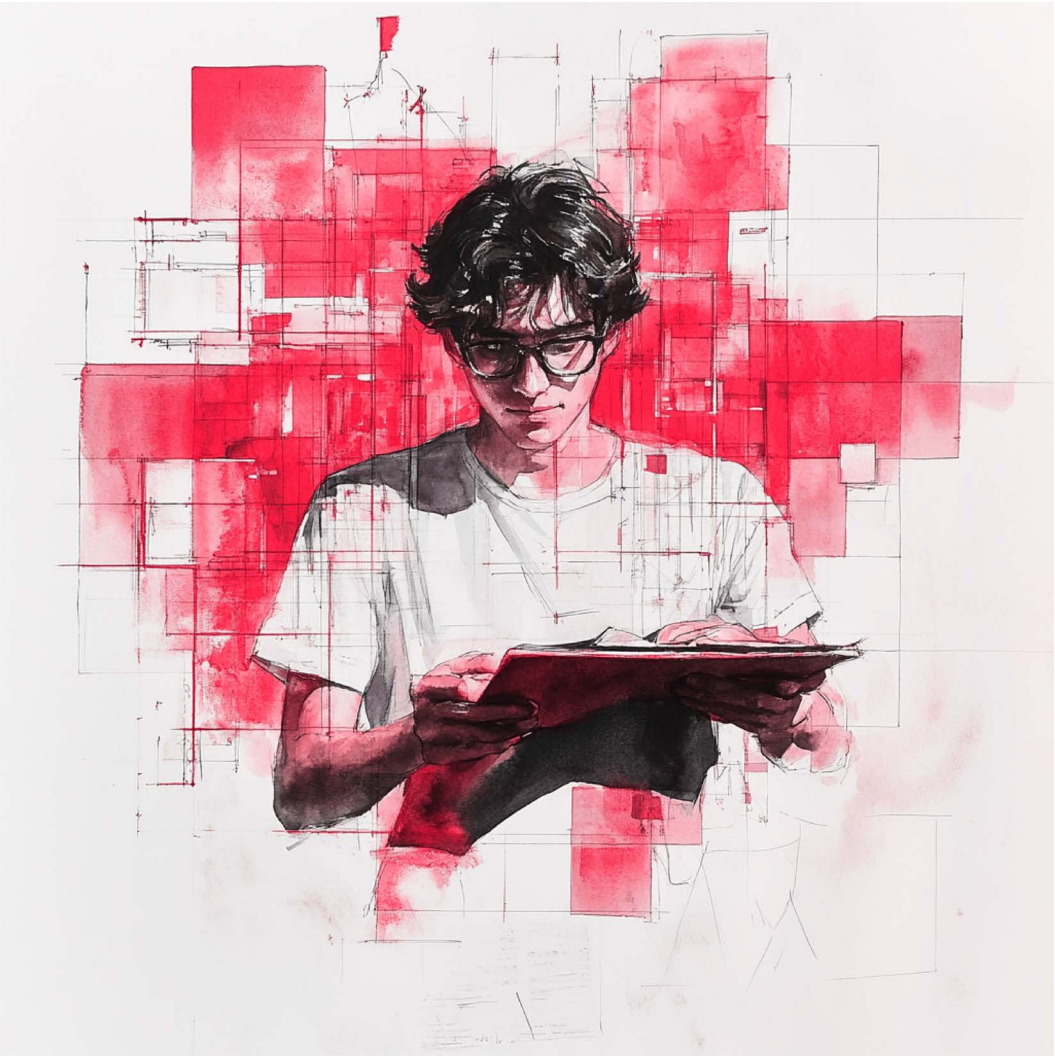
Summary of Leverage Introspection Research

Precursors, methods, and findings



Executive Summary

The mind, as an object of study, has challenged researchers for years. Originally, the mind was studied by philosophers, including Plato, Descartes, Spinoza, and Kant, whose treatises covered a range of topics, including the mind. Later attempts, such as by Freud, Titchener, Watson, and others, tried unsuccessfully to turn the study of the mind into a science.

The Leverage introspection research program, informed by a variety of sources, studied the mind using a variety of methods and techniques. Some of these were familiar, though refurbished, while others were new. The central activity involved eliciting verbal reports to create models of people's beliefs and goals. These models were then used, in combination with theoretical claims about the mind, to derive predictions about the results of interventions. The belief/goal models could then be improved in accuracy, as measured ultimately by success in predicting the outcomes of interventions meant to change observed behavior.

These methods yielded the identification of many patterns and the confirmation of others, with the result being an understanding of the mind that can be explained by analogy to the physical sciences of physics, chemistry, biology, and ecology. The physics of the mind involves a small set of basic terms and a small number of postulated universal rules. These rules yielded the centrality of *belief* in explaining behavior and led to patterns in belief and belief updating dynamics being the central objects of study.

Beliefs, according to the model developed, occur in repeating, recognizable patterns. The chemistry of the mind involves cataloguing these belief structures and formations, of which one of the most important are those that redirect attention. Patterns of attentional redirection yield a practical distinction between a conscious and subconscious mind; the subconscious mind was found to be vast, as was the belief system in general. Consciously accessible goal systems were found to be sufficiently simple to study; this corresponds to the biology of the mind. The mind in interconnection with other minds via high-bandwidth non-verbal communication then constitutes the ecology of the mind, with features with surprising explanatory power.

There are many potential applications for greater knowledge of the mind. Leverage's researchers focused on education and everyday mental health, finding the former the most tractable. Other topics were studied as well. Evidence for the system described will come from a variety of sources, the most stable being an ongoing data stream generated by new researchers who explore similar research avenues.

A New Approach to the Mind

The mind has been a particularly challenging object of study. The Leverage introspection research program, using new instruments and theory, managed to produce a stream of sufficiently reliable data to begin mapping out the human mind. This research, which took place between 2012 and 2019, yielded a picture of the mind as a system that operates according to simple rules which yield identifiable patterns at the micro- and macro-levels.

> Leverage's introspection research found the mind to be governed by simple rules.

The following exposition begins with a summary of previous research programs that studied the mind, then describes the primary methods used by Leverage's introspection researchers, then covers the central findings, organizing them in terms of an analogy to the physical sciences: physics, chemistry, biology, and ecology. Applications are then briefly covered as well as evidence, followed by suggestions for further reading.

Previous Approaches

Before describing the Leverage introspection research program, both its methods and results, it will be useful to provide some context about the study of the mind. The mind has been studied from many angles for many years; it is thus valuable to understand some of those approaches, so that one has some reference point from which to judge a proposed new approach. The following is a brief summary of research programs and traditions that focused on the mind, with notes about the advantages and disadvantages of each, especially as they relate to the Leverage program.

Mental philosophy. A variety of philosophers, from Plato to Kant, studied different aspects of the mind using different methods. This yielded a trove of useful observations and analyses. Picking through the sources, one can find the difference between beliefs (i.e., "ideas") and imagination in René Descartes' Meditations, the recognition of the cognitive element of emotion in Spinoza's *Ethics*, and the recognition of a non-conceptual space (i.e., a "form of intuition") relating objects of visual experience in Kant's *Critique of Pure Reason*. The main challenge, as is natural to philosophy, is that different readers will derive different lessons from a given treatise. The works are thus a valuable source of insight, but not the basis for a joint study of the mind.

Psychoanalysis. Sigmund Freud and his successors pioneered another approach to the mind, which involved analyzing verbal patterns and dreams for clues to the content of the unconscious mind. It was proposed, in particular, that the mind has discoverable structures which can be analyzed and altered in repeatable ways. This approach had the benefit of correctly identifying that the mind has intelligibly ordered mental content, only some of which is easily available to introspection. The primary drawback was that its methods did not yield agreement among practitioners as to the contents of a given mind.

Structuralism. Wilhelm Wundt and Edward Titchener developed a further approach, which was the use of introspection to identify the fundamental aspects of human experience. Taking inspiration from chemistry, introspectors were given careful instructions so they could break down their experiences into the smallest, unanalyzable parts. This approach, which was called "structuralism," had the merit of using introspection, though focused its use on conscious sensory experience, rather than on beliefs. Sensory experience is fleeting, however, which makes it difficult to study through introspection alone.

> Different approaches to the mind have encountered different challenges.

Behaviorism. Dissatisfied with various aspects of other research programs, John Watson and others proposed investigating the mind by examining behavior. In some cases taking inspiration from the study of animal behavior, "behaviorism" sometimes banned the mention of mental entities, such as thoughts or ideas. This approach had the virtue of grounding research in observables, and moreover, observables accessible to third parties. Mental entities, however, turned out to be ineliminable in most cases, with the result that behaviorism was soon abandoned as a research effort.

Modern experimental psychology. After the demise of behaviorism, the mainstream study of the mind split into two research programs, which became dominant in the second half of the twentieth century. The first is modern experimental psychology, which uses statistical methods to analyze data gathered through survey or experiment. This approach has the advantage of exhibiting the form of a science. The drawback, however, is that its researchers are not equipped with fruitful ways to generate probable hypotheses. The result is a high non-replication rate. That, combined with the comparatively high cost of experimentation, has left the field in crisis.

Modern cognitive science. The second dominant research program has been modern cognitive psychology. The idea, it is proposed, is that there is an analogy between the mind and a computer. On the strength of this analogy, one can then use ideas from the study of the mind to try to build artificial intelligence and the results of attempting to build AI to learn about the mind. This research program has been very generative, providing a large number of hypotheses for researchers to investigate. Unfortunately, those hypotheses have not yet been seen to pan out, with successful approaches in AI taking inspiration more from the structure of the brain than the structure of the mind.

Ideally, of course, one would want the benefits of these approaches without the drawbacks: a generative research program that focuses on sufficiently stable, sufficiently investigatable elements of the mind, using introspection where possible, but still grounding out claims in predictions about behavior, with sufficiently inexpensive, sufficiently reproducible results. This is what we believe we have achieved. Before describing the Leverage program, it is worth adding three additional research programs, which, while small, are nevertheless noteworthy.

Focusing. "Focusing" is an introspective approach to therapy pioneered by Eugene Gendlin. Gendlin describes a way to engage in introspection, which he calls "Focusing," and provides a codification. Focusing has the advantage of involving introspection (like structuralism) and focusing on beliefs (like psychoanalysis). The primary disadvantage, from a research perspective, is the difficulty translating the data into a form that permits generalizations about the mind.

Internal Family Systems. "Internal Family Systems," or IFS, is a form of psychotherapy that posits that the mind is composed of a number of "parts" or "sub-agents" which exhibit mind-like qualities. The deep investigation of the mind sometimes turns up elements that can easily be interpreted as "mind-like parts." For this reason, IFS is a useful reference point for anyone investigating the mind.

Coherence Therapy. Coherence therapy is another system of psychotherapy; it proposes that psychological symptoms arise from a coherent, underlying perspective on the world, and that those symptoms resolve when they are no longer necessary.

> Leverage's introspection research benefitted from several previous efforts.

Of these, the largest contributors to the Leverage introspection research were mental philosophy, especially Descartes, Spinoza, and Kant; psychoanalysis, especially the evidence from Freud of intelligible and analyzable mental structure; and Focusing, as an example of proceduralized introspection. IFS was a useful reference point during the research; coherence therapy had noted points of similarity but little causal influence. Some of the other research programs were useful as foils, some were only investigated after the fact.

Research Methods

Researchers in the Leverage introspection program used a variety of methods. The central methods are summarized here. This list is not meant to be exhaustive: researchers also employed philosophical methods of analysis and careful description of phenomenology. Rather, these are the methods that established the central feedback mechanism that permitted the acquisition and examination of a large amount of shareable and reproducible data.

Eliciting verbal reports. One of the best ways to gather information about the mind is to ask people questions. Our researchers asked people many questions, with the most common being: "What is an action you take?" "What is another action you take?" "What is good about that?" "What else is good about that?" "If that happens, what happens next?" "Do you have the ability to do that?" "Would it be bad if that were not the case?" and variations.

Building belief/goal models. Through information gained from verbal reports, it is possible to build models of people's behavior. The primary way our researchers did this was by determining which behavior should be counted as "actions" and positing beliefs and goals to explain those actions. The resulting belief/goal models then serve as a basis for explaining a person's behavior, with much behavior explained as the natural result of a person with relevant beliefs pursuing the relevant goals. For instance, a person's behavior of going to professional conferences would be identified as an action, explained by them having various goals that they believed would be achieved by going to those conferences.

> Verbal reports are useful and can be improved with a good reporting intention.

Distinguishing beliefs and endorsements. Verbal reports are not necessarily the best indicators of a person's beliefs. Put differently, belief/goal models built by taking verbal reports at face value do not necessarily have the highest degree of predictive power. Rather, our researchers found that by distinguishing "beliefs" and "endorsements," it was possible to identify verbal reports that provided much better indications of a person's (actual) beliefs. That some verbal reports rather than others were more reliable was tested in a variety of ways, to be described below. Beliefs and endorsements were distinguished, prior to the development of belief reporting, by a number of factors, mostly centrally intuitive fit with the person's emotional responses.

Belief reporting. In particular, our researchers found that instructing people to "maintain the intention to tell the truth" while stating a proposition (and often, stating the proposition's negation), yielded characteristic patterns of physiological and phenomenological response which could be distinguished by both the researcher and the person themselves. These characteristic patterns of physiological and phenomenological response intuitively indicated "yes" (e.g., feeling of resonance, smooth ability to speak) or "no" (e.g., hesitation, shrugging, mental resistance, change of intention), and could then be correlated to the presence or absence of the relevant belief.

Charting. Our researchers found that it was possible to organize information elicited via verbal reports in the form of diagrams that represented the person's putative system of goals. These goal system representations, called "charts," include the notable actions a person takes, the sequences of instrument goals served by those actions, and the basic goals served by the instrumental goals. In the process of building a chart, our researchers would frequently encounter pockets or clusters of beliefs that were surprising or noteworthy, especially given what one would expect a person to believe, given their evidence. These beliefs were noted on charts, typically in ways that indicated how the beliefs affected the way the person sought to achieve their goals.

White chaining. The simplest part of charting was the production of what were called "white chains." A "white chain," so-called because of the background color of boxes used in making the chart, consists of actions at the top and then instrumental goals in the appropriate sequence, leading to basic goals at the bottom. Information for white chains was typically elicited by asking questions like "What is an action you take?" to get an initial action, then repeated instances of "What is good about that?" to elicit the next step down in the chain, with belief reporting used to increase accuracy.

> Information from verbal reports can be organized into representations of goals.

Gray chaining. Another part of charting, the most technically difficult for most researchers, was the production of what were called "gray chains." A "gray chain," named thus because of the background color of the relevant boxes, consists of actions that the person did not believe themselves capable of taking or states of the world they did not believe they had the power to bring about. These indications of perceived powerlessness were important determiners of chains of instrumental goals, i.e., of white chains. Information for gray chains was typically elicited by questions like "Do you have the ability to do that?" or variations; one could then ask the same question about putative components or prerequisites of a given ability. Belief reporting, again, was used to increase accuracy.

Thought experiments. Our researchers found that people sometimes did not answer the intended question, instead, for instance, swapping in a related question or adding an unnecessary assumption. In these cases, thought experiments were often helpful for eliciting relevant information from people. People were often able to report on beliefs from circumstances conceived via thought experiment, though it was important to make sure that the apparatus of the thought experiment did not in some way lead to the wrong beliefs being elicited. Thought experiments were especially helpful in ascertaining basic goals and in constructing gray chains.

Inducing belief change. Charts could be constructed in accordance with belief reports and carefully checked by repeated questions and a variety of thought experiments. The goal standard, however, for the correctness of a chart was its use in causing deterministic belief change. Using a chart, as well as a posited set of dynamics for belief updates, it is possible to make predictions about how a person's beliefs will change in response to different types of information. Information of that type can then be supplied, and it can be seen whether the person's beliefs actually changed.

Checking belief reports. The fastest way to check whether a belief has changed is to check the person's belief reports before and after an intervention. This can also be done with regular verbal reports, though that process can be less reliable, depending on the details of the case. If one uses belief reports, a person may initially belief report that working on a given project is bad, but then, after an intervention is done, belief report that working the relevant project is good; this indicates a change in beliefs. For thoroughness, one can also check surrounding beliefs, keeping in mind that drawing attention to beliefs can, under some circumstances, cause those beliefs to update.

Checking changes in behavior. Belief/goal models imply that some changes in beliefs should lead to changes in action, and hence, behavior. Changes in verbal reports are one subset of these changes, but typically the goal of interventions is to change more than just verbal reports. One can, therefore, look at the actions that are meant to change as the result of changes in belief from a given intervention, and see whether those actions change as predicted. This provides a way of verifying that the effects of interventions designed using charts are real. The reality of the effects will be especially obvious if the changes are large, lasting, and abrupt departures from previous observed behavior.

Mental information can be gathered by imagistic and other modalities.

Testing interventions. By checking changes in behavior, belief report, and other relevant reports (e.g., of emotional responses), it is possible to test a wide variety of interventions. This is made easier by charting, where having an explicit chart enables one to have an estimate of the person's relevant mental states before the intervention and, using theories of belief and action, predict the results of many interventions. One could, e.g., get belief reports about a person's plans for work, and then have them do jumping jacks, and then get belief reports about the same propositions to see, e.g., if doing jumping jacks caused the person's beliefs to change in an expected or unexpected way.

Imagistic exploration. While mental content considered in propositional form, or explicitly expressed in verbal form, is often the easier to work with, it is also possible to examine mental content in other sensory modalities. The most common is visual, where a person can be directed to introspect on visual images and give verbal descriptions of those images. Images can then be translated into beliefs, which can be organized in a chart in the standard manner. The accuracy of material discovered using imagistic methods (e.g. Mythos) can be cross-checked against belief reports or checked directly by making a chart, designing an intervention, and making and testing predictions about interventions.

Attention tracking. In many cases, it is valuable to understand another person's pattern of attention. Some people have an intuitive sense of this, but our researchers found it is possible for many people to learn to track another person's attention by paying attention to them with the correct intention. One's attention then follows the other person's attention; our researchers found that this can be achieved while visually observing the person, with a stationary or semi-stationary physical contact point (e.g., with one's hand on the person's shoulder or back) with one's eyes closed, or remotely. Accuracy can be checked by giving reads and asking the person to report on their pattern of attention.

Intention reads. Our researchers also found that it was possible for some people to articulate information about others' beliefs, either remotely or with a stationary or semi-stationary physical contact point. Such methods were dubbed "intention methods," since the information gained appears to depend on the intention of the reader. Accuracy of intention reads can be checked by directing the person to introspect on relevant mental content, assessing fit with the person's chart, or performing interventions and seeing whether predicted changes in behavior occur.

Removing lens elements. Intention reads were found to be lossier than information gathered via belief reports, with the largest source of inaccuracy for a sensitized reader being concepts in the reader's intention which conflict with identifying particular types of content. When people's reads differed, it was often possible for one or the other to identify a "lens element," i.e., a concept that the other person had fixedly present in their attention. Removing lens elements then increased the degree to which different people gave the same or similar intention reads.

> Different sources of mental information can be cross-checked against each other.

Mental pointing. In some cases, researchers who were using intention methods found it useful to direct each other's attention to particular mental content. This can be achieved by paying attention to the relevant content and then drawing another person's attention to the same content. Consistency or compatibility of intentions was necessary here; sensitized people sometimes could not pay attention to the same things, and this was explained via the incompatibility of their intentions.

Theory generation. In addition to gathering and organizing data, Leverage's researchers also spent considerable time developing theories, specific and general, about mental phenomena and the operation of the mind. Some theories were easily tested, while others (e.g., theories about the relations between skills and beliefs) required substantial effort to test. Theory and experiment formed complementary information sources, including theories about how different instruments and information gathering modalities (e.g., belief reporting, attention tracking) worked.

Central Findings

There are various ways to describe the central findings of Leverage introspection research. The following includes ten major claims that can be thought of as scaffolding the research conclusions, with a variety of details included under each major claim. The order of the claims follows an analogy with other, more familiar fields: physics, chemistry, biology, and ecology. The findings are the result of the application of the research methods described previously; it is expected that others who apply those research methods will reach the same conclusions.

1. The human mind can be adequately described using a language that contains a small number of concepts.

The human mind presents itself in a large variety of ways. There is a similarly large vocabulary one could use to describe it: fleeting feelings, deep-seated concerns, ideas, thoughts, notions, and the like. The team at Leverage found that it was possible to describe the mind using just a very narrow set of basic concepts, with others built up from them.

The narrow set of basic concepts included sensation, concept, belief, goal, action, attention, and intention; a few relational concepts, such as awareness, space (i.e., the spatial relations between visual sensations), and application (i.e., the interpretation of a sensation via the application of a concept); and a few additional more general concepts. Other concepts, such as "perception" and "imagination," could then be defined, for instance as "sensation caused by an external thing it resembles" and "sensation caused by a mental action."

> We found that mental states can be described in a simple vocabulary.

This ontology, which is reminiscent of Descartes' and Kant's, though dissimilar from Hume's, permits one to characterize everyday experience in terms of sensations (e.g., red, blue, hot, cold, sweet, bitter, etc.) interpreted by concepts (e.g., "the apple I am holding," "a frozen blueberry," "a summer breeze," etc.). Many mental phenomena, such as "hopes" or "concerns," can then be characterized in terms of a feeling component, which is one or more sensations, and a cognitive component, which is one or more concepts or beliefs.

This ontology is reminiscent of that of Descartes and Kant, and is dissimilar to that of Hume. It is not meant to be as parsimonious as possible; for that, one could replace "concept" and "belief" with "representation," and state the difference between them, or could do away with the distinction altogether. The concept of "attention" might also be able to be defined, as well as "intention," which was typically characterized in terms of what one believed would happen, especially in the near term and especially as a result of one's actions.

Despite the various imperfections, this set of concepts was sufficient for Leverage's researchers, at least insofar as they followed the central thread of research. The main concepts were: sensation, concept, and belief. Later, as the role of attention and intention were appreciated, these were added and rose in prominence. Whether these new concepts were fundamental was not a matter of central concern; the purpose was to have a parsimonious but practically useful set of concepts to deploy.

2. Most, if not all, of the human mind can be adequately described by a small number of simple rules.

Even if the mind can be adequately described with a simple vocabulary, it may nevertheless seem impossible that the mind itself could function in accordance with simple rules. First, the mind is typically thought to be tightly related to the brain, one of the most complex objects of human study, which arose through an evolutionary process that evidently did not choose maximum simplicity. Second, the mind as it is experienced frequently appears chaotic, with thoughts arising and falling away, images appearing, and actions being taking in patterns that are often difficult to understand.

> We found that the mind follows a small number of simple rules.

Nevertheless, our researchers found that the relevant parts of the mind could be described by the following rules:

1. A person's beliefs update:
    - ...in order to explain their sensations,
    - ...elegantly, i.e., towards mutual coherence,
    - ...locally, i.e., towards local rather than global elegance, and
    - ...only where they pay attention.
2. A person always believes their basic goals will be achieved.
3. A person's basic goals do not change.
4. A person has exactly those concepts included in their beliefs.
5. People act in accordance with what they believe will happen.
6. A person's intention is included in their attention.

Slightly different formulations of these rules are possible. No violations of these rules were identified during the research. Readers familiar with Leverage's research will recognize the first two points as a reformulation of the "belief rule" and "attention rule" from connection theory, and the fourth, fifth, and sixth point as the "concept rule," "action rule," and "intention claim," respectively. The third point is a different formulation than that of connection theory, and more closely matches the rule the researchers actually considered in practice.

One may think of this, provocatively, as a *physics* of the mind. It is, however, not a complete physics. It does not include an explanation for emotion, i.e., when feelings attached to the relevant cognitive states (e.g., in happiness, in anger) occur. It also does not include a statement about the pattern of attention, which may relate to beliefs in some way. It assumed that sensations in general do not follow rules, apart from the fact that some of them are caused by the action of the mind and the rest arise from the external environment and/or the functioning of the brain, as the case may be.

3. The dynamics of the human mind are largely, if not completely, driven by beliefs.

From the preceding rules, it follows that the human mind runs on beliefs and sensations. Sensations and the present state of a person's belief system together yield the next state of the person's belief system, and the process repeats. Emotions decompose into sensations and beliefs, which then feed into beliefs updating as usual. Since people act in accordance with what they believe will happen, actions also arise from beliefs.

It is possible that there are other inputs, though the only available place that fits the preceding rules would be in the pattern of attention. It seems likely that the pattern of attention is itself determined by sensations, concepts, and beliefs, and according to the preceding rules, concepts are determined by beliefs. It thus seems that the dynamics of the human mind are driven entirely by beliefs and sensations; if not, then largely so. Of course, the role of sensations, according to the rules just stated, is to cause changes to beliefs.researchers actually considered in practice.

> We found that beliefs play a central causal role in the mind.

The centrality of beliefs meant, in theory and in practice, that our researchers spent most of their time studying patterns in beliefs. By altering beliefs, one can change other beliefs, or one can change actions, including both mental and physical actions. Though we did not reach agreement on the patterns behind emotion or attention, changes to belief likely cause changes in each. There is thus the challenge of determining how precisely beliefs work. The rules on the preceding page state belief dynamics in a very general sense; this leaves open how beliefs work on a slightly more concrete level.

4. Beliefs in the human mind form a relatively small number of recognizable patterns.

Looking into the dynamics of beliefs, our researchers found what appeared to be a relatively small number of recognizable patterns that repeated in many, many places. If the rules on the preceding page are a *physics* of the mind, the recognizable and repeating patterns of beliefs can be considered a *chemistry*.

The recognizable patterns we found included beliefs holding each other in place (i.e., "roots"), large numbers of nearly duplicate beliefs (i.e., "thickets"), places inside otherwise densely related sets of beliefs where a person lacked a belief (i.e., "gaps"), places where clusters of beliefs on related topics were only sparsely interconnected, yielding compartmentalization (i.e., "zones"), and cases where large amount of behavior seemed to change at once, in apparent violation of the locality of belief updating, because a few lynchpin beliefs were changing (i.e., "modes").

The largest variety of structures at this level were series of mental actions designed to produce transformations of mental content (i.e., "intellectual processes" or "IPs"). Many of these were used for intellectual purposes, but some were used for the purpose of maintaining motivation or simply directly altering beliefs (i.e., "belief injection"). We catalogued hundreds of IPs, including ones that interfered with introspection.

> Our researchers found a large number of basic, recognizable belief patterns.

Recognizing that mental actions could interfere with introspection was one of the first big breakthroughs necessary before being able to quickly and reliable examine belief clusters. This in turn made it much easier to run predictive tests on interventions designed to cause belief change. Other structures also interfered with introspection, including beliefs that knowing one's beliefs would be bad (i.e., conflict with the achievement of one's basic goals), beliefs that knowing why knowing one's beliefs would be bad would be bad, and so forth.

One of the most important mental structures we encountered was the attentional redirect. In some cases, attentional redirects were implemented by IPs, though in other cases they appeared to be a result of the belief structure itself. Attentional redirects were also barriers to introspection, and together, certain aggregates of attentional redirects created a *de facto* distinction between the conscious and subconscious mind. This was then relevant to the question of what parts of the mind could be explicitly mapped out using which techniques.

5. Some aspects of human belief systems can be mapped out explicitly; other parts are intractably complex.

Most of the work our researchers did testing the basic rules and working out the basic mental structures was done in the context of trying to map out as much of the human mind as possible. In the process of doing this, we frequently encountered overwhelming complexity. The key question was which parts of the mind could be mapped out, and especially which could be mapped out easily and quickly. It was found that some parts of the mind are tractable, while others are intractably complex.

From the beginning, Leverage's researchers found that people's consciously accessible goal systems were mappable. Being "consciously accessible" is not a natural kind; rather, it is easier or harder to direct a person's attention to particular topics and have them report beliefs on those topics. There is a point, however, after which directing a person's attention using normal conversational methods becomes impractical; things prior to this point can be designated "consciously accessible."

Goal systems themselves include basic goals and instrumental goals, with instrumental goals forming chains or sequences that terminate in the basic goals. Basic goals and instrumental goals are both embedded in beliefs, especially beliefs about what is good and what will happen. A full representation of a goal system includes both chains of instrumentality, ending in basic goals, and the actions, mental and physical, that a person regularly takes.

In the process of mapping out goal systems, it is natural to encounter beliefs that are unexpected or unusual, given one's understanding of the person's evidence. Mapping the entire belief system, even just appreciable parts of the consciously accessible aspects of the belief system, was not found to be feasible. Small belief clusters, were entirely tractable. Adding annotations about unusual or unexpected beliefs, in the form of notes about belief clusters, to the representation of a goal system resulted in maps containing anywhere from 150 to 950 nodes. This meant that "charting," which was the process of creating maps of the consciously accessible elements of people's goal systems, was a difficult and intensive, though practical completable, activity.

> We found that it was feasible to create maps of the conscious goal system.

Charting originally took approximately 40 hours per chart. Through serious effort, charting technique improved and it became possible to produce reasonably complete charts with the important elements displayed in a few hours. Being "complete," in this context, means that the chart includes, from what is consciously accessible, all of the actions a person regularly takes, all of the chains of instrumental goals that lead from them, all of the relevant basic goals, and annotations noting and explaining all of deviations in belief from what one would expect given the person's evidence.

As our researchers found, the limitation to the conscious elements of the belief system is important. The subconscious elements are decidedly more vast and complicated than the conscious elements, and mapping those was not considered tractable in a normal sense. A more ambitious project to map the entire belief system, including conscious and subconscious elements, can certainly be envisioned, though this would be a grand undertaking that, if it is possible to complete, would certainly be orders of magnitude more difficult than creating a standard chart.

6. Many of the specific details of human belief systems are simple and common enough to be fruitfully studied.

The study of individual human minds is, to continue the analogy with the physical sciences, much like the *biology* of the mind. Individual minds have belief systems which are composed of beliefs that form commonly repeating structures. This, as noted, corresponds to the *chemistry*. Those elements, then, do not form mere compounds, but instead compose vastly larger structures, some of which are sufficiently simple to be studied directly.

As in biology, there are a vast number of interesting and tractable questions that can be investigated about the individual human mind. For a given person, one may want to know what their basic goals are or what the intermediate steps are that they believe will transpire on the way to the achievement of those basic goals (i.e., the chains of instrumental goals, or "paths"). One may want to know particular beliefs, or particular sets of beliefs, or the causes of particular beliefs, especially in cases where beliefs do not update normally (i.e., elegantly) in response to evidence.

There are also many interesting questions that can be examined about topics of belief, where it will be necessary to study the belief systems of many people in order to find representative or sufficiently exhaustive answers. For instance, one may want to know about the origin of religious belief or the degree to which adolescent rebellion is socially conditioned. One may want to know why people believe that 2 + 2 = 4 or the degree to which people believe the deliverances of science on the basis of reason, demonstrated power, or testimony from trusted authorities.

> Our researchers found evidence pertaining to many important questions about belief.

In some cases, answers can be deduced from the rules of the mind. For instance, it is relatively easy to enumerate the causes of irrational belief (i.e., belief that deviates from global elegant updating on evidence). An empirical study is necessary, however, to determine how frequently people's beliefs on a topic deviate from the evidence due to constraint, which arises from the need to believe basic goals will be achieved, entrenchment, which arises from the locality of updating, or inattention.

The complexity of the belief system is a barrier to answering some questions, but not all. Furthermore, while human beliefs do change, both individually and on a group level, our researchers found a surprising degree of stability. Individual people's beliefs, especially in the context of stable life circumstances, were found to remain the same on time-scales of at least months. It is presumed that belief formations frequently last much longer than that. This means that, in practice, questions about individual, group, and society-wide belief can be investigated and answered in a way that retains utility despite ongoing changes in belief.

7. Human perception includes high-bandwidth non-verbal communication.

The existence of attentional redirects naturally raises the question of whether there are ways to bypass those redirects, or whether there are parts of the human mind that are entirely shut off from observation. Leverage's researchers struggled with this question on a practical level, making a small amount of progress through both diligence and the attempt to remove higher-level blocks. The development of intention methods then made it possible to bypass attentional redirects in a wholesale way, with a corresponding, though not critical, increase in the noisiness of data collection.

In particular, our researchers found that it was possible to read information about others' patterns of attention, and then beliefs, through touch, and then without touch. Different researchers postulated different theories of how this was possible; the simplest was to posit that humans are in high-bandwidth non-verbal communication with one another, that the results of this communication are frequently not attended to, and that with care it is possible to learn to pay attention to it. This is a commonsensical set of viewpoints, which is neutral on deeper metaphysical questions and has surprising consequences once combined with the belief dynamics described earlier.

> We found that people's beliefs are affected in lasting ways by non-verbal information.

Non-verbal communication is widely believed to occur; in poetic or literary contexts, one may speak of two people "having a silent conversation." The postulate of high-bandwidth non-verbal communication amounts to the proposal that such communication is happening constantly and produces lasting changes to people's beliefs. The primary question, then, will be as to the expressiveness of the non-verbal language; our researchers found that it was highly expressive, at least on a par with verbal communication.

Our researchers found, surprisingly or not, that the same belief dynamics, including the physics and the chemistry, were exhibited by beliefs, whether the communication was verbal or non-verbal in nature. Non-verbal communication, however, permitted bypassing attentional redirects and made it much easier for one person to access consciously what was for another person subconscious. This made non-verbal reads, gathered via what our researchers called "intention methods," valuable for continuing the examination of the mind beyond what is easily consciously accessible.

There were three main barriers to the use of intention methods. First, the methods must be learned, in a process that may be dubbed "sensitization." Second, the process of sensitization can be deeply disturbing to the person learning to do intention reads; there are a host of practical challenges here. Third, and most relevant to the findings, is that information gathered by intention methods is lossier than information gathered via conscious, correctly intentioned verbal report (i.e., by belief report). It was found that, in practice, sometimes with difficulty, these barriers could be overcome.

Fortunately, the lossiness of attention reads was found to have a pattern. Concepts can sometimes become stuck in the attention of the reader (i.e., the "lens"), which makes it difficult to identify mental content that discoheres with those concepts. The intention of the reader was found to behave similarly, so that readers with different intentions would get different reads. In practice, it was possible to reduce the magnitude of these effects, causing readers' reports of people's beliefs to become more similar. Also, in many cases it was simply possible to direct the person's attention carefully to the relevant beliefs, bypassing relevant attentional redirects, and getting direct confirmation from the person themselves.

8. High-bandwidth non-verbal communication yields an information ecosystem with arguably magical properties.

The existence of high-bandwidth non-verbal communication, on a basic level, can simply be thought of as people having a more complex set of sensations, or of being more complexly responsive to the same set of sensations, than would be expected otherwise. It thus poses little problem on a theoretical level. In terms of consequences for the study of the mind, however, it indicates a number of important challenges.

First, it follows that in many contexts, investigators are thoroughly entangled with the people they are studying. Investigators may then have effects on those people; in some cases, for instance, it was found that investigators could change a person's belief reports simply by changing their own (that is, the investigator's) intention. This problem can occur when dealing with conscious content, though the effects there are fairly easy to identify (e.g., mode switching or belief injection). In practice, conscious content is much more stable, making investigation easier, and, importantly, the physics and chemistry of the mind remain the same even if the investigator is inadvertently affecting a person's beliefs.

Second, it reveals, on an empirical level, a vastly more complicated landscape of beliefs than is visible to regular introspection. The conscious mind is large and intractably complex in some regards; however, the goal system and small belief clusters are simple enough to map. When subconscious content (i.e., content importantly hidden by attentional redirects) is taken into account, it is clear the mind is shockingly vast. This may be discouraging to researchers who have hopes of mapping the entire terrain, though of course, a fairly complete, useful map of the whole is not fully inconceivable.

> Non-verbal communication both poses challenges and provides opportunities.

While the existence of high-bandwidth non-verbal communication poses challenges to research, it is also another mental phenomenon that can be fruitfully studied. Leverage introspection research ended before many of these phenomena could be characterized. Nevertheless, it is possible to give a brief characterization of general dynamics, along with examples of phenomena that can be explained.

High-bandwidth non-verbal communication, combined with many people having basic goals pertaining to connection, which includes mutual understanding or shared belief, results in humanity itself becoming a massive information processing unit, with the whole seeking information equilibration or the reduction of information asymmetries. Because much of that processing is subconscious, this provides a way to explain a number of phenomena that might otherwise be difficult to describe.

For instance, Carl Jung noted the occurrence of synchronicity, which involves seemingly coincidental events happening a rate greater than chance. Such a phenomenon can easily be accounted for if one posits that people are engaged in highly expressive, detailed communication on a subconscious level, since people may, in effect, already all know each other's schedules. Taking into account the importance of intention, which drives belief updating by being a constant part of the attention, and which influences others via multiple means (e.g., the drive for connection), it is then possible to explain purportedly "magical" phenomena like the Evil Eye or the power of positive thinking.

The effect of high-bandwidth, non-verbal communication is to bind together human beings or, more accurately, to recognize the way in which we are already bound together. It thus yields a transition from the *biology* of the mind to the *ecology* of the mind, with a corresponding increase in complexity and interest.

9. The results reached here are sufficient to enable a bridge between the study of the mind and various other fields.

There are obvious potential overlaps between the study of the mind, whether that occurs under the auspices of psychology or cognitive science, and those of many other fields. One can easily imagine connections to sociology, anthropology, and organizational management on side and neuroscience and physiology on the other. The primary difficulty in fruitful cross-pollination has been the lack of solidity on one or another side: it is difficult, for instance, to correlate brain states and mental states if one does not know enough about how to describe or classify the mental states.

> The study of the mind has links to other fields, which our researchers explored.

Our researchers found that the models they developed and conclusions they reached, from the "physics" through the "chemistry" and "biology" to the "ecology" of the mind, useful informed investigations of several other fields. Those most explored were extensions to the study of group dynamics and society, though some researchers also investigated connections to neuroscience and the study of human physiology. Indeed, in some cases it was found that the study of the mind directly, via the methods our researchers used, had important limitations, and that our researchers' findings helped build a bridge to other fields, without replacing the need for essential input from those fields.

10. Many questions about the mind remain open.

There also remain a very large number of questions to answer pertaining to the mind, and which might be thought of as falling within the fields of psychology, cognitive science, or education. At the most basic level, there is a question of whether the patterns identified or confirmed by Leverage's researchers will be identified by other researchers. This is something about which one should expect confirmation or disconfirmation in the natural course of further study.

Assuming the results will hold up, largely or completely, there are then many remaining questions to answer. With respect to the "physics" of the mind, there is the question of the pattern of emotions and the pattern of attention. Researchers at Leverage made various conjectures or came to different theories, but these were not agreed upon. Regarding the "chemistry" of the mind, it is likely that there are structures that show up more frequently on a subconscious level that our researchers did not encounter or only encountered briefly. These and other structures can be catalogued and described.

> If our results are upheld, many questions at many different levels remain open.

The "biology" of the mind has a very large number of unanswered questions. The actual substance of most people's interest in the mind — the character of the beliefs of various groups of people, the most effective therapeutic methods, the most effective educational methods, and more — need to be investigated. Here, special attention will need to be paid to the likely patterns of similarity and dissimilarity across minds: the more variation observed in a sample, the more variation one should expect outside of that sample.

With respect to the "ecology" of the mind, there is a question of how much can actually be determined, and whether there are limitations on knowledge imposed by the object of study itself. Limits on study are common across fields, in some cases arising from limitations of the instruments, in other cases being a consequence of the relevant theories. The ecology of the mind borders on sociology and political science; there is then the important question of how much self-knowledge individual people, and humanity itself, can achieve.

Prospective Applications

It goes without saying that an improved understanding of the human mind has many likely applications. Some of these will be sketched here, especially those indicated by Leverage's introspection research. In each case, it will be noted roughly how much effort our researchers put into examining each application and what results they found.

Education. The primary focus of Leverage's introspection research was education. Here, the primary questions were whether skills can be decomposed in belief correlates, whether relevant belief correlates can be gained, and whether, as a result, it is possible for people to gain new skills or learn new subjects more quickly than they would otherwise. Our researchers found that, in general, beliefs are sufficiently mutable that, in principle, anyone can learn anything, but that there are very strong predispositions to learning some things rather than others. They also found that people could dramatically increase the pace at which they learn and the range of things learned and, in general, become substantially smarter.

Mental health. The second main focus of Leverage's research was mental health. Leverage's researchers were not licensed therapists and did not advertise as such. Rather, in the course of examining beliefs it was inevitable that one would find a very large number of belief and goal clusters that were causing problems for people in their everyday lives and on which it was possible to intervene. The result was the development of the ability to help people "solve issues." This was very successful in a general sense, though specific issues were seen to vary in underlying structure widely. The result was that no "silver bullets" for well-known mental problems were discovered, and that helping a person deal with one or another problem could take from a few minutes to many years.

> Our researchers explored many different applications, for instance, in education.

Culture. The next main focus of the introspection research was culture. Just as one encounters a large number of individual personal issues in the course of examining the mind, one also finds a large number of interpersonal and group dynamics issues. Some of these end up affecting what one might think of as the "culture" of groups or organizations, and it is possible to design psychologically-informed interventions that are meant to help improve culture or reduce problems that arise in culture. This research was also successful, though it was discovered that developing a good culture is a very hard problem, especially when intention phenomena are taken into account.

Organizational design. A related topic of study was organizational design. Researchers at Leverage did experiment with organizational design, with this experimentation informed by the introspection research. However, a clear predisposition towards individualism existed in the group, which led to a deprioritization of the study of group phenomenon until near the end of the research program. This yielded a bias against believing in the reality of organizational forms, which impeded the study of organizational design.

Communication. A further topic of interest was communication. Rather than looking for new marketing tricks, which have been thoroughly explored by existing individuals and organizations, most of Leverage's focus on communication was on understanding the patterns of adoption for ideas, what would allow or prevent them from spreading. This, for the most part, turned into the study of clear communication, which involves understanding the subject matter, the audience, and the way the human mind interacts with information.

Social analysis. Concurrent with its introspection research, Leverage also supported sociological research. The findings and themes from the introspection research were an important influence on the sociological research, and at least one important effort was made to bridge from an understanding of the mind to the identification of macro-scale patterns in society.

Physical health. Another potential application for knowledge of the mind is in physical health. Our researchers identified a link between the mind and muscular control and developed procedures to help people relax unintentionally tense muscles. Some researchers looked into other relations between the mind and the various systems in the body (e.g., the immune system), and reported that they had found connections.

> Physical health and social analysis are two particularly promising applications.

The foregoing is a sketch; the results of investigating each of the applications, as well as the progress made by Leverage's researchers after the end of the introspection research program, should be thoroughly documented. Each of the applications is likely to be highly valuable, though people's judgments of which are the most valuable will depend on their different beliefs and goals.

Discussion of Evidence

Since the claims in this document involve a new set of research methods and set of new results, occurring in a field that has a history of misfires, it should not be supposed that sufficiently solid evidence could be provided on the basis of previous data. Instead, the claims here are offered for consideration in anticipation of confirmation or disconfirmation by evidence gathered by future researchers.

In this connection, it is important that it is relatively easy to replicate the basic techniques and reproduce the basic findings described here. Belief reporting is frequently easy to learn; if it is hard for a person to learn, another person for whom it is easier can be selected. Charting is also not difficult: if one encounters a difficult part of the mind, one can choose a simpler part and observe the dynamics there. Even inducing belief change, via a method that uses white chain, gray chaining, flipping a belief at the top of a gray chain, and propagating the belief change in a way that alters the white chain, is not unduly difficult.

> Ongoing evidence for the claims will come from data gathered by new researchers.

What one then finds is confirmation that a particular set of patterns exist or can be observed in the mind, and that those patterns hold for a limited number of human minds, and moreover, a limit part of those minds. One will then push further, investigating further people's minds, perhaps including one's own, and find the same patterns hold. Counterexamples or anomalies to general claims will be discovered but, on brief examination, then be discovered to not actually be counterexamples.

As one increases the number of people one is working with and the number of different topics or phenomena investigated, one will then find some phenomena that fit easily into the preceding patterns and some that do not. Over time, as one makes clearer distinctions and becomes able to identify more subtle patterns (e.g., the operation of a mental action that redirects attention), one will find the original patterns upheld more and more completely. There will then arise the question, as it did for us, whether the patterns are actually universal, holding not just for a small group, but for all people, and not just for some parts of the mind, but across all of them.

Different researchers will have different predilections for theory or experiment. Some will infer quickly to a single shared basic goal; others will hold out for more empirical evidence. Some will accept the ontology of beliefs, concepts, sensations, actions, and so forth, while others will develop variations. Some will expect patterns to hold with true universality, while others will expect that there are exceptions that occur with increasing frequency as one approaches the edge of "normal experience."

There will be opportunity for crucial experiments. Some people will find, for instance, the question of whether "psychopaths" have the same mental dynamics as other people to be important. It will be very informative to see what happens when people with varying degrees and types of brain damage are charted. Many of the "tests" of the relevant claims will come from practical use, where educational methods based on knowledge of the mind will have the opportunity to compete against those that do not heed such knowledge.

Some of the original data from Leverage's introspection research is available, having been gathered and suitably anonymized. It will be cheaper and more evidentially compelling, for new sources of data to be developed. But old data can be presented, if it is of interest.

There will be opportunity for crucial experiments. Some people will find, for instance, the question of whether "psychopaths" have the same mental dynamics as other people to be important. It will be very informative to see what happens when people with varying degrees and types of brain damage are charted. Many of the "tests" of the relevant claims will come from practical use, where educational methods based on knowledge of the mind will have the opportunity to compete against those that do not heed such knowledge.

Some of the original data from Leverage's introspection research is available, having been gathered and suitably anonymized. It will be cheaper and more evidentially compelling, for new sources of data to be developed. But old data can be presented, if it is of interest.

> We encourage researchers with an interest to experiment with the basics methods.

It is valuable to note that while learning the basics of the experimental methods described earlier in this document is relatively easy, achieving mastery in the investigation of the mind is difficult, and it should be expected that getting to the true cutting edge, and pushing further, will take many years of study. This should not discourage everyone; for some people, the mind is an enduring object of interest, and the potential benefits that come from better educational or therapeutic methods, or from other applications, should absolutely be pursued by some.

For those with a serious interest, it should be noted that investigating the mind comes with risk. These risks are covered in some of the further reading at the end of this document. In general, experimentation with the mind should be thought of as experimentation with the body. The risks and benefits should be carefully weighed by prospective researchers.

Acknowledgements

This report was prepared by Leverage staff, with Geoff Anders as author, in advance of the 2024 Bottlenecks in Science and Technology workshop. Melinda Bradley and Oliver Carefull provided commentary. Geoff created the cover image. Opinions in the document are those of the institute.

Our great and enduring thanks to all of the researchers who continued to Leverage's introspection research in 2012-2019. A more detailed record of who contributed to which discoveries will be covered in a lengthier treatment. We also thank the operational staff and all team leaders who managed Leverage during the time of this research, as well as the funders who made it possible.

Further Reading

Burns, Brian and James Dama. *Chart Logic and Core Mechanics*, Paradigm Academy, October 2022 (originally published August 2018). https://bit.ly/chart-logic-and-core-mechanics

Leverage Research. "What an Actual Science of the Mind Would Look Like," March 2023. https://bit.ly/what-an-actual-science-of-the-mind-would-look-like

Vaughan, Kerry. "How to Belief Report," Leverage Research, November 2022. https://bit.ly/how-to-belief-report

Vaughan, Kerry. "Introspection Safety for Researchers," Leverage Research, November 2022. https://bit.ly/introspection-safety-for-researchers

Vaughan, Kerry. "On Intention Research," Leverage Research, April 2022. https://bit.ly/on-intention-research