Prediction, Control, and the Regulating Mind

The Short Version: A Working Framework

Dr. Greg Moody

August 8, 2026

1. What this adds to work you're already doing

Before I was a licensed psychotherapist and ran businesses and did martial arts and blah blah blah… I was an aerospace engineer. Part of this is about guidance, navigation, control. Feedback loops with an estimate on one end and a correction on the other, and a short list of ways they fail.

One day, I realized these two areas fit as I came across 2 concepts…

First, I was reading the great and fun (and not academically backed up) work by the weirdly named "Slime Mold Time Mold" guys… "The Mind in the Wheel", which is a theory of mind laid out in a cybernetic format.

A word on **cybernetic**, or **cybernetics**, since it has been dragged off in the wrong direction. It has nothing to do with cyborgs, and it predates them. It comes from the Greek *kybernetes*, the steersman on a ship, and Norbert Wiener took it in 1947 for the study of how any system steers itself by feedback. A man at a tiller, a thermostat, a cruise control, a body holding its temperature. The system senses where it is, compares that to where it means to be, and corrects. Wiener chose the word because the first serious paper on the subject was Maxwell's 1868 study of *governors*, the spinning weights that hold a steam engine at speed.

The argument, across a prologue and twelve parts, is that psychology never had a paradigm because it never proposed actual entities and rules, only abstractions. Their candidate entity is the negative feedback loop, which they call a **governor**: a sensor, a set point, a comparator, and an output that acts on the world. Their headline claim is that **an emotion is the error signal in a behavioral control system**. Hunger, thirst, fear, shame, all error signals from different governors, all competing for one body through an arbitration layer they call the selector. And **error** here doesn't mean bad. It's a signal that something is off, that's all, and it comes out as a feeling. From there they rebuild personality as the parameter settings on your governors, and recast depression not as one disease but as several mechanically distinct faults that happen to share a surface. It's self-published, it has essentially no new data, and it's the most complete recent statement of a control-theoretic psychology anyone has written. But it resonated with me.

Then, completely separately, one day I was interested in what causes "stimming" in autistic people. I came across the idea of predictive error models in psychology.

The claim there is that the brain doesn't receive the world, it guesses at it. Predictions, with estimations of the expected environment, run down the hierarchy, and only the mismatch (the prediction error) runs back up. What you end up perceiving is a settlement between what you expected and what arrived, weighted by how much the system trusts each one. That weighting is called precision, and it's the parameter that does all the work.

> *Aside… if you have read Plato's The Allegory of the Cave, this is kind of congruent with that… but enough intellectualizing here….*

Figure 1.1. The allegory of the cave. The prisoners see shadows, not the objects casting them, and take the shadows for the world.

I realized MAYBE the cybernetic idea had merit and it was **modified by** the predictive error model. Both are systems, and both predict emotion and behavior, but while the control system model (cybernetic) ran off of a set point, the predictive model may **establish** the set point.

I see this as a complement (not a replacement) to cognitive behavioral therapy. CBT has decades of outcome data behind it. A control-theoretic account of its mechanism is just an idea that may fit.

What it offers is a layer underneath. CBT tells you to examine the thoughts and schema, and CBT is my favorite model. But "test the thought" covers at least three mechanically different situations, and nothing in the standard model separates them for you.

The belief can be inaccurate.

The belief can be accurate and acted on with more force than the situation calls for.

The belief can be roughly right and held with more certainty than it deserves.

Three different problems. Three different levers. CBT already contains all three, which is most of why it works: cognitive restructuring, arousal and distress-tolerance work, behavioral experiments. What this new model may help with is understanding about the feedback and resetting the mind may go through.

Let's look at both models in a little more detail and then put it together… see what we get!

2. The mind as a control system

Cybernetics says the mind is a regulator. It has targets, it senses how far it is from them, and it acts to close the gap.[1]

The vocabulary is small and worth having.

Set point. The target value the system defends. **Sensor.** What reports the current state. **Error signal.** The gap between the two. **Controller gain.** How hard the system acts on that gap. **Damping.** Whether corrections settle or overshoot. **Delay.** How stale the information is by the time it drives a correction.

The clinically interesting move is that **an emotion is the error signal in a behavioral control system**. Hunger is what being below a set point feels like from the inside. Fear is what it feels like to have too much estimated danger against your tolerance for it.

Then what is happiness?

It's the first question anybody asks, and it's a good one. Hunger, fear and shame all behave like error signals. Joy doesn't. Nothing in you is working to drive joy to zero.

The Slime Mold answer is that happiness isn't an emotion at all.

> "Emotions are easy to identify because they are errors in a control system. Like any error in a control system, successful behavior drives the error to zero. This means that happiness is not an emotion."

What happiness is, on their account, is what it feels like when a governor's error gets corrected. Their line: "Happiness is what happens when a thirsty person drinks, when a tired person rests, when a frightened person reaches safety." Correct a large error, or correct one quickly, and you get more of it than from a slow incremental fix. That matches ordinary experience better than it has any right to. The best meal of your life was eaten hungry.

---

[1] If the word is ringing a bell from a paperback, that's *Psycho-Cybernetics* (Maltz, 1960), which sold north of thirty million copies borrowing the same engineering vocabulary for a self-help program. It's a real ancestor of this document and not a hostile one, but it's missing the one piece this whole argument turns on. Appendix A works through the difference.

It also predicts something odd that I think is right: there's no such thing as unhappiness. Happiness can't go negative. What gets called unhappiness is an uncorrected error somewhere, which is a different thing wearing the same word.

And they give it a job. Happiness marks which behaviors worked so the system repeats them, and helps calibrate how much to explore versus stay with what already works.

Now the hole, because there is one.

That account handles joy, relief, satisfaction and pleasure by making them all the same thing. Their words: "there's only one way to feel good," and "all of our words for positive emotion, joy, excitement, pride, are really referring to the same thing, just in different contexts."

That's asserted rather than argued, and it leaves amusement out entirely. What error gets corrected when something is funny? What set point does a joke defend? Nothing obvious, and the series never raises it. Amusement, awe and aesthetic pleasure don't appear anywhere in fourteen posts.

They're straighter about curiosity, which they flag themselves as an enigma: "acting on your fear should make you less afraid, acting on your thirst should make you less thirsty, but acting on your curiosity often seems to make you more curious." A signal that grows when you act on it runs backwards from every other loop in the model.

So my position is narrower than theirs. Error signals account for the emotions that push you toward a target. Happiness is a reasonable read of what arriving feels like. Amusement, awe and curiosity are unaccounted for, and I'd rather leave them that way than stretch the model to cover them, which is the exact flexibility problem Section 9 is about.

Hold that one loosely for now. It gets a good deal more precise in Section 4, once the estimator is bolted on, and the precise version is the one that matters clinically.

If the list of regulated variables reminds you of Maslow, the resemblance is real but the logic runs the other way. Maslow arranged needs into a hierarchy of priority. This arranges them as parallel control loops competing for one body, which is why a person can be hungry and lonely and cold at the same time without any of it queuing politely.

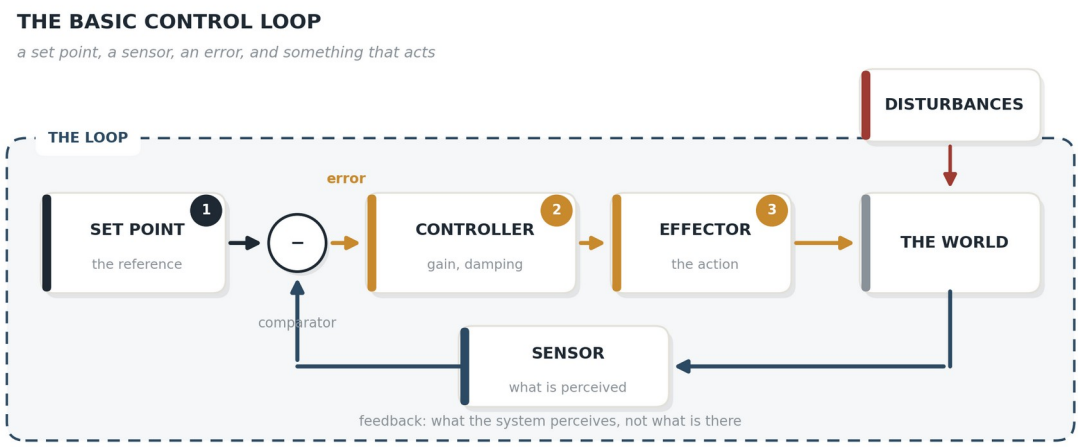

Figure 2.1. The basic control loop.

What this buys you is a failure taxonomy. A loop can break in a small number of specific ways: the sensor misreports, the set point is wrong, the gain is mis-set, the damping is insufficient, the effector fails, or in a system with competing loops the arbitration fails. Those are different problems with different interventions, and symptom-level diagnosis doesn't distinguish them.

The one that matters most here is gain and damping. A loop with high gain and insufficient damping doesn't just respond too strongly. It **oscillates**. It overshoots, corrects, overshoots the other way, and either settles slowly or never settles at all. Wiener identified cerebellar tremor as exactly this problem in 1948. Somebody with cerebellar damage reaches for a glass and the hand doesn't travel smoothly to it. It overshoots, comes back too far, overshoots again, and hunts around the target, getting worse the closer it gets. Wiener's point was that this isn't weakness or paralysis. The muscles work. What's broken is the correction: it arrives too hard and too late, which is the textbook signature of a feedback loop with its damping removed. He was looking at a neurological sign and seeing a control failure, and that is the moment the whole field starts.

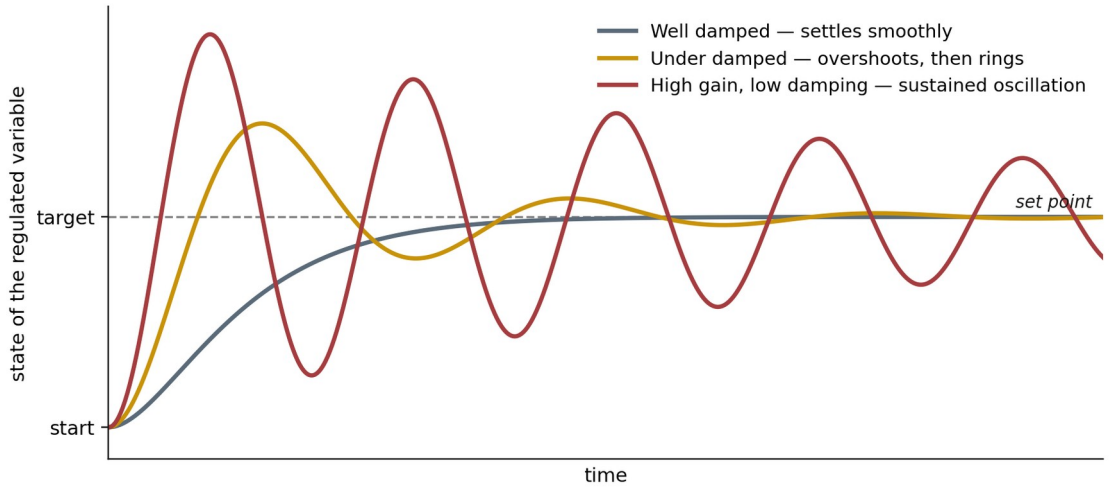

Figure 2.2. Gain and damping. Same system, three parameter settings.

Where it falls short. A thermostat reads a number and takes it at face value. No biological system gets that deal. Its sensors are noisy, slow, and measure proxies rather than the regulated variable. Plasma osmolality is not thirst. A racing heart is not danger. The control account has a sensor-shaped hole in it.

2a. The human thermostat

Before going further, work the thermostat all the way through. It's the cleanest way to see why a controller alone isn't enough, and it's the example I use when I'm explaining this out loud.

A thermostat on your wall already has predictions built into it. Not obvious ones. It waits a little while before it adjusts. This is called **hysteresis**, which means the point where it switches on isn't the same as the point where it switches off. Once it turns the air conditioning on it holds until the temperature passes the set point, rather than flipping the instant the number crosses. It waits a few degrees on the way back. None of that is comfort engineering. It's there so the unit isn't cycling on and off and on and off on sensor noise.

Now take the thermostat off the wall and put a person in its place.

Hand somebody the dial. Their only sensor is their own body. They feel warm, so they crank the AC up. Then they feel cold, so they crank it down. They do it the instant they feel it.

Here's what they don't know at the beginning. There's a delay between turning the dial and their body registering the change (it takes a bit to turn the AC on, then for the air to circulate and get the temperature reduced and then for their body to sense it).

So they crank it up when they're hot, and nothing happens, so they crank it further. Then the room goes cold, and they crank it way down. Then it's hot again. They keep cranking,

harder each time, because they aren't getting the response they expect on the timeline they expect it.

That person doesn't have a broken thermostat. They have a **broken model of the system.** They haven't learned that the environment has a lag in it.

Now let them learn. Once they know about the delay, the behavior changes completely: crank it up a little, wait, then decide whether to adjust again. Same dial, same body, same room. What changed is the prediction, and the prediction is what damped the controller.

That splits the job in two, and the split is the whole point of this document.

The estimator's job is to predict what's going to happen in the environment given the sensory input coming in. **The controller's job** is to decide how much to adjust.

Two different failures live in those two jobs, and they look similar from outside.

Go back to the dial. Suppose the sensor itself is bad. The thermostat reads 70, then 71, then 68, then 71, then 68, and it reacts to every reading. Now the AC cycles constantly, and nothing is wrong with the decision rule at all. The reading is the problem.

Or suppose the sensor is perfect and the built-in patience is missing. It's exactly 70.0. At 70.1 it turns on, at 69.9 it turns off. Perfect information, no prediction, and the same cycling.

A worked contrast, because it isn't all failure. Put your hand on a knife and cut yourself. The pain arrives large and the response is immediate, which is the controller doing its job correctly, because that's a case where fast and hard is the right answer. Meanwhile the other half of the system is doing something slower and more useful. It's taking that sensory input against what you expected and learning from it: *that hurt because the knife is sharp.* Now you carry a prior that the knife is sharp, and your caution around knives is set at a more accurate level than it was an hour ago. One event, two jobs. The controller handled the moment. The estimator changed the setting.

Which is also why the same parameter is right in one loop and wrong in another. For pain and danger, act fast. For whether the room is a little warm, you can afford to be slow. React to discomfort on the timescale you'd react to danger and you're the person cranking the dial.

Two different things can go wrong here, and Section 4 gives them names. For now: one is a problem with the estimate, and one is a problem with the response. Hold both.

3. The mind as a prediction machine

Predictive processing says the brain doesn't receive the world so much as guess at it, continuously, correcting the guess only where the evidence insists.

Predictions flow down the hierarchy. Only the mismatch flows up. What reaches you is a reconstruction, not a readout.

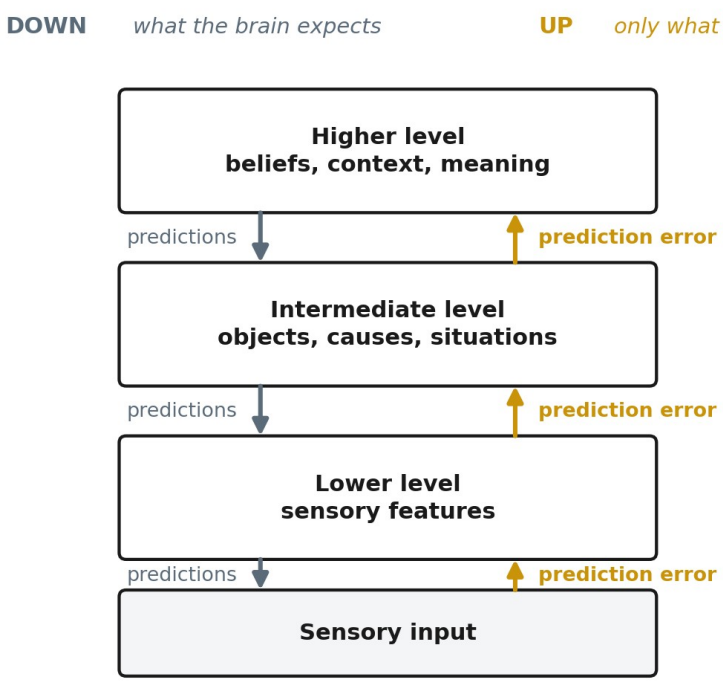

Figure 3.1. Predictions descend. Only prediction error ascends.

The parameter that does the work is **precision**, meaning how much the system trusts a signal relative to its own expectation. Same evidence, same prior, different precision settings, and you get a different percept.

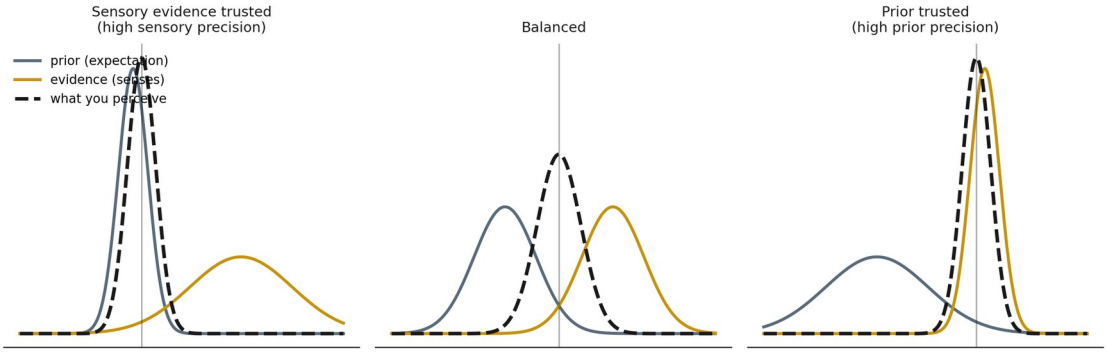

Figure 3.2. Precision weighting. Identical inputs, three different settings.

This is also the perceptual version of something I've argued about decision-making for years.

> **Expectation runs in front of evidence. We decide and then manufacture the logic.** A Vulcan couldn't run a business because rationalization almost always arrives before reasoning does, and the antidote is engineering discipline: estimate, test, and let the results outrank the argument you fell in love with.

Writers have a version of this, and it arrives with a lesson attached. It's usually quoted as Faulkner: *in writing, you must kill all your darlings.* Faulkner never said it. It's Arthur

Quiller-Couch, lecturing at Cambridge on January 28, 1914: “Whenever you feel an impulse to perpetrate a piece of exceptionally fine writing, obey it, wholeheartedly, and delete it before sending your manuscript to press. *Murder your darlings*” (Quiller-Couch, 1916). The Faulkner version got popular through screenwriting guides in the 1990s and has been loose ever since. Samuel Johnson gave the same advice 150 years earlier and credited it to an unnamed old tutor of a college. So the advice about killing your darlings arrives with a darling of its own that nobody has managed to kill.

What predictive processing adds is that this isn’t only true of reasoning. It’s true one level down, in perception itself. The expectation arrives before the evidence does, and what you end up seeing is the settlement between them.

Where it falls short. It explains beautifully how an estimate gets built and says very little about why the organism cares, or how hard it acts once it knows. And its evidence base is thinner than its citation counts suggest. Section 8 gives the numbers.

4. Putting them together

A more thorough model is when we combine them.

Predictive processing describes how the estimate gets made. Control theory describes what happens to the error once the estimate exists. Bolt them together and you get a control loop with an inference engine on the front.

Two things about how it’s wired, and both matter clinically.

The estimate establishes the target. The estimator combines what you expected with what you sensed and produces an estimate of the current state. That estimate is also your updated prediction of what should be happening. And that prediction is what the controller compares against. There’s no separate goal sitting off to the side. The thing you predict IS the set point you defend.

The sensory signal goes to both halves. Your action changes the world, the world sends back sensation, and that same sensation does double duty. It feeds the estimator as evidence to weigh against the prior. It also feeds the controller, where it gets compared against the set point. One signal, two jobs.

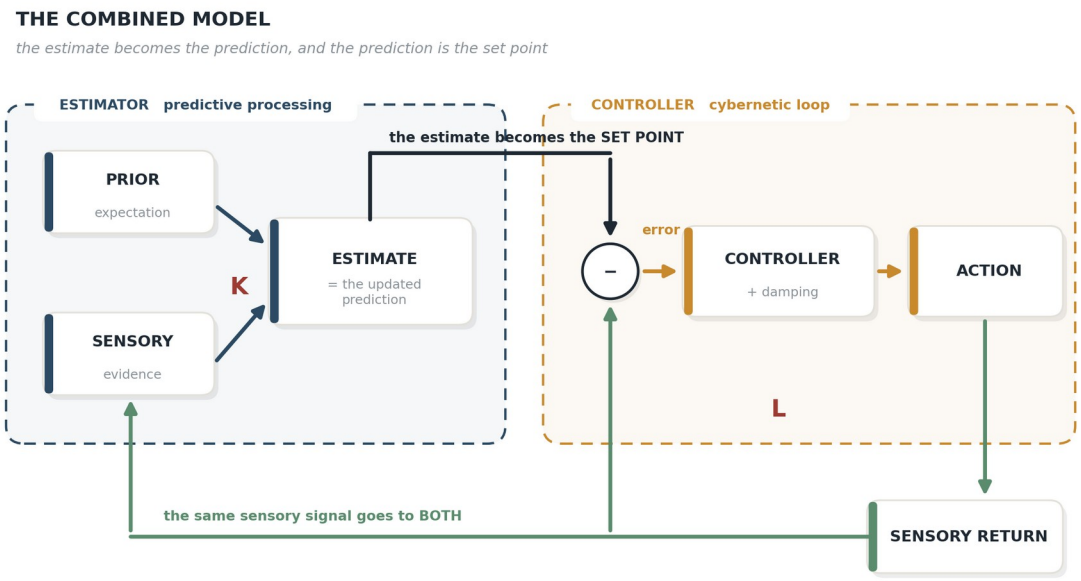

Figure 4.1. The combined model. The estimate of state becomes the updated prediction, which serves as the set point. The sensory return from action feeds both the controller and the estimator.

That system has **two gains, not one**.

Observer gain (K). How far new sensory evidence moves your estimate relative to what you expected. This is precision. It's the entire prior-versus-evidence argument compressed into one parameter.

Controller gain (L). How forcefully you act once the error is on the books. Nothing to do with belief. Two people can reach the identical read of a situation and respond with wildly different intensity, and that difference is L.

Back to the dial. K is how fast you update your model of the room. High K and you learn the delay in one cycle, or you learn the wrong lesson in one cycle, because fast learning isn't automatically accurate learning. Low K and you keep cranking, because your model never catches up to the room you're standing in.

L is how hard you turn the dial once you've decided it's wrong. The knife gets high L and should. Room temperature gets low L and should. Run comfort at knife settings and you're the person cranking.

Do the two gains coexist, or does one feed the other?

Both, and the diagram is easy to misread on this point.

They're in series in the sense that K's output becomes the thing L works on. The estimate sets the target; the controller drives toward it. Change K and you move the target itself. Change L and you change how hard the system pushes toward a target that hasn't moved.

They run simultaneously in the sense that this isn't a relay race. Every cycle, sensation arrives, the estimate updates, the target moves with it, the error recomputes, and the action adjusts. Continuously, at the same time, in a loop that never stops to hand off.

Which is why the clinical distinction holds. A person can have a badly placed target and a calm response, or a well-placed target and a violent one. Those are different problems even though both parameters are always running.

What this architecture commits us to

Once you say the prediction IS the set point, the error the controller sees is a discrepancy between predicted and actual sensation, not a gap between a state and some independently specified goal.

That's the active-inference architecture. It's the side Friston takes (Friston, 2011; Adams, Shipp & Friston, 2013; Friston, Daunizeau, Kilner & Kiebel, 2010), and it's a real position in a live argument. Classical control engineering keeps the goal separate from the estimate, which is what allows an engineer to tune the estimator and the controller independently. Kennaway (2025) argues the two frameworks can't be reconciled at all.

I've built it this way because it's what an organism actually does. Nobody hands you a reference signal. You infer what should be happening and then act to make the world match (or adjust based on your system and what you can control), and the same sensation that tells you whether it worked is the sensation you used to form the belief in the first place.

Emotion, precisely

If an emotion is the error signal, and the estimate feeding that comparison is an *inference* rather than a measurement, then two things follow. Both of them happen constantly, and anyone who does this work has watched them happen.

The reason emotions respond to reframing at all is because inferences move when the evidence or the expectation moves. A thermostat doesn't feel differently about 62 degrees once you explain the situation to it.

And reframing has a ceiling.

> **Changing a high-level belief doesn't automatically change a low-level estimate. "I know it's irrational and I still feel it" isn't a failure of insight.** That sentence is what it feels like when one level of the system has updated and a lower level has not. The part of you that talks has changed its mind. The part of you that runs the estimate is still using the old weighting, and it's the one generating the feeling.

The learning theory

None of these parameters are fixed at birth…there may be an initial set point but then each is estimated from experience. That is, if the system is working properly.

Volatility teaches observer gain. In a stable world a surprise is noise and should be discounted (your system predicts this and dampens a reaction). In an unpredictable one, it's a signal and should move you a long way (your system knows it needs to react to keep up). People demonstrably track this. Someone who grew up in a chaotic house learned to weight incoming evidence heavily, because in that house it WAS heavily weighted evidence.

Consequence teaches controller gain. Where small problems (or errors or trauma) *reliably* became large ones, high gain isn't a malfunction. It's the correct setting for that environment. It was learned correctly.

Overshoot teaches damping. A system punished for waiting learned not to wait.

Volatility and consequence set those parameters together, not separately, and laying them on two axes is more useful than either one alone.

	Low consequence	**High consequence**
Stable environment	Dampened reaction, little prediction adjustment	Reliable prediction, strong reaction
Unstable environment	Fast prediction updating, reaction either way	Rapid prediction updating, strong reaction

Stable, low consequence. The commute is the same every day and being five minutes late costs nothing. Correct setting: barely update, barely react. This is most of ordinary life and it needs no attention.

Stable, high consequence. A surgical checklist. A pre-flight. The world holds still and the cost of error is large, so you want an accurate model and a decisive response. This quadrant produces the most reliable performance of the four, and it's where training and drilling actually pay.

Unstable, low consequence. A new hobby, an unfamiliar city, small talk with strangers. Update fast because the ground keeps moving, and the reaction can be whatever you like because nothing much rides on it. This is the healthiest place to learn anything.

Unstable, high consequence. The ground moves and the cost of being wrong is real. A volatile household. A business quarter that could go either way. Correct setting is rapid updating and strong reaction, and it is exhausting to run for any length of time.

> *An aside… attachment theory fans may consider this model in development of stable, anxious and avoidant styles.*

The two problem quadrants aren't the ones people expect.

The dangerous one is unstable-high-consequence held too long. It is the correct setting while it lasts. Run it for a childhood and you've learned high K and high L on a threat channel, and both keep running after the environment stabilizes. That person doesn't have a distorted view of danger. They have last decade's correct view of danger.

The quiet one is stable-high-consequence learned so well it stops updating. The model is accurate, the reaction is strong, and K has dropped near zero because it never needed to move. Then the world changes and nobody notices, because a prediction that's been right for fifteen years doesn't get checked. That's the expert who gets blindsided, and it's a low-K failure rather than a high-K one.

Which gives the claim:

Most of what presents as dysregulation is a correctly learned parameter running in an incongruent environment.

Incongruent, not wrong. The setting was right where it was learned. The environment changed and the setting didn't. Maybe when we were young we learned the best possible reaction to the environment… then it persists even though the environment has changed.

It shows up two ways. An ineffective K means the estimate itself is off, so the target sits in the wrong place and everything downstream is aimed at it. An ineffective L means the estimate is fine and the correction is disproportionate, or adjusts in the wrong direction entirely. Same presentation, different repair.

Aaron Beck, the psychiatrist who built cognitive therapy in the 1960s and gave us the idea that what you believe about a situation drives how you feel about it, would call the learned part a **schema**: a stored belief about yourself or the world, formed early and applied automatically ever after (Beck, 1967, 1979). He'd be right. What the model adds is where a schema comes from mechanically: it's the stored prior, built from the statistics of an environment the person actually lived in, still being emitted as a prediction in an environment that no longer matches it.

It carries a warning. A parameter learned over fifteen years from consistent evidence won't be revised by one good afternoon of argument. Which is why insight is a poor lever and repeated disconfirming experience is a good one.

5. Fitting this into CBT

Before going anywhere near autism, this has to line up with cognitive behavioral therapy, because CBT is the model I like, it's the model many people reading this relate to, and a framework that can't be stated in its terms isn't worth much.

It lines up pretty well. Term for term, with one piece left over that turns out to matter.

The model I teach

Here's the diagram I've learned from my mentor Ryan Sheade, LCSW. I have put in front of students and clients for years.

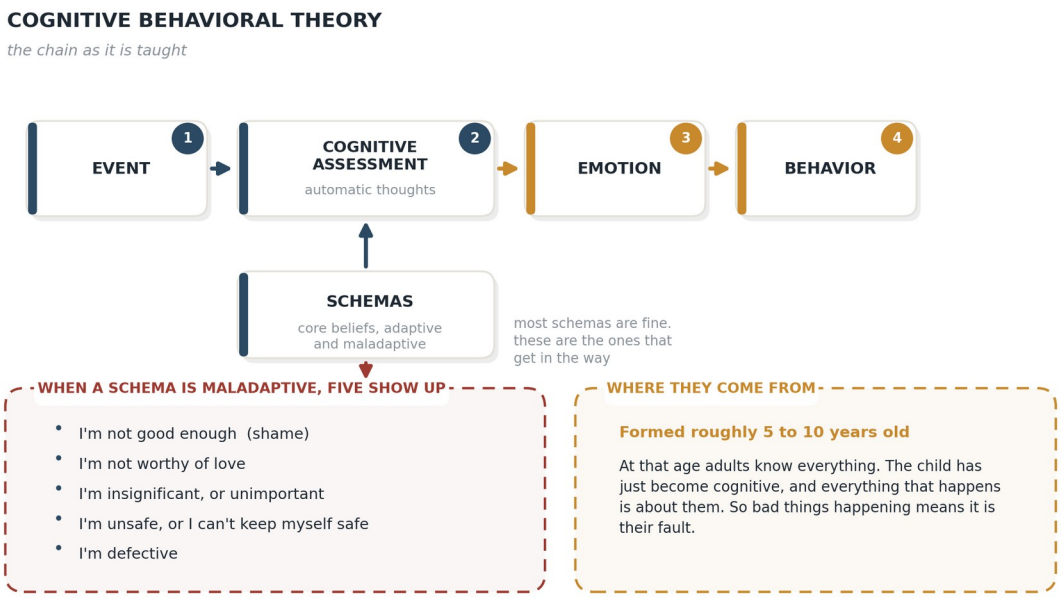

Figure 5.1. The cognitive behavioral model as I represent it.

An event happens. You run a cognitive assessment of it, which shows up as automatic thoughts. That assessment produces an emotion. The emotion produces behavior.

Underneath the assessment sit the schemas, the core beliefs, and those are what the assessment is running on. When they're maladaptive, five show up over and over:

I'm not good enough. That one arrives as shame. **I'm not worthy of love. I'm insignificant, or unimportant. I'm unsafe, or I can't keep myself safe. I'm defective.**

They form early, roughly five to ten years old. From the child's point of view, at that age adults know everything. The child has just become cognitive enough to build general rules out of specific events, and not yet cognitive enough to consider that a rule might be about somebody else. So everything that happens is about them. Bad things happening means something is wrong with them, and the rule gets written down and filed. To be clear, not all core beliefs are maladaptive, it's the maladaptive ones that often get in our way.

The same diagram, drawn as a loop

Now redraw it.

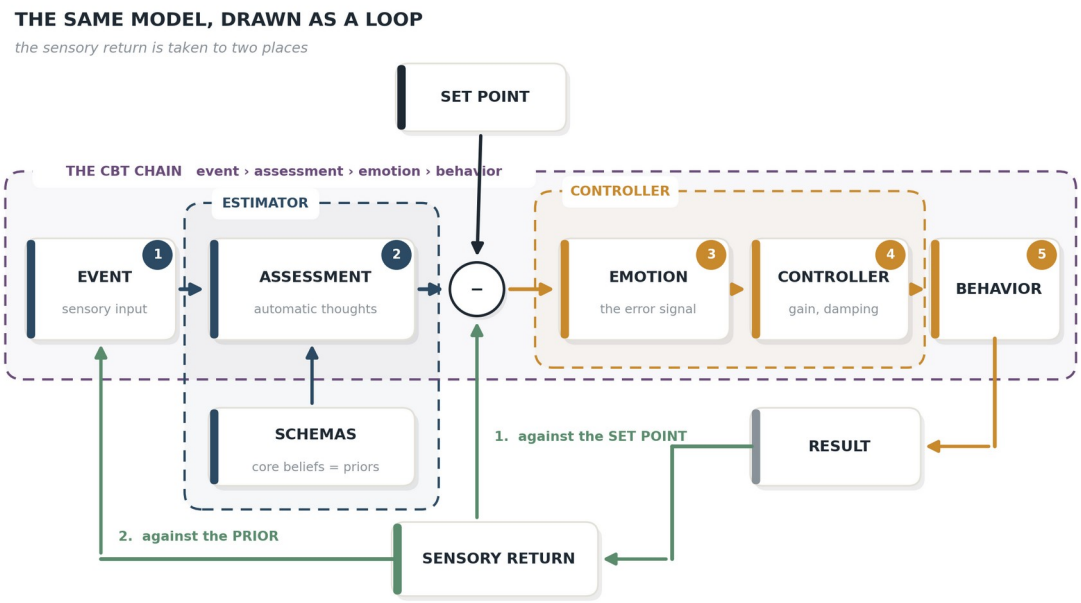

Figure 5.2. The same model as a control loop.

Every term carries over with nothing left standing.

The event is sensory input. Something arrives.

The cognitive assessment is the estimator. It takes the input, weighs it against a prior, and produces an estimate of the situation. Not a description of the situation. An estimate. **The schemas are the priors.** That isn't a metaphor and it isn't a rename, it's the same job: a stored expectation that shapes what ambiguous evidence gets read as.

The emotion is the error signal. It's the gap between the estimate and the set point the system is defending. (The positive emotions, and why happiness may not be an emotion under this account, are in Section 2.)

The behavior is the controller's output. What gets done about the gap, and how hard.

Then the piece the standard diagram doesn't have. The behavior produces a **result**, and the result is the next event. The chain isn't a chain. It's a loop, and it runs continuously.

Two things follow immediately, and the second one is why the redraw was worth doing.

What the loop makes visible

The step from assessment to emotion stops being a mystery. In the standard teaching that arrow is the part that gets waved at, and students accept it because the sequence obviously works even when the mechanism is vague. Under the control reading the emotion is a comparison, and its magnitude is the size of a discrepancy. Which is why the same thought produces a different feeling in two people, and a different feeling in one person on two different days. The thought didn't change. What it was being measured against did.

That also gives the four of the five core beliefs a job, and separates out the fifth. *Not good enough, not worthy of love, insignificant, defective* are all priors. They bias what the estimator concludes from thin evidence. But *I'm unsafe, or I can't keep myself safe* isn't doing that. It's setting how much estimated danger is tolerable before the system calls it an error. It's a set point, not a prior, and it produces a different presentation and needs a different intervention. Under the standard model all five sit in the same box.

And the loop explains why a wrong belief survives contact with reality. This is the part the left-to-right chain can't show, because a chain ends at behavior and reality never gets to answer.

Work an example. The prior is *I'm not worthy of love.*

The event is a partner who's quiet at dinner. The estimator, running on that prior, reads quiet as withdrawal, and withdrawal as the front edge of leaving. The set point is being securely connected. The gap is large, so the emotion is large, somewhere between fear and shame. The controller has high gain and no damping, so the reaction is to press. What's wrong. Are we okay. You'd tell me, right. Three hours of it.

The result is a partner who was tired at the start of the evening and is now irritated and genuinely pulling back.

That result is the next event.

Nobody in that loop did anything irrational. Every step follows correctly from the step before it. The prior was wrong when the evening started and accurate by the time it ended, because the loop went and made it accurate.

That's what a maladaptive core belief is, mechanically. Not a wrong idea sitting in storage waiting to be corrected. A prior running a loop that manufactures its own evidence. Which is also why arguing with the belief so often does nothing: you're disputing a conclusion the system re-derives from fresh data every week.

DON'T say: "You're being irrational." SAY: "I think your instrument is reading off."

> **The anxious person usually isn't over-reacting to their situation. They're reacting accurately to an estimate of their situation that happens to be wrong.**

Does the assessment do both jobs?

Here's the question I keep coming back to, and I don't have an answer to it.

The clean version of Figure 5.2 gives the cognitive assessment one job, which is to produce an estimate. But look at what an assessment actually delivers in a session. It tells you what's happening and how much it matters, in the same breath. *He's pulling away* is an estimate. *He's pulling away and this is the end of everything* is an estimate with a gain setting attached to it.

Appraisal research has said something close to this for decades. The appraisal is held to determine the intensity of the emotion and not only its content (Smith & Ellsworth, 1985). If that's right, the assessment box is doing two mechanically different jobs, and the standard model is treating them as one.

Which opens something the diagram doesn't show. A maladaptive core belief may be able to enter the loop in three different places.

As a prior. *I'm defective* biases what the estimator concludes from thin evidence. Ambiguous feedback gets read as criticism.

As a set point. *I'm unsafe, or I can't keep myself safe* changes what counts as an error at all. Less estimated danger is tolerable, so the same situation opens a gap that wouldn't open in somebody else.

As a gain. *I'm not worthy of love* may not distort the reading whatsoever. It may set how hard the system acts once an error exists, because it reads the stakes as total. The estimate is accurate. The response is maximum.

Same five beliefs. Three mechanisms. Three different treatments, which is the argument of this entire document arriving from a direction I hadn't planned on.

I can't settle it from here and as far as I can tell nobody has. What's worth saying is that it's separable in principle. A belief operating as a prior predicts biased interpretation of *ambiguous* evidence and roughly normal handling of unambiguous evidence. A belief operating as a gain predicts accurate interpretation across the board and an outsized response to both. Hand the same person ambiguous and unambiguous versions of the same situation, and the two accounts come apart.

That's a study, and it's a small one.

The vocabulary, side by side

CBT term	Control term	What the translation adds
Activating event	Sensory input	Nothing. Same thing.
Core belief, schema	Prior	Little on its own. Vocabulary substitution unless precision comes with it.
Automatic thought	The prediction the estimator emits	Separates the sentence from the estimate driving the feeling
Conviction rating, 0 to 10	Where the estimate currently sits	Says nothing about whether it can move. That's a second parameter.
Emotion	Error signal = Emotion	Gives the assessment-to-emotion arrow an actual mechanism

Coping behavior	Corrective action	Ties intensity to gain rather than to the belief
Safety behavior	A correction that removes the signal without testing the estimate	Explains why comfort can be worse here than distress
Behavioral experiment	Prediction-error protocol	Explains precisely how one fails
Exposure	Forced error generation	Moves the active ingredient off habituation
Response prevention	Damping	Explains why the first week gets worse

Most of that right-hand column is a rename. Saying so up front is cheaper than being caught at it later. Three of the rows aren't.

Three places the translation pays

Conviction is not movability. In the cognitive model a belief has content and strength, and strength gets, for example, a number out of a hundred. In the predictive account, the confidence a prediction is held with is a separate parameter from what the prediction says, and it can move without the content moving at all. Calling a schema a prior and stopping there is vocabulary substitution. Starbucks calling a small a tall doesn't put more coffee in the cup.

Two clients both say people think they're stupid. Both rate it 7/10. In the first, the content is wrong and contrary evidence gets in, so restructuring moves the estimate. In the second, the content may be partly accurate somewhere, and the prediction is held so tightly that nothing moves it. Restructure that one and you change the words while the parameter sits untouched. The client agrees with you and feels exactly the same, and both of you leave the session confused about why.

> *Aside… a similar concept: "Affirmations only work if you believe them" – Jim Rohn*

I read conviction ratings for years as though they told me how hard a belief would be to shift. They don't. They tell you where the estimate is sitting. Whether it can be moved is a different question, and it's the one you're actually being paid to answer.

Expectancy violation and prediction error are the same quantity. Craske, Treanor, Conway, Zbozinek and Vervliet (2014) argue that exposure works by building a new inhibitory association rather than erasing the old fear, and that the target is the mismatch between what the client expected and what happened, not the reduction of fear inside the session. Read that in control terms and it isn't an analogy, it's the same quantity under a different name. Craske says the therapeutic work is done by the size of the gap between what somebody expected and what actually happened. This model says the estimate updates in proportion to the size of the prediction error. Those are two descriptions of one

number. She got there from exposure research and I got here from control engineering, and we are both pointing at the gap.

The control account doesn't improve on Craske. It agrees with her, which is the most useful thing it does in this entire section, because her version already has evidence behind it.

DON'T ask: "How anxious are you now compared with when we started?" SAY: "What did you predict would happen? What actually happened?"

The same reading sharpens why a behavioral experiment fails, and there are only three ways. The prediction was never stated precisely enough to be violated, because "it will go badly" can't generate a clean error. The prior was held at such high precision that the discrepancy got absorbed as a fluke. Or a covert safety behavior meant the experiment tested a different prediction from the one written on the form. All three are prevented by the same step, which is writing the prediction down before anything happens. Skip it and you get the Home Depot outcome. One trip for the part you thought you needed, two more for the parts the job actually required.

"I know it's irrational but I still feel it." Stott (2007) called this rational-emotional dissociation, and clients produce it in close to those exact words. The predictive reading is specific. Verbal work can move the high-level prior, the one that generates sentences, and leave the precision on lower-level predictions exactly where it was. The explicit model updated. The model actually running the interoceptive estimate did not. Both are running. Only one of them talks.

Which fits something worth stating plainly: people decide emotionally and then assemble the logical case afterward. A Vulcan couldn't run a real business. Every clinician reading this has watched somebody build an airtight argument for a conclusion they had reached before they sat down.

~~I am afraid to be wrong in front of these people.~~ "I just want to make sure we're all aligned before I commit."

The argument isn't the cause. It's the transcript.

Four intervention families instead of two

Separating the estimator from the controller splits the standard toolkit into four targets, and the reason to bother is that two of them present identically.

Target	What it changes	Interventions
Prior, content	The expectation being emitted	Restructuring, thought records, psychoeducation, formulation, meaning-making
Observer gain, precision	How tightly the prediction is held, independent of content	Behavioral experiments, exposure designed for expectancy violation,

		attention training, tolerating uncertainty
Controller gain	Intensity of the response, without touching the belief	Arousal regulation, paced breathing, applied relaxation, distress tolerance skills
Damping	Whether corrections converge or oscillate	Delay, urge surfing, response prevention, scheduled worry time, ritual postponement

The rows aren't clean and I'm not going to pretend they are. A behavioral experiment moves precision and usually content with it. Mindfulness reallocates attention and lowers arousal at the same time. The question the table answers is what an intervention is mainly for.

Rows two and three are where it matters. A person with an accurate belief and a controller that overshoots, and a person with an inaccurate belief held at high precision, both arrive distressed, both describe an intense reaction, and both may rate conviction at 80. Restructuring is close to useless for the first and central for the second. Arousal work is central for the first and a distraction for the second. Standard assessment doesn't separate them, because standard assessment asks what somebody believes and how strongly, and both of those questions land on the same row.

DON'T stop at: "What do you believe, and how convinced are you?" ALSO ASK: "What would it take to change your mind?" (precision). "What did you do about it?" (controller gain). "How many times did you do it?" (damping).

All three are answerable inside an ordinary intake, and they point at different treatments.

One warning falls straight out of the table. Breathing retraining sits in the controller row, and in panic it can quietly become a safety behavior, a correction that removes the error signal and blocks the test. Salkovskis, Clark and Hackmann (1991) treated panic successfully with cognitive therapy and no breathing retraining at all, which is worth remembering. Lowering controller gain helps when the estimate is accurate. It hurts when it becomes the thing preventing the estimate from being corrected.

What this doesn't do

It doesn't add a technique. Every intervention in that table already existed, most with better evidence behind it than anything in this document. What changes is the selection rule. This is an effort to improve the model (or maybe just add more detail).

And the dismantling data should keep anyone honest about how much that's worth. Ougrin (2011) pooled 20 randomized trials, 1,308 participants, comparing cognitive therapy against exposure, and found no significant difference in panic, PTSD or OCD, with cognitive therapy ahead only in social phobia. Öst, Havnen, Hansen and Kvale (2015) found the same

inside OCD specifically: exposure and response prevention versus cognitive therapy came out at an effect size of 0.07, which is nothing.

If two routes reach the same outcome, a model explaining the mechanism of one route has not thereby explained the disorder's treatment. Whatever is doing the work is reachable from both directions, and the model I've built predicts a separation that decades of outcome data have not produced.

The clean answer is that the trials never sorted people by which parameter was broken, so the two groups washed out against each other. That answer is also exactly the kind of unfalsifiable patch Section 9 says should make you suspicious. It's testable, which is the only thing that redeems it, and until somebody runs it, CBT works better than this model can currently explain.

6. Autism under this model

Autistic sensory-perceptual differences sit on the observer side. Repetitive behavior, insistence on sameness, and difficulty settling sit on the controller side.

The autism literature has spent fourteen years arguing about K. Controller gain has never been modeled, measured, or discussed.

That matters, because repetition is what an oscillating loop produces. Four findings already fit, from independent groups, each reported as a separate deficit:

Excess low-frequency force oscillation below 0.25 Hz during **dynamic** tracking only, specific to autism against ADHD and fetal alcohol controls (Lidstone et al., 2020; n = 22 autistic, 18 ADHD, 17 FASD, 22 TD). Increased postural trembling, **worse with eyes open** than closed (Bloomer et al., 2025; n = 49 autistic, 94 controls). Reduced dynamical complexity of center-of-pressure trajectories (Pettinato et al., 2024, a feasibility study, n = 11 per group), independently pre-replicated a decade earlier by Fournier et al. (2014). And at high visual display gain, force output that became **more regular**, alongside elevated variability at both the lowest and highest gains (Mosconi et al., 2015; n = 28 autistic, 29 controls).

Worse with more feedback is not a sensory-overload signature. It's a stability signature.

Now the part I got wrong in an earlier draft, corrected here because it's the kind of error that should be printed rather than quietly fixed. I originally described Mosconi as showing force irregularity that worsened as visual gain rose. It shows the opposite. Approximate entropy went **down**, meaning the output got more regular and more periodic, most markedly at the highest gains. And force variability was U-shaped, elevated at the smallest **and** the largest gains.

That matters, because a simple over-gained loop predicts that lowering the gain should improve stability. Mosconi's data show instability returning at low gain too. That's a system that can't optimize across a range of feedback conditions, which is not the same claim I was making.

What this changes in practice

Predictability and sensory intensity are different variables. Intervene on them separately. A quiet room that's unpredictable can be worse than a noisy one that isn't. That single distinction is the most actionable thing in this document, and it doesn't depend on the model being right.

Suppressing stimming removes a regulatory output. If the behavior is correcting an error, taking it away without addressing the error leaves the error uncorrected and removes the correction.

That statement is worth more than one line, because most clinicians were trained to treat stimming as a symptom to reduce and were never told what it does.

What stimming actually does

Rocking, hand-flapping, spinning, pacing, finger-tapping, repetitive vocalizing, skin-picking. The behaviors are familiar. The function is not, and when autistic adults are asked directly, they describe at least five distinct jobs. All quotations below are from published interview research.

It regulates arousal in both directions. Not calming. Regulating. One participant in Kapp et al. (2019) put both poles in a single sentence: *"[s]timming is just a release of any high emotion, so really anxious, really agitated, really happy, really excited, just any high emotion, that's when I stim."* A clinician watching only for distress will miss half the instances and misread the other half.

It sets a rhythm the body can organize around. One participant described the mechanism better than the literature does: *"it sort of metronomes everything in your body to sort of go at that speed … So it just sort of helps quell everything, because you're at the same rhythm with everything."* That is a person describing a reference signal they generate for themselves.

Which is where the phase-locked loop comes in. A phase-locked loop regulates the timing of an internal oscillator against an external rhythm. It has a capture range, meaning it can only lock onto inputs close enough to its own natural frequency. And when it can't lock at all, it doesn't go quiet. It **free-runs at its own natural frequency**.

Read stimming that way and the metronome quote stops being a metaphor. A loop that can't capture external structure runs at its own rate instead. That predicts something the plain predictability account doesn't: stimming should INCREASE when external rhythm is present but unlockable, meaning irregular or at the wrong tempo, and decrease when rhythm is either absent or well matched. Worst case is structure that almost works.

It also restates the double empathy problem in one line. Two oscillators with mismatched natural frequencies can't lock to each other, while two matched ones lock easily whatever their frequency. Crompton et al. (2020) found exactly that: information transfer in autistic-to-autistic chains was as effective as in non-autistic chains, and both beat mixed chains.

Untested, and one result cuts against it. Beker, Foxe and Molholm (2021) found neural entrainment to rhythm *reduced* in autistic children, which supports reduced capture but leaves the free-running half needing an intact internal oscillator nobody has measured.

It filters and protects attention. Another: it *"helps me keep me calm … it cuts down what is going on around, it helps me focus."* Stimming here isn't competing with concentration. It's what makes concentration possible.

It heads off escalation before it starts. One participant found this by accident: *"it actually managed to help me stave off some panic attacks. For example, I never used to wave my hands that much, but I've started doing it more and it actually helps, like if I'm in a crowded elevator or something."* That's not soothing after the fact. That's a correction applied early, while the error is still small.

It communicates. Some stims are read by people who know the person as a signal that the load is climbing. Removing the behavior removes the signal, and the first thing anyone learns about is the meltdown.

And it has a genuine cost, which the same interviews report plainly. One participant on skin-picking: *"because you can get into a loop and you can start really making your fingers leathery if you're not careful."* Some stims injure. That's a real clinical problem and it deserves a real answer.

The answer the model gives is specific. **If a stim is causing harm, the target is the harm, not the regulation.** Find what the loop is correcting and give it another output that does the same job. Substitute, don't subtract. A person who paces because pacing sets their rhythm needs a different rhythm, not stillness.

Note: nobody has measured autonomic state during spontaneous stimming in autistic people. The evidence above is what autistic adults report about their own experience, which is the best evidence that exists and is not the same as a physiological demonstration.

What follows from that is a posture rather than a technique. The burden of proof sits with suppression. If you're going to remove a behavior a person uses to regulate, you should be able to say what it was doing and what's replacing it.

Treat the model as a hypothesis. The four supporting findings are reinterpretations of data collected for other purposes. That's legitimate. It's also weaker than predicting a result in advance.

The counter-arguments…

A clinician reading this will find these, so they belong in the document rather than in a rebuttal.

Reduced entropy reads as rigidity, not ringing. This is the hardest one, because it sits inside the supporting evidence rather than outside it. Two of the four findings are entropy reductions, and in the postural literature reduced complexity is standardly interpreted as

constrained, over-regularized control (Fournier et al., 2014). An under-damped loop and an over-constrained loop both lower entropy. Nothing in these datasets distinguishes them.

The frequency data don't look like resonance. A classically under-damped system produces a resonant peak at a characteristic frequency. What the autism data show is power shifting into the lowest band with *reduced* power in the mid bands (Mosconi et al., 2015, group × frequency bin $F(1,55) = 15.87$, $p = 10^{-4}$), alongside reduced entropy and, in Mosconi's own words, "increased reliance on slower feedback mechanisms." Slower and more regular is at least as consistent with sluggish low-bandwidth control as with ringing.

Proprioceptive over-reliance explains all four findings without any gain term. Five studies, three labs, sixteen years: autistic motor learning weights proprioception, the sense of where your own limbs are without looking, over vision (Haswell et al., 2009; Izawa et al., 2012; Marko et al., 2015; Hirai et al., 2021; Lidstone et al., 2025), with a cerebellar volumetric correlate. Worse dynamic visual tracking, worse when vision dominates, worse when somatosensory input is corrupted, degraded visuomotor transformation at extreme gains. The loop isn't over-gained on that account. It's listening to the wrong sensor. This is the account the model has to beat, not merely mention.

Motor difference isn't autism-specific. Depending on the instrument, 86.9% (Bhat, 2020, n = 11,814) or 35.4% (Licari et al., 2020, n = 2,084) of autistic children screen positive for motor impairment. A base rate that swings by a factor of 2.5 can't anchor a precision argument. Motor and postural differences are also well documented in ADHD, where higher joint torques maintain the same posture, which is itself an over-gain description in a different diagnosis.

IQ and sensory severity outpredict diagnosis. Surgent et al. (2021, n = 110) found motor ability "best characterized by individual variation in sensory features and cognitive abilities rather than diagnostic group."

The findings contradict each other. Stania et al. (2023) used Bloomer's exact rambling–trembling decomposition and found the reverse: rambling differed, trembling did not. And Lim et al. (2020) found no postural group difference at all in children, from the same lab that published the meta-analysis the field cites for autistic postural impairment.

Some of it normalizes with age. Shafer et al. (2025, n = 54 autistic, 31 neurotypical) found force variability and regularity differences present in childhood that resolved by adolescence. A constitutional loop parameter shouldn't do that.

And the strongest evidence for a gain abnormality puts it somewhere else. Doumas et al. (2016) experimentally manipulated sensory gain and found autistic participants at gain 1.0 performed like controls at gain 1.6, which quantifies hyper-reactivity at roughly 1.6×. Their conclusion locates it in **sensory integration**, not motor output. A hyper-reactive sensor and an under-damped controller produce similar traces, and nothing here separates them.

Where that leaves it: the over-gained, under-damped reading is an untested hypothesis rather than an inference the data support. It generates falsifiable predictions, which is worth something. It doesn't currently have the evidence behind it, which is worth saying.

7. What it means for everyday practice

The chain the model proposes:

> learning history → prior → prediction → the estimate the system regulates against → error signal → corrective action → the action prevents disconfirmation → the prior is reinforced

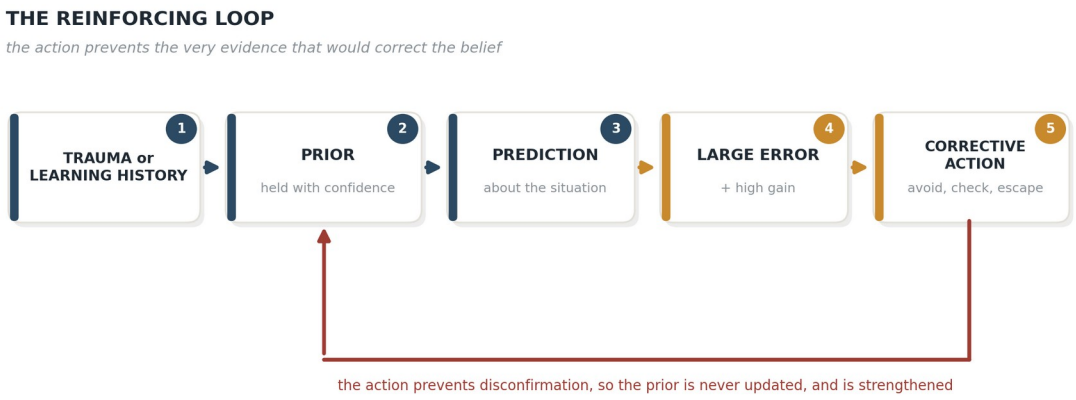

Figure 7.1. The reinforcing loop.

Most of that is standard cognitive therapy in different words, and I'd rather say so than oversell it. Beck's schemas are priors. Safety behaviors preventing disconfirmation is Salkovskis, tested experimentally, and control theory contributed nothing to it. The overlap isn't a weakness in the argument. It's the reason to trust the argument, because a mechanical account that contradicted decades of outcome data would be the account with the problem.

What the layer adds is a separation the cognitive model runs together.

DON'T say: "You're being irrational." SAY: "I think your instrument is reading off."

That difference isn't cosmetic, conceptually or in the room.

> **The anxious person usually isn't over-reacting to their situation. They're reacting accurately to an estimate of their situation that happens to be wrong.**

The four questions

Is the belief inaccurate, or accurate and acted on too hard? Restructuring addresses the second and does little for the first. The first needs work on response intensity: arousal regulation, distress tolerance, deliberate delay.

The client who says their boss is unhappy with them, and the boss is in fact unhappy with them. Nothing to restructure. The estimate is correct. What's disproportionate is the response, which is three days of rumination and a resignation letter drafted at midnight. Restructuring here tells an accurate person they're wrong, and they will correctly conclude you're not listening.

Is the belief wrong, or held with too much confidence? Content and confidence come apart. A moderately accurate belief held with total certainty behaves worse than a somewhat inaccurate one held loosely, because certainty blocks updating. Behavioral experiments work on confidence. That's why they beat argument.

The client who is sure they'll be fired. Maybe they're right, maybe not. Arguing the content invites them to defend it, and they'll win, because you don't have the data either. Asking what would count as evidence, and then going and getting some, works on the certainty instead. The belief may survive. The grip loosens.

Is the person failing to update, or being prevented from updating? Safety behavior stops the error signal from ever being generated. The prior isn't stubborn. It's starved.

The client who's had thirty good presentations and still believes they'll humiliate themselves. Look at what they did in each one. Over-rehearsed, read from the page, left immediately. They never ran the experiment. They ran a controlled version that's fully compatible with the belief.

A phobia is a great example case. Say I'm afraid of heights. The prediction is that height means danger, the error is large (emotion=fear!), and the corrective action is to get away from the edge. The action works, in the narrow sense that the fear drops. It also guarantees I never collect the evidence that would revise the estimate, which is standing near an edge and not falling (it's not unsafe). Every successful escape confirms the model rather than providing new data to revise the estimate. Thirty years of that and the prior is untouched, not because it's stubborn but because it has never once been given new data that revises the estimate.

Flooding and systematic desensitization work by forcing the data through. Stay near the edge long enough, without the escape, and the sensory return finally contradicts the prediction. The estimate moves, the prediction resets, and the error the controller sees gets smaller on its own. You didn't argue anyone out of anything. You restored the evidence stream the avoidance (the learned emotion which was driving behavior) was blocking.

What environment taught this setting, and does it still exist? Usually answerable, usually informative, and it changes the conversation from correction to recognition.

The client who reads a two-word text as a catastrophe. Ask where a two-word message used to mean something was badly wrong. Frequently there's an immediate answer. The setting was correct in that scenario. The current sender (new environment) is just terse.

One more thing worth noticing

Rumination, worry, and compulsions all look like oscillation. A loop that can't settle, cycling at a characteristic frequency, each pass generating a correction that doesn't resolve the error.

Under that reading, "not acting" is therapeutic in a specific mechanical sense. You're adding damping.

Take the client who checks the lock. The urge arrives, the correction is immediate, the urge returns in ten minutes, and the cycle runs all evening. Standard response prevention asks them to feel the urge and not check. In control terms, you've inserted a delay between the error and the correction, which is the definition of damping. The first few cycles get worse, because an under-damped loop always overshoots before it settles. Then the oscillation loses amplitude and the intervals stretch out.

The same logic covers the worry spiral where you agree to postpone the worrying to a set hour, and the argument where the rule is to wait a day before replying. In every case the intervention isn't insight and it isn't a better answer. It's slowing the loop down enough to let it settle.

This also explains why telling someone to stop worrying doesn't work, it just raises the gain, because now failing to stop is a second error to correct. Adding damping works. Adding pressure adds gain.

8. What the evidence actually supports

The citation counts in this area may overstate the evidence, and anyone using this framework should know the numbers.

The largest synthesis of Bayesian accounts of autism, 83 studies, reports that a **slight majority find no group difference** in the integration of priors at all (Angeletos Chrysaitis & Seriès, 2023). The second largest, 47 studies, reaches the opposite conclusion (Cannon, O'Brien, Bungert & Sinha, 2021). The two biggest reviews of the same literature disagree about what it shows, which is worse for the field than either verdict alone.

The first meta-analysis of the simple model pooled 24 effect sizes and found a small one, $g = 0.37$, with heterogeneity that no moderator could explain, concluding that the evidence offers "only limited support for a universal 'simple Bayesian model' of autism" (Cui et al., 2026).

The flagship computational finding, elevated volatility estimation in autistic adults (Lawson, Mathys & Rees, 2017), **failed a preregistered test** that used the identical computational model in a new task domain, finding no credible group difference in either

volatility or learning-rate updating (Plank et al., 2026). That is an extended replication rather than a direct one, which the authors say plainly, and it is still the closest thing to a replication attempt the finding has had.

The mismatch negativity, described everywhere as the neural signature of aberrant prediction error, pools to **no significant effect** (standardized mean difference 0.15, 95% CI [−0.01, 0.32], $p = .07$), and the sign reverses with age: the youngest autistic cohorts show smaller amplitudes than controls and adult cohorts show amplitudes equal to or larger than their peers, with age accounting for a quarter of the variance in effect size (Schwartz, Shinn-Cunningham & Tager-Flusberg, 2018). A larger meta-analysis puts a number on each side of the flip: reduced in autistic children and adolescents ($g = 0.25$), increased in autistic adults ($g = −0.26$) (Sapey-Triomphe et al., 2025).

The predictive account of stimming has been asserted in high-profile theory papers since 2014 and never tested. Nobody has measured autonomic state time-locked to spontaneous stimming in autistic people. The nearest attempt is from 1998.

The endorphin explanation is worse than untested. The historical opioid theory ran the opposite direction, proposing chronically **elevated** opioid tone, which is why the drug tested was an opioid *antagonist*. Naltrexone got three decades of randomized trials and failed on core symptoms and stereotypy, with one well-powered trial finding stereotypy got worse.

Sensory integration therapy shows an inverse relationship between study quality and positive findings. When rigor and results run in opposite directions, the literature is telling you something.

9. Where the model falls apart

It's flexible enough to fit anything. Free parameters for set point, two gains, damping, delay, and arbitration threshold can accommodate almost any behavior. That's the same objection this document levels at the literature it reviews, and adding a control layer makes it worse, not better.

The defense against that objection is not argument but falsifiability. A framework that can accommodate any possible observation makes no empirical claim, because a claim is defined by what it rules out. The criterion is Popper's (2002/1935) and it applies here without modification: the content of a hypothesis lies in the observations it forbids. A model carrying this many free parameters therefore has to specify those observations in advance, before the data are in, or it is a post-hoc interpretive scheme rather than a scientific proposal.

Here are four. **The model forbids all of them:**

- **Repetitive behavior with no frequency structure.** If stereotypy has no characteristic frequency, and no relationship to motor oscillation measured in the same person, there's no oscillating loop and the central claim is dead.

- **Damping interventions failing while sensory reduction succeeds.** If adding lag and smoothing feedback does nothing, and simply turning the volume down works, then the problem was sensory load all along and the control reading is decoration.
- **Observer and controller differences that never dissociate.** If belief accuracy and response intensity always move together across people, there aren't two dials. There's one, and this whole document is a long way of describing it.
- **Gain manipulations leaving repetitive behavior untouched.** Amplify displayed error and repetitive behavior should rise. If it doesn't move, the mechanism isn't what I've claimed.

If those results come in, the model is wrong and either needs adjustment or just throw it out.

The identifiability problem survives. In a Gaussian update only the ratio of sensory to prior precision enters the posterior. "They trusted their senses too much" and "they held their expectation too loosely" are the same claim from a different point of view for most data.

The autism evidence is reinterpretation, not prediction. Four findings that fit aren't impressive if there are forty that don't and nobody counted.

The separation may not be real. Friston argues there's no separate controller to bolt an estimator onto, because the goal is a prior inside the model rather than a cost outside it (Friston, 2011, *Neuron* 72(3), 488–498; Adams, Shipp & Friston, 2013, *Brain Struct Funct* 218(3), 611–643). Kennaway (2025) argues from the other direction that the two frameworks are fundamentally incompatible. If either is right, the central move here is an engineering convenience projected onto a system that doesn't work that way.

It says almost nothing about meaning. Control systems regulate variables. Much of the work is narrative, relationship, identity, and grief. This describes a regulatory layer. It isn't a theory of persons.

Nobody has tested any of it. Not the oscillation account, not the observer/controller dissociation, not the clinical chain. This is a framework proposal.

10. Where this goes

Six studies would settle most of it, and two need no new participants: a frequency-domain reanalysis of existing motor data to separate noise-tracking instability from resonant instability, and adding repetitive behavior as a dependent variable to the visual-gain paradigm that already exists.

The one that would change practice is a clinical dissociation study. If observer gain and controller gain separate in real people, then belief accuracy and response intensity should dissociate across clients, and those two groups should respond differently to cognitive versus arousal-focused intervention. That's a testable prediction about treatment matching.

Three things don't require waiting:

Ask the two questions separately. How confident is this belief, and how hard is this person acting on it.

Treat predictability and intensity as distinct variables in sensory distress.

And when a pattern looks maladaptive, ask what environment would have made it correct.

A closing note

Here's where I'm supposed to hand you the framework that ties it together and tell you it changes everything, but that wouldn't help.

Ask yourself if this seems to fit, and if you can explain it in a therapeutic way. Is it the predictive system (learned behavior, schema) or the controller (the reaction to the input is too high)?

I wanted to write this down and invite disagreement, comment and independent opinions. Perhaps this can be useful.

Full treatment, including published first-person accounts from autistic adults, the complete treatment-evidence audit, the falsification criteria, and the phase-locking extension, is in the long version. The underlying evidence audit is in the companion literature review.

Appendix A. How this differs from *Psycho-Cybernetics*

Anyone who has spent time in the self-help section will have noticed that a plastic surgeon got to this vocabulary in 1960 and sold more than thirty million copies with it. Maxwell Maltz's *Psycho-Cybernetics* borrowed Wiener's engineering language, applied it to the self, and became one of the most influential business and sports psychology books ever written. It deserves a straight answer rather than a sneer, so here it is.

What Maltz got right. His central claim is that behavior organizes itself around a defended reference, and that the reference sets the ceiling rather than effort does. That's a set point, described in a mass-market paperback a decade before control theory reached clinical psychology in any serious way. He also saw that the system is goal-seeking and self-correcting rather than driven, which is the same insight Powers formalized in 1973. The observation was sound. Coaches have been getting mileage out of it for sixty years for a reason.

What's missing is the estimator. Maltz has a servo-mechanism and a target and essentially nothing between the world and the target. There is no separate machinery that takes sensory evidence, weighs it against a prior, and produces an estimate that the set point then tracks. Which is precisely the hole this document exists to fill. Without an estimator you cannot distinguish a person whose reading of the situation is wrong from a person whose reading is right and whose response is too large, and that distinction is the entire clinical payload of Section 5. Maltz collapses both into self-image.

Three further differences follow from that one.

One master set point instead of many. Maltz runs everything through a single global reference, the self-image, and treats it as the thing that determines outcomes across domains. The account here has many parallel governors, each defending its own variable, competing for one body. That's why a person can be professionally confident and socially terrified at the same time without any contradiction, which a single-self-image model has to explain away.

Set points move on imagination. Maltz's mechanism of change is mental rehearsal: picture the outcome vividly enough and the servo will steer toward it. Under this account the estimate moves on evidence weighted by precision, which is why visualizing yourself calm at the edge of a roof does close to nothing and standing at the edge of a roof does a great deal. The distinction isn't academic. It is the difference between a technique with fifty years of outcome trials behind it and a technique without.

There is no evidence base. Maltz worked from case observation, uncontrolled, with no falsification criteria and no way to be wrong. Section 9 of this document lists four ways the model proposed here could be shown false. *Psycho-Cybernetics* offers none, and a system that cannot fail a test is not making a claim about the world.

The twenty-one days

Maltz is also the origin of the most repeated false number in behavior change, and the story is worth telling accurately because it is the same failure this whole document keeps finding in the research literature.

What Maltz actually wrote is that it requires a minimum of about twenty-one days for an old mental image to dissolve and a new one to jell. He was describing how long his surgical patients took to stop seeing their old face in the mirror, and how long amputees took to stop feeling a limb that was gone. A careful observation about psychological adjustment to a changed body, hedged twice, in the same sentence.

What got repeated dropped "minimum," dropped "about," and swapped "mental image" for "habit." A clinical field note became a law of behavior change, and it is now printed on wall calendars.

The actual figure comes from Lally, van Jaarsveld, Potts and Wardle (2010), who tracked 96 people adopting a new daily behavior and measured the time to automaticity. The median was 66 days. The range ran from 18 to 254. There is no single number, which is the finding, and any program built on a fixed one is built on a misquotation of a surgeon's observation about faces.

Maltz was more careful than the people who cite him. That pattern shows up in Section 8 of this document, and on nearly every page of the companion literature review. The damage is almost never done by the person who made the original claim. It is worth noticing that the most durable damage in a field usually isn't done by the person who made the original claim.

Key sources

Bloomer, B. F., et al. (2025). Postural sway dynamics in adults across the autism spectrum. *Molecular Autism, 16*(1), 44. https://doi.org/10.1186/s13229-025-00676-y

Kapp, S. K., et al. (2019). "People should be allowed to do what they like": Autistic adults' views and experiences of stimming. *Autism, 23*(7), 1782–1792. https://doi.org/10.1177/1362361319829628

Adams, R. A., Shipp, S., & Friston, K. J. (2013). Predictions not commands: Active inference in the motor system. *Brain Structure and Function, 218*(3), 611–643. https://doi.org/10.1007/s00429-012-0475-5

Beker, S., Foxe, J. J., & Molholm, S. (2021). Oscillatory entrainment mechanisms and anticipatory predictive processes in children with autism spectrum disorder. *Journal of Neurophysiology, 126*(5), 1783–1798. https://doi.org/10.1152/jn.00329.2021

Crompton, C. J., Ropar, D., Evans-Williams, C. V., Flynn, E. G., & Fletcher-Watson, S. (2020). Autistic peer-to-peer information transfer is highly effective. *Autism, 24*(7), 1704–1712. https://doi.org/10.1177/1362361320919286

Friston, K. (2011). What is optimal about motor control? *Neuron, 72*(3), 488–498. https://doi.org/10.1016/j.neuron.2011.10.018

Friston, K., Daunizeau, J., Kilner, J., & Kiebel, S. J. (2010). Action and behavior: A free-energy formulation. *Biological Cybernetics, 102*(3), 227–260. https://doi.org/10.1007/s00422-010-0364-z

Kennaway, R. (2025). Perceptual control theory and the free energy principle: A comparison. *Current Opinion in Behavioral Sciences, 65*, 101575. https://doi.org/10.1016/j.cobeha.2025.101575

Lidstone, D. E., et al. (2020). Children with autism spectrum disorder show impairments during dynamic versus static grip-force tracking. *Autism Research, 13*(12), 2177–2189. https://doi.org/10.1002/aur.2370

Bhat, A. N. (2020). Is motor impairment in autism spectrum disorder distinct from developmental coordination disorder? A report from the SPARK study. *Physical Therapy, 100*(4), 633–644. https://doi.org/10.1093/ptj/pzz190

Doumas, M., McKenna, R., & Murphy, B. (2016). Postural control deficits in autism spectrum disorder: The role of sensory integration. *Journal of Autism and Developmental Disorders, 46*(3), 853–861. https://doi.org/10.1007/s10803-015-2621-4

Fournier, K. A., Amano, S., Radonovich, K. J., Bleser, T. M., & Hass, C. J. (2014). Decreased dynamical complexity during quiet stance in children with autism spectrum disorders. *Gait & Posture, 39*(1), 420–423. https://doi.org/10.1016/j.gaitpost.2013.08.016

Haswell, C. C., Izawa, J., Dowell, L. R., Mostofsky, S. H., & Shadmehr, R. (2009). Representation of internal models of action in the autistic brain. *Nature Neuroscience, 12*(8), 970–972. https://doi.org/10.1038/nn.2356

Izawa, J., Pekny, S. E., Marko, M. K., Haswell, C. C., Shadmehr, R., & Mostofsky, S. H. (2012). Motor learning relies on integrated sensory inputs in ADHD, but over-selectively on proprioception in autism spectrum conditions. *Autism Research, 5*(2), 124–136. https://doi.org/10.1002/aur.1222

Licari, M. K., et al. (2020). Prevalence of motor difficulties in autism spectrum disorder: Analysis of a population-based cohort. *Autism Research, 13*(2), 298–306. https://doi.org/10.1002/aur.2230

Lim, Y. H., et al. (2020). Effect of visual information on postural control in children with autism spectrum disorder. *Journal of Autism and Developmental Disorders, 50*(9), 3320–3325. https://doi.org/10.1007/s10803-019-04182-y

Marko, M. K., Crocetti, D., Hulst, T., Donchin, O., Shadmehr, R., & Mostofsky, S. H. (2015). Behavioural and neural basis of anomalous motor learning in children with autism. *Brain, 138*(3), 784–797. https://doi.org/10.1093/brain/awu394

Mosconi, M. W., Mohanty, S., Greene, R. K., Cook, E. H., Vaillancourt, D. E., & Sweeney, J. A. (2015). Feedforward and feedback motor control abnormalities implicate cerebellar dysfunctions in autism spectrum disorder. *Journal of Neuroscience, 35*(5), 2015–2025. https://doi.org/10.1523/JNEUROSCI.2731-14.2015

Shafer, R. L., et al. (2025). Visual feedback and motor memory contributions to sustained motor control deficits in autism spectrum disorder across childhood and into adulthood. *Journal of Neurodevelopmental Disorders, 17*(1), 26. https://doi.org/10.1186/s11689-025-09607-7

Stania, M., et al. (2023). Modulation of center-of-pressure signal in children on the autism spectrum: A case-control study. *Gait & Posture, 103*, 67–72. https://doi.org/10.1016/j.gaitpost.2023.04.018

Surgent, O. J., Walczak, M., Zarzycki, O., Ausderau, K., & Travers, B. G. (2021). IQ and sensory symptom severity best predict motor ability in children with and without autism spectrum disorder. *Journal of Autism and Developmental Disorders, 51*(1), 243–254. https://doi.org/10.1007/s10803-020-04536-x

Pettinato, F., et al. (2024). Dynamical complexity of postural control system in autism spectrum disorder. *Journal of NeuroEngineering and Rehabilitation, 21*(1), 225. https://doi.org/10.1186/s12984-024-01520-9

Pezzulo, G., Rigoli, F., & Friston, K. (2015). Active inference, homeostatic regulation and adaptive behavioral control. *Progress in Neurobiology, 134*, 17–35. https://doi.org/10.1016/j.pneurobio.2015.09.001

Salkovskis, P. M., et al. (1999). An experimental investigation of the role of safety-seeking behaviours in the maintenance of panic disorder with agoraphobia. *Behaviour Research and Therapy, 37*(6), 559–574. https://doi.org/10.1016/S0005-7967(98)00153-3

Stephan, K. E., et al. (2016). Allostatic self-efficacy: A metacognitive theory of dyshomeostasis-induced fatigue and depression. *Frontiers in Human Neuroscience, 10*, 550. https://doi.org/10.3389/fnhum.2016.00550

Sterling, P. (2012). Allostasis: A model of predictive regulation. *Physiology & Behavior, 106*(1), 5–15. https://doi.org/10.1016/j.physbeh.2011.06.004

Bennett-Levy, J., Butler, G., Fennell, M., Hackmann, A., Mueller, M., & Westbrook, D. (Eds.). (2004). *Oxford guide to behavioural experiments in cognitive therapy*. Oxford University Press. https://doi.org/10.1093/med:psych/9780198529163.001.0001

Craske, M. G., Treanor, M., Conway, C. C., Zbozinek, T., & Vervliet, B. (2014). Maximizing exposure therapy: An inhibitory learning approach. *Behaviour Research and Therapy, 58*, 10–23. https://doi.org/10.1016/j.brat.2014.04.006

Lally, P., van Jaarsveld, C. H. M., Potts, H. W. W., & Wardle, J. (2010). How are habits formed: Modelling habit formation in the real world. *European Journal of Social Psychology, 40*(6), 998–1009. https://doi.org/10.1002/ejsp.674

Maltz, M. (1960). *Psycho-cybernetics*. Prentice-Hall.

Ougrin, D. (2011). Efficacy of exposure versus cognitive therapy in anxiety disorders: Systematic review and meta-analysis. *BMC Psychiatry, 11*, 200. https://doi.org/10.1186/1471-244X-11-200

Salkovskis, P. M., Clark, D. M., & Hackmann, A. (1991). Treatment of panic attacks using cognitive therapy without exposure or breathing retraining. *Behaviour Research and Therapy, 29*(2), 161–166. https://doi.org/10.1016/0005-7967(91)90044-4

Stott, R. (2007). When head and heart do not agree: A theoretical and clinical analysis of rational-emotional dissociation (RED) in cognitive therapy. *Journal of Cognitive Psychotherapy, 21*(1), 37–50. https://doi.org/10.1891/088983907780493313

Öst, L.-G., Havnen, A., Hansen, B., & Kvale, G. (2015). Cognitive behavioral treatments of obsessive-compulsive disorder: A systematic review and meta-analysis of studies published 1993–2014. *Clinical Psychology Review, 40*, 156–169. https://doi.org/10.1016/j.cpr.2015.06.003

Smith, C. A., & Ellsworth, P. C. (1985). Patterns of cognitive appraisal in emotion. *Journal of Personality and Social Psychology, 48*(4), 813–838. https://doi.org/10.1037/0022-3514.48.4.813

Angeletos Chrysaitis, N., & Seriès, P. (2023). 10 years of Bayesian theories of autism: A comprehensive review. *Neuroscience & Biobehavioral Reviews, 145*, 105022. https://doi.org/10.1016/j.neubiorev.2022.105022

Cannon, J., O'Brien, A. M., Bungert, L., & Sinha, P. (2021). Prediction in autism spectrum disorder: A systematic review of empirical evidence. *Autism Research, 14*(4), 604–630. https://doi.org/10.1002/aur.2482

Cui, K., Lin, X., Gao, R., Jing, S., Luo, F., & Wang, J. (2026). A systematic review and meta-analysis of empirical evidence for the simple Bayesian model of autism. *Neuropsychology Review, 36*(2), 184–195. https://doi.org/10.1007/s11065-025-09672-8

Lawson, R. P., Mathys, C., & Rees, G. (2017). Adults with autism overestimate the volatility of the sensory environment. *Nature Neuroscience, 20*(9), 1293–1299. https://doi.org/10.1038/nn.4615

Plank, I. S., Pior, A., Yurova, A., Nowak, J., Shi, Z., & Falter-Wagner, C. M. (2026). Predicting emotional valence in autism: A preregistered study in the Bayesian Brain framework. *Molecular Autism, 17*(1), 32. https://doi.org/10.1186/s13229-026-00730-3

Popper, K. R. (2002). *The logic of scientific discovery*. Routledge. (Original work published 1935)

Sapey-Triomphe, L.-A., Bouet, R., Mattout, J., Sonié, S., Schmitz, C., & Lecaignard, F. (2025). Systematic review and meta-analysis of mismatch negativity in autism: Insights into predictive mechanisms. *Autism Research, 18*(12), 2431–2450. https://doi.org/10.1002/aur.70131

Schwartz, S., Shinn-Cunningham, B., & Tager-Flusberg, H. (2018). Meta-analysis and systematic review of the literature characterizing auditory mismatch negativity in individuals with autism. *Neuroscience & Biobehavioral Reviews, 87*, 106–117. https://doi.org/10.1016/j.neubiorev.2018.01.008

Beck, A. T. (1967). *Depression: Clinical, experimental, and theoretical aspects*. Harper & Row.

Beck, A. T., Rush, A. J., Shaw, B. F., & Emery, G. (1979). *Cognitive therapy of depression*. Guilford Press.

Quiller-Couch, A. (1916). *On the art of writing: Lectures delivered in the University of Cambridge, 1913–1914*. Cambridge University Press. https://www.gutenberg.org/ebooks/17470

Slime Mold Time Mold. (2025). *The mind in the wheel* [Blog series]. https://slimemoldtimemold.com/2025/02/06/the-mind-in-the-wheel-prologue-everybody-wants-a-rock/