan, M. M., Karim, F., & Ripon, M. B. B. (2025). Behavioral economics and consumer decision-making: An integrative review. *Business and Social Sciences*, 3(1), 1–8.

Hornecker, E., Hogan, T., Hinrichs, U., & van Koningsbruggen, R. (2023). A design vocabulary for data physicalization. *ACM Transactions on Computer-Human Interaction*, 31(1), 1–62.

Instituto para la Economía Social [IPES]. (2018). *Shopper Study of District Marketplaces* [Estudio de Shopper Plazas de Mercado Distritales]. Alcaldía Mayor de Bogotá.

Jansen, Y., Dragicevic, P., Isenberg, P., Alexander, J., Karnik, A., Kildal, J., Subramanian, S., & Hornbaek, K. (2015). Opportunities and challenges for data physicalization. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (pp. 3227–3236). ACM.

Kahan, D. M. (2013). Ideology, motivated reasoning, and cognitive reflection. *Judgment and Decision Making*, 8(4), 407–424.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2), 263–291.

Krasny-Pacini, A., & Evans, J. (2018). Single-case experimental designs to assess intervention effectiveness in rehabilitation: A practical guide. *Annals of Physical and Rehabilitation Medicine*, 61(3), 164–179.

Latkin, C. A., Edwards, C., Davey-Rothwell, M. A., & Tobin, K. E. (2017). The relationship between social desirability bias and self-reports of health, substance use, and social network factors among urban substance users in Baltimore, Maryland. *Addictive Behaviors*, 73, 133–136.

Lim, W. M. (2025). Sample size in qualitative research: Moving from data saturation to theoretical saturation. *Journal of Global Marketing*. https://doi.org/10.1080/08911762.2025.2590757

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Sage Publications.

Löhr, K., Weinhardt, M., & Sieber, S. (2020). The "World Café" as a participatory method for collecting qualitative data. *International Journal of Qualitative Methods*, 19, 1–15.

Magnani, R., Sabin, K., Saidel, T., & Heckathorn, D. (2005). Review of sampling hard-to-reach and hidden populations for HIV surveillance. *AIDS*, 19, S67–S72.

Michie, S., van Stralen, M. M., & West, R. (2011). The behaviour change wheel: A new method for characterising and designing behaviour change interventions. *Implementation Science*, 6, 42.

Mullainathan, S., & Shafir, E. (2013). *Scarcity: Why having too little means so much*. Times Books, Henry Holt and Company.

Pastor-Bernier, A., Plott, C. R., & Schultz, W. (2017). Monkeys choose as if maximizing utility compatible with basic principles of revealed preference theory. *Proceedings of the National Academy of Sciences*, 114(10), E1766–E1775.

Rahimi, S., & Khatooni, M. (2024). Saturation in qualitative research: An evolutionary concept analysis. *International Journal of Nursing Studies Advances*, 6, 100174.

Secretaría Distrital de Desarrollo Económico [SDAE]. (2019). *Results of the pedestrian-flow measurement in district marketplaces* [Resultados de la medición del flujo de personas en plazas de mercado distritales]. Alcaldía Mayor de Bogotá.

Simon, H. A. (1955). A behavioral model of rational choice. *The Quarterly Journal of Economics*, 69(1), 99–118.

Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: Improving decisions about health, wealth, and happiness*. Yale University Press.

Tipoe, E., Adams, A., & Crawford, I. (2021). Revealed preference analysis and bounded rationality. *Oxford Economic Papers*, 74(3), 882–901.

Tourangeau, R., & Yan, T. (2007). Sensitive questions in surveys. *Psychological Bulletin*, 133(5), 859–883.

Zhao, J., & Vande Moere, A. (2008). Embodiment in data sculpture: A model of the physical visualization of information. In *Proceedings of the 3rd International Conference on Digital Interactive Media in Entertainment and Arts* (pp. 343–350). ACM.

Zhu, O. Y., Greene, D., & Dolnicar, S. (2024). Should the risk of social desirability bias in survey studies be assessed at the level of each pro-environmental behaviour? *Tourism Management*, 103, 104933.

Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. PublicAffairs.