TECHNICAL APPENDIX

The evidence behind Hooked

Claims, source checks, scoring, and disclosures

Nir Eyal, with Ryan Hoover · Evidence cutoff: 13 September 2026

Report ID: TBS-HOOKED-APPENDIX

58 min read

This is the supporting record for the reader review. It retains the detailed analysis rather than requiring every reader to work through it. Use the claim register to find a finding, then follow its sources and qualifications. The common tables standardize navigation; the original audit designs and identifiers remain distinct.

Overall 50% · Scientific Accuracy 50% · Reference Accuracy 50% · Practical Value 50%.

About This Review

This review combines factual reporting with editorial judgments about the identified book and its evidence. Ratings apply our published rubric; they are not percentages of truth or judgments about an author's honesty. Criticism of support does not itself show that a claim is false. Evidence is assessed through 13 September 2026, with source-access limits disclosed. Preparation used AI, and independent human fact-checking or legal clearance is not documented. Read the methodology, book-specific disclosures, and full editorial notice with the review. Corrections and substantive responses are welcome.

Find the supporting evidence

EVIDENCE MAP

Claim register

This common register makes the detailed evidence easier to navigate. Each row identifies a scoped proposition or source relationship, a verdict, and the treatment it warrants. A label applies to that row—not to the whole book or its author. The full wording, sources, qualifications, and original audit identifiers remain in the linked exhibits.

> **Verdicts:** Supported; Partly supported; Overstated; Incorrect; Unresolved; Practical judgment. "Incorrect" requires an identified factual or measurement problem; "unresolved" is an access or verification limit, not a failed claim. "Practical judgment" identifies advice or a safeguard rather than a tested effect. These routing labels add no numerical score. Definitions.

ID	Claim or focal issue	Verdict	Treatment and supporting detail
C01	A formation timetable that varies rather than a universal rule	**Supported**	Retain the bounded finding and its stated outcome. Full assessment and sources
C02	Flossing supports repetition, not a complete product formula	**Partly supported**	Retain the supported core; carry the qualifications into the advice. Full assessment and sources
C03	Habit cues do not have to become invisible emotions	**Overstated**	Narrow the inference to what the study measured. Full assessment and sources
C04	Fogg's model is represented substantially faithfully, but "formula" overpromises	**Partly supported**	Retain the supported core; carry the qualifications into the advice. Full assessment and sources
C05	The scarcity result is real, but conditional	**Partly supported**	Retain the supported core; carry the qualifications into the advice. Full assessment and sources

ID	Claim or focal issue	Verdict	Treatment and supporting detail
C06	The wine effect survives; the account of the procedure does not	**Partly supported**	Retain the supported core; carry the qualifications into the advice. Full assessment and sources
C07	Endowed progress is a reasonably faithful field example	**Supported**	Retain the bounded finding and its stated outcome. Full assessment and sources
C08	Three factual errors in the foundational brain-stimulation story	**Incorrect**	Correct the identified description or measurement; do not discard unrelated valid findings. Full assessment and sources
C09	Anticipatory reward activity does not establish that pleasure never motivates action	**Overstated**	Narrow the inference to what the study measured. Full assessment and sources
C10	Berns measured an fMRI signal, not dopamine release	**Incorrect**	Correct the identified description or measurement; do not discard unrelated valid findings. Full assessment and sources
C11	A title's word "variable" is not evidence of variable-ratio reinforcement	**Overstated**	Narrow the inference to what the study measured. Full assessment and sources
C12	A clinical neurocognitive account is not a typical-user experiment	**Overstated**	Narrow the inference to what the study measured. Full assessment and sources
C13	The original "but you are free" finding was reported too loosely	**Partly supported**	Retain the supported core; carry the qualifications into the advice. Full assessment and sources
C14	Labor can increase valuation; completion is a tested boundary	**Partly supported**	Retain the supported core; carry the qualifications into the advice. Full assessment and sources
C15	Foot-in-the-door evidence is not evidence of habitual product use	**Overstated**	Narrow the inference to what the study measured. Full assessment and sources
C16	Computer reciprocity cannot establish a mandatory order of stages	**Overstated**	Narrow the inference to what the study measured. Full assessment and sources
C17	The denominator changes the ethical conclusion	**Overstated**	Narrow the inference to what the study measured. Full assessment and sources
C18	The self-disclosure source was not a meta-analysis	**Partly supported**	Retain the supported core; carry the qualifications into the advice. Full assessment and sources

01 The verdict in context

The Hook Model is more defensible as a set of design questions than as an established causal recipe for habit formation. Trigger, action, reward, and investment provide a memorable way to inspect a product. That organizational usefulness is distinct from evidence that the complete sequence is necessary, sufficient, or superior to alternatives.

The broad cue-and-repetition foundation is the strongest part. The privileged role assigned to variability, the interpretation of investment, and several neuroscience explanations require substantial qualification. Frequent use is not automatically habitual use; retention may reflect genuine utility, social participation, defaults, or switching costs.

Some cited studies support narrow illustrations well. Others do not measure the process claimed for them. A later qualitative application of the model is not an outcome trial, and a small reward-schedule trial with no significant difference does not establish equivalence.

Scientific assessment; 18-pair audit.

Best use: turn each design question into a falsifiable product experiment, measure user benefit separately from engagement, and preserve control over reminders, stopping, and exit. The ethical test must concern users' outcomes and choices, not only the designer's intentions.

> **Recommendation:** Use as a product-design checklist, not a validated causal model or an ethical clearance test.

Confidence: Moderate confidence in the overall appraisal; high confidence that design examples and brain-imaging associations do not by themselves validate the complete sequence.

Ease of application: Easy to understand; meaningful implementation needs product research and testing.

Best use: Use the framework to ask testable design questions, not to claim a proven causal mechanism.

Rating profile: Overall 50% · Scientific Accuracy 50% · Reference Accuracy 50% · Practical Value 50%. Full rationale.

02 Scope and method

Question: How well do the reviewed book's important scientific claims hold up, how faithfully does it use sources, and what can readers responsibly do with its advice? The intended audience is an interested general reader, with enough detail for researchers and editors to audit the reasoning.

Edition and scope. 2014 Portfolio / Penguin ebook; ISBN 9780698190665. Coverage comprises 18 purposively selected claim–source pairs; all eight substantive chapters. Book citations identify chapters and named sections of the reviewed edition. The source copies have not been authenticated against publisher-controlled masters.

Preparation. The source basis includes the chapter assessments, source records, every registered claim, additional identifiable named studies, and all nine scoring inputs under the stated rubric. This is a targeted critical narrative review, not an exhaustive systematic review or an independent reproduction of datasets. The shared methodology explains the evidence and scoring rules.

Claim selection. Three central propositions summarize the book's organizing argument, its proposed mechanism or intervention, and important practical extensions. Selection is purposive and retrospective, not random or preregistered. All chapters remain covered to

expose peripheral but consequential claims. The three proposition grades are not a sample-based estimate of the truth of the whole book. Narrow existence claims are easier to support than universal superiority claims; the stated proposition must always accompany the grade.

Evidence rules. The analysis distinguishes source identity, what the cited study actually found, the strength of its design, independent corroboration, and transfer to the book's practical claim. A controlled component study does not validate a branded package; an association does not identify an intervention effect; failure to locate evidence is not proof of its absence. Null findings, attrition, selection, selective reporting, measurement limitations, and competing explanations matter where they change an inference. Evidence available at publication and later updates are identified separately.

Ratings. The three categories are Scientific Accuracy, Reference Accuracy and Practical Value. Each uses three explicitly anchored judgments. Category percentages are calculated from the original inputs; the overall rating gives each category equal weight. Final displays use whole-number percentages, with all rounding performed after calculation. Ease of application and consequential cautions are reported separately. Section 09 and the series methodology provide the exact inputs, rules and limits. These are editorial judgments, not probabilities, estimates of the percentage of true content, or validated psychometric measurements.

Reference checks. Targeted source audits are stress tests, not prevalence estimates. The number of discrepancies in a purposive sample cannot estimate the proportion of a book's references that are wrong. The Grit report additionally preserves an explicitly restricted-frame seeded selection, with its missing-source cases and inherited inclusion coding disclosed. It is not the same audit design as the other books.

Relationship to Red Pen Reviews. The series adapts the separation of central scientific claims, reference support and practical consequences from the public Red Pen Reviews process and method, version 2.2. It uses a behavioral-science Practical Value category instead of nutrition-specific Healthfulness. Reference Accuracy is a common three-criterion assessment, not Red Pen Reviews' ten-reference random-sample score. The three category percentages are averaged, but there is no claim of an independent second expert reviewer. These departures mean the scores are not directly interchangeable with official Red Pen Reviews scores. This review is not affiliated with or endorsed by Red Pen Reviews.

Research cutoff: 13 September 2026. Dated source checks are listed in Section 10. Source-access dates are distinct from the papers' publication dates. The search does not guarantee exhaustive database coverage or detection of every notice. This is an evidence date, not a publication date.

Research search record

The passage below describes the documented research. Not every listed search, source read, or assessment was independently repeated; dated retrieval and access limits are listed in Section 10.

Search scope and decision rules

The evidence search was conducted on **8 September 2026**, using public web search and scholarly/source-focused retrieval. It began with the book's named studies and endnotes, then sought later evidence relevant to the core mechanisms and the complete Hook Model. Search terms included the exact titles of the cited papers and combinations equivalent to "Hook Model empirical validation randomized automaticity," "variable reinforcement schedule app randomized trial," and "but you are free meta-analysis re-examination." These describe the actual research directions; they are not a preregistered search protocol.

Priority went to original papers, publisher records, author-hosted manuscripts, primary institutional records and the original study reports reproduced in accessible repositories. Sources were included when they could clarify a consequential claim, provide a relevant comparison, or expose a limitation in the review's own inference. General web criticism and a researcher's reputation were not treated as substitutes for inspecting the actual claim and study.

The model-level search found a directly relevant 2023 comparative case study, foundational and naturalistic habit research, and component-level intervention evidence. It did not yield a convincing direct test of the full four-stage package with the necessary comparison and habit-specific outcomes. Database completeness, unpublished studies and inaccessible records remain limitations. **"Not located" is not "proved nonexistent."**

03 The strongest fair reading

The Hook Model, faithfully reconstructed

The central organizing device is a four-stage loop: **trigger → action → variable reward → investment**. External triggers initially prompt the user. Over repeated cycles, the product becomes associated with internal triggers such as thoughts, feelings or an existing routine. The action should be easy enough to perform with little deliberation. The reward should address the user's need while retaining an element of uncertainty. Investment means putting something into the service that improves subsequent experiences or sets up the next trigger. (B, Introduction.)

The four stages are not the book's entire message. Chapter 1 introduces the **Habit Zone**, a conceptual relationship between frequency and perceived utility. Chapters 2–5 explain the stages. Chapter 6 asks whether designers should use these techniques and offers the **Manipulation Matrix**. Chapter 7 illustrates the framework through YouVersion's Bible App. Chapter 8 proposes **Habit Testing**: identify devoted users, codify their common paths, and modify the product accordingly. (B, ch. 1; chs. 6–8.)

The strongest reasonable reading

On a charitable reading, the model is a checklist for making a recurring useful experience easier to discover, perform and repeat. "Investment" directs attention beyond the immediate conversion toward cumulative value. "Variable reward" asks whether the experience can remain interesting. "Internal trigger" asks whether the product becomes a remembered solution rather than depending indefinitely on paid acquisition. This is an intelligible and potentially useful product language.

The book also supplies important brakes on simplistic application. It rejects one-size-fits-all habit design. It says variable rewards are not a cure for a product nobody wants. It warns against excessive investment demands, privacy violations and deceptive invitations. It makes user testing a necessity rather than treating the diagram as an oracle. These qualifications deserve weight, not a footnote added after a hostile reading. (B, chs. 1–5; ch. 8.)

Where the model becomes scientifically slippery

The same terms often do more than one job. "Habit" can mean automatic responding, frequent engagement, commercial loyalty or difficult-to-break dependence. "Reward" can mean a pleasurable outcome, information, social recognition, task completion or relief. "Investment" includes rationally useful data, transferable or nontransferable skill, psychological commitment and an action that prompts someone else to reply.

These mechanisms can coexist, but combining them under one heading does not show they share a cause. A person can deliberately return to a valuable database without having an automatic checking habit. A saved contact list can make switching expensive without changing automaticity. A notification can cause another visit without demonstrating the promised transition to unprompted engagement. These are logical distinctions that matter when deciding what the book's examples actually demonstrate.

Reading the figures correctly

Figure 1, ch. 1: the Habit Zone boundary has no numerical calibration. The text expressly calls it a guiding theory. It should not be read as an estimated response surface, validated threshold or formula for the trade-off between utility and frequency.

Figure 20, ch. 4: tribe, hunt and self are overlapping reward categories. Eyal introduces them as a proposal. They can support brainstorming without being an exhaustive, independently validated taxonomy.

Figure 28, ch. 5: the origami valuations correspond approximately to the cited IKEA-effect experiment. The graph supports a local valuation result, not a general claim that any added work increases retention.

Figure 36, ch. 6: the Manipulation Matrix sorts designers by their own use and their belief in the product's benefits. It is an introspection aid, not a measured classification of actual user outcomes. [19]

04 Three central scientific claims

These are the three rated propositions, stated at the level assessed—not three verbatim quotations or an exhaustive list of the book's claims.

CLAIM 1 · 75%

Frequent, cue-linked repetition can make specific product actions more automatic.

The underlying cue–repetition principle is reasonably supported. Retention and frequency remain imperfect proxies for automaticity, and no universal usage threshold follows.

CLAIM 2 · 50%

Variable rewards are a privileged driver of repeat engagement and habit formation.

Reward schedules and uncertainty can matter. The evidence does not establish variability as necessary or generally superior across products, rewards, and outcomes.

CLAIM 3 · 25%

Investment and the complete Hook sequence reliably create durable product habits.

The full sequence is not validated by component experiments or retrospective company cases. Investment may create value or switching costs without demonstrating the proposed habit mechanism.

Three propositions, three different evidence levels

The three claims below were selected for their importance to the book's argument, not because they exhaust its content. Their formulations preserve a supported foundation, the model's distinctive reward claim, and the stronger claim about investment and the complete sequence. They overlap conceptually; the summary appraisal is an editorial overview, not a statistical synthesis.

Claim 1 · Frequent, cue-linked repetition can make product-related actions increasingly automatic

75% · Reasonably supported, with important qualifications
Confidence: high in the broad mechanism; moderate in transfer to a particular product or metric.
What the book says. Habits develop through repeated behavior and associations with cues; frequently used products have an advantage in forming them. The Habit Zone adds perceived utility as the other major dimension. The author correctly rejects a universal formation timetable. (B, Introduction; ch. 1; ch. 2.)

What supports it. Research on habits connects repeated responding in recurring contexts with learned cue–response associations. Lally and colleagues studied attempted habit formation in everyday behaviors, rather than merely intentions to change. Judah and colleagues studied flossing linked to an existing toothbrushing routine. These are relevant to the book's basic proposition, even though neither tests a digital-product framework. Oulasvirta and colleagues are more directly relevant: their work examined smartphone checking and included a controlled field experiment involving readily accessible informational rewards. [2,3,4,5,6]

Later synthesis — 2024: Singh and colleagues reviewed 20 health-behavior studies involving 2,601 participants. Habit-strength changes were favorable overall, but 11 studies were rated high risk of bias and only four provided formation-duration estimates. Their pre/post synthesis is not a randomized causal effect of the complete Hook Model and not an exact independent replication of Lally. It strengthens the bounded case for repetition and context while retaining uncertainty about timing, population, measurement, and transfer. The earlier Wood, Quinn, and Kashy diary studies concern recurring behavior in stable contexts; their self-reported frequency and context are not an objective automaticity standard. [35; 44]

What does not follow. Frequency is not the same construct as automaticity. More opportunities to repeat can facilitate learning, but a frequent deliberate action remains possible. Context consistency matters; raw visit counts do not capture it. A product's apparent loyalty advantage can also reflect superior utility, defaults, network participation or switching costs. The book usually does not separate these explanations in its company examples.

The external-to-internal story is also too categorical when treated as a requirement. Established habits can continue to be cued by surroundings or preceding actions. Becoming independent of a company's marketing messages is not the same as becoming independent of external environmental cues. An app icon or the end of a meeting can remain an effective cue without invalidating habit formation. [2]

Why not the highest rating? The underlying principle is stronger than the operational claims built on it. The uncalibrated Habit Zone does not supply a validated threshold. The studies cited do not establish a general dose–response rule for app visits, a universal minimum use frequency, or a necessary emotional trigger. Research uncertainty is not resolved by labeling all frequent users "habitual."

Better formulation. Repetition in recurring contexts can increase the automaticity of a specific action. For a product, measure the action, relevant opportunities and automaticity separately; do not assume retention is proof of the mechanism.

Claim 2 · Variable rewards are a privileged driver of repeat engagement and habit formation

50% · Plausible component, overgeneralized mechanism
Confidence: high that reinforcement schedules and uncertainty can matter; low that variability is necessary or generally superior for durable product habits.
What the book says. The variable-reward phase distinguishes the Hook Model from a simple feedback loop. Predictable loops supposedly do not create desire; variability drives wanting and ongoing novelty sustains attention. Rewards of tribe, hunt and self organize the design examples. (B, Introduction; ch. 4.)

The legitimate foundation. Experimental learning research makes it reasonable to investigate reward schedules. Neuroscience also shows responses related to predictability and uncertainty. Fiorillo and colleagues directly recorded dopamine-neuron responses while manipulating reward probability; the study distinguishes signals related to probability and uncertainty. This is a genuine basis for saying uncertainty can affect reward processing—not a reason to claim every uncertain app outcome disables judgment. [13]

Interpretation exchange — 2005: Niv, Duff, and Dayan proposed that asymmetric temporal-difference prediction errors, averaged across trials, could account for the reported uncertainty-related ramping. Fiorillo, Tobler, and Schultz replied that individual-trial and task-specific evidence favored their uncertainty interpretation. Both primary abstracts and analytical passages were examined. This is a computational interpretation dispute, not an empirical failed replication, retraction, or demonstration that dopamine was not recorded. Original spike data were not independently reproduced, and neither interpretation establishes the Hook Model's effect on product habits. [36; 37]

The important contrast is between a **candidate mechanism** and a **general product prescription**. The book slides among uncertainty about whether a reward arrives, how much arrives, when it arrives, which content appears, whether the user succeeds at a task, and how others respond. A manipulation of one cannot automatically validate the others. A person might return for information that changes because the world changes, rather than because variability itself has an independent reinforcing effect.

The strongest citation problem. Berns and colleagues measured fMRI responses to predictable and unpredictable delivery of juice and water. Aharon and colleagues examined the rewarding value of different face categories using behavior and fMRI. Neither directly

measured dopamine release. The latter's title uses "variable reward value" to describe differences between stimuli, not to report a variable-ratio reinforcement experiment. Their findings cannot be translated into observed dopamine spikes during feed browsing. [8,9]

What direct product evidence adds. A six-week, three-arm randomized study of 61 participants using a health-related app compared fixed and variable reward arrangements. It did not detect superior engagement under the variable schedules; one arrangement had lower costs. Because the study was small and the conditions differed in more than an abstract "variability" parameter, this is neither a decisive disproof nor evidence of equivalence. It does show why the broader design claim needs testing rather than assumption. [16]

A useful internal counterpoint. The book's car-wash loyalty example rewards a known number of purchases, yet produces greater completion when participants receive an apparent head start. That study is evidence about goal pursuit under a predictable contingency. It does not establish automatic habit, but it illustrates why repeat behavior should not be attributed exclusively to unpredictable rewards. [23]

Credit the boundary condition the book does supply. Eyal explicitly warns that rewards must fit the user's need and that gamification is not a universal fix. That prevents the most extreme reading—"any randomness will hook anyone"—from being a fair summary. The remaining problem is the stronger language about necessity, craving and general neurological effects. (B, ch. 4.)

Better formulation. Relevant uncertainty can sometimes increase attention or persistence. Whether it improves a product should be tested against a predictable, useful alternative at comparable expected value and burden. Neither excitement nor extra sessions alone establishes a lasting habit or a user benefit.

Claim 3 · Investment and the prescribed four-stage sequence produce durable, unprompted engagement

25% · An attractive design hypothesis, not established as a general causal package
Confidence: high in the gap between the supplied evidence and the strongest claim; uncertain about average incremental effectiveness in practice.
What the book says. Investment is described as a final critical phase and a requirement for repeat use. Users should invest after receiving the variable reward. Their accumulated work, data, content, reputation or relationships make the product more valuable and help load future triggers. Repeated cycles are said to yield unprompted engagement. (B, ch. 4; ch. 5.)

What is genuinely promising. A saved collection or preference profile can improve future service quality. A useful reply can elicit another reply. Skill can reduce future interaction costs. These are coherent reasons for a product to improve with use. They do not depend on an exotic psychological explanation, and a product team should consider them even without accepting the complete Hook Model.

The empirical examples are narrower. The IKEA-effect experiments show higher valuation of successfully self-made objects under certain conditions. Computer-reciprocity research shows increased helpful behavior after assistance from a computer. Neither establishes that a

separate post-reward investment step is necessary for automaticity, nor that the prescribed sequence outperforms other onboarding or engagement designs. [19,21]

Several distinct explanations are bundled together. Users may remain because they now receive better recommendations; because moving their data is costly; because they infer value from past effort; because others reply; or because a cue automatically activates a response. These imply different product decisions. For example, transferable data preserves value without necessarily creating lock-in. Labor that adds no value could make a product worse even when successful construction increased valuation in a laboratory task.

What the model-level search found. A 2023 peer-reviewed article explicitly analyzes Hook Model applications in Uber and Instagram. Its method is qualitative feature and use-case decomposition, not randomized comparison. It is evidence that the framework can describe products, not that the four stages caused their success. This review did not locate a study establishing the complete sequence's incremental causal benefit with habit-specific measures and a suitable comparison condition. That is a bounded search conclusion, not a universal assertion that no relevant study exists. [14]

Why evidence of some digital habits does not settle the question. Gera and colleagues used a smartphone task to study habit induction and outcome devaluation. It is valuable evidence about human action control outside a conventional laboratory session. It does not independently isolate the contribution of the Hook Model's stages or their ordering. A successful experiment using rewards is not automatically a test of every framework that contains a reward box. [15]

What would justify a stronger score? A prospective experiment could compare the full package with a well-designed simpler alternative, hold reward value and exposure reasonably comparable, establish actual receipt of the manipulations, and follow behavior after reminders or incentives change. Component tests would then be needed to distinguish investment's practical value from the claim that its position after reward is essential. Replication in more than one product category would matter.

Better formulation. Useful user contributions can improve future value or create legitimate reasons to return. The complete sequence is a hypothesis-generating framework. Its necessity, optimal ordering and incremental effectiveness require direct evidence.

The conceptual issue beneath all three claims

An app can create **retention without habit**, **habit without benefit**, or **benefit without more engagement**. Those are not semantic technicalities. A successful task tool may reduce sessions by completing work faster. A dating service may serve a user well when the user leaves after finding a partner. Conversely, a system may increase checking without improving the outcome that originally brought the user there. The appropriate success metric depends on the product's job; a rising engagement curve cannot determine that by itself.

05 Chapter-by-chapter assessment

What to keep, qualify and discard

Introduction · A memorable map, an inflated promise

Scope: Introduction. The introduction makes the problem vivid and gives the reader an unusually memorable structure. Its real strength is linking a recurring user problem to the design of repeated experiences. Its rhetorical weakness is the suggestion that product makers have engineered users' actions with something approaching predictable control. The fact that a successful product can be redescribed using four labels is weaker evidence than a demonstrated ability to predict which new products will succeed.

The explicit practical-tool disclaimer in the Introduction is important. It justifies evaluating the model as a heuristic, but does not immunize empirical claims in the surrounding explanation. **Keep the four questions as prompts; reject the implication that a neat diagram settles the causal science.**

Chapter 1 · The Habit Zone

Scope: ch. 1. The distinction between businesses that need recurrent use and those that do not is one of the book's most valuable strategic qualifications. Its discussion of accumulated value and switching also identifies plausible commercial advantages. The unscaled Habit Zone graph should remain what the author says it is: a guiding idea, not a calculated boundary.

The chapter sometimes shifts from habit to loyalty, from switching costs to automaticity, and from clinical or weight-related changes to a general account of behavioral persistence. Its "last in, first out" analogy is too sweeping to adopt as a law. The 13 September source check identifies the intended extinction, relapse-curve, and weight-maintenance references, but none establishes a universal law for every patient or product habit. The proposed definition involving discomfort when an action is omitted is also narrower than the automaticity definition with which the book begins. Keep recurrence and context; do not make discomfort or compulsive persistence the test of whether a habit exists. [2; 5; 17; 39; 40; 42]

Clinical source scope: Bouton's extinction review concerns context-dependent learning, not a universal human medical rule. Kirshenbaum and colleagues synthesized 53 relapse curves from 20 alcohol/tobacco studies, using selected reporting intervals and digitized graphs; those aggregate first-event patterns do not identify one mechanism in every individual. Zywiak and colleagues' later analysis of 550 participants favorably reproduced the broad first-drink curve while showing recovery and drinking transitions hidden by that summary. Jeffery and colleagues reviewed the difficulty of maintaining weight loss, not a trial showing that everyone regains everything. A later BMJ synthesis of 249 studies finds regain patterns that depend on time, comparator, and program features; long follow-up remains sparse, and across-study feature comparisons are not randomized mechanism tests. These sources justify bounded claims, not deterministic predictions about an individual. [39; 40; 41; 42; 43]

Chapter 2 · Trigger

Scope: ch. 2. Separating acquisition prompts from cues embedded in ongoing experience is operationally useful. So is observing what happens immediately before a target action. The book's emphasis on actual behavior rather than only stated preferences can improve the questions a team asks.

The risk lies in turning a plausible story about hidden feelings into a discovered causal truth. The "5 Whys" exercise is a hypothesis generator, especially when its answers are imagined for a persona. Eyal acknowledges that different narratives could produce different answers; that caveat should be central to its use.

An additional factual problem appears in the depression example. The cited research describes **a month-long Internet-usage study and a CES-D symptom survey**, not a year-long study identifying depression through visits to university health services. The association does not demonstrate that technology use relieved symptoms. Eyal does call the relief explanation a hypothesis, which is appropriate, but the method description should be corrected. **Observe cues before inventing emotions; test the inferred need.** [27]

Chapter 3 · Action

Scope: ch. 3. This is among the most practically useful chapters. It makes designers inspect time, money, effort, complexity and context rather than blaming users for lacking motivation. The discussion of privacy concerns around social login recognizes that "fewer clicks" is not always subjectively easier.

The overreach is the universal priority given to ability. When an action is already easy but the offer is irrelevant or untrusted, another removed step may accomplish little. This is a logical boundary on the prescription, not evidence that friction never matters. The historical Twitter redesigns do not isolate simplicity from changing public awareness, product quality or distribution; the text itself recognizes part of this confounding.

The heuristics section is best treated as a source of experiments. Correct the wine-study procedure, preserve the car-wash result's scale, and do not infer an effect on retention from an effect on immediate valuation. **Keep bottleneck diagnosis; replace "always simplify first" with "test the currently binding constraint."** [7,22,23,24]

Chapter 4 · Variable Reward

Scope: ch. 4. The tribe–hunt–self vocabulary can help a team notice different forms of value. The warnings against irrelevant badges, points and monetary incentives are a strength. The emphasis on autonomy and the risks of forced exposure of browsing information is also substantive.

This is nevertheless the chapter with the greatest scientific repair burden. Its neurobiological account contains multiple specific citation mismatches. Its examples often do not distinguish variability's contribution from the underlying reward, the availability of new information, community participation or the challenge of skill development. Its finite-versus-infinite variability distinction can support content planning, but does not establish a universal law that predictable experiences lose all appeal.

Keep reward relevance and user choice. Treat novelty as one possible design feature, not the universal ingredient that converts feedback into habit. The later evidence on the compliance phrase should replace the assurance of a reliable doubling. [8,9,10,11,12,13,16,25,26]

Chapter 5 · Investment

Scope: ch. 5. This chapter's strongest contribution is the idea that the next interaction should benefit from the current one. Content, data, reputation and skill are concrete categories a team can inspect. Staging contributions in small, manageable steps is more sensible than demanding a large commitment before users understand the value.

Its scientific weaknesses are the claim of necessity and the merger of several mechanisms. Rational future benefits, costly switching, reciprocity, effort-based valuation, and automatic responding should not be collapsed into one process. The original IKEA experiments supply a useful completion boundary, but later conceptual replication does not support presenting completion as a universal necessity. The computer-reciprocity example supports more helpful behavior, not proof that reward–investment ordering is optimal.

Keep investment that genuinely improves the service or helps the user return for a real purpose. Reject busywork justified solely by "labor leads to love." Offer contributions proportionate to the benefit, and test whether they improve outcomes rather than just predicting them among already motivated users. [19,20,21]

Chapter 6 · What Are You Going to Do with This?

Scope: ch. 6. The book does not ignore ethics. It distinguishes helpful habits from destructive dependency, explicitly raises manipulation, rejects exploitation, and says companies should identify and assist users forming harmful attachments. These are significant merits.

The proposed test remains too dependent on the designer's perspective. Personally using a product and believing it improves life do not establish that other users benefit. Conversely, designing for a need one does not personally have does not imply a lack of empathy or inferior ethical standing. The categorical association between designer identity and likely business success is not validated by the matrix.

The 1% passage is especially problematic because it uses a population estimate to reassure about product-user risk. **Keep the obligation to protect users; replace self-certification with measurable outcomes, meaningful consent and practical exit options.** A moral intuition is a starting point for governance, not an audit result. [28]

Chapter 7 · Case Study: The Bible App

Scope: ch. 7. This chapter is useful precisely as a case study. Portability, shorter reading tasks, optional audio, reminders, saved annotations and existing social distribution provide plausible design lessons. The book also acknowledges first-mover timing and the importance of congregations and religious leaders, which are alternative or complementary contributors to growth.

There is no counterfactual showing what would have happened without the complete Hook sequence. Many users reportedly never register or follow a reading plan; that is a reason not to assume the detailed loop is required for all engagement. The self-disclosure paper is mislabeled, and a streak is not the same manipulation as an artificially endowed head start. [23,29]

One further numerical statement deserves caution, not an invented correction. Chapter 7 says 66,000 people open the app every second. Taken literally, that implies **5.7024 billion opens per day**. The original interview data and intended unit were not independently recovered. A concurrency-versus-rate confusion is possible, but is **not established**. This report therefore flags the claim as unverified and does not substitute a guessed figure. **Keep the design observations; do not treat the success narrative or this statistic as model validation.**

Chapter 8 · Habit Testing and New Opportunities

Scope: ch. 8. The identify–codify–modify cycle is a useful call to observe real users and iterate. The author explicitly says there is no way to know which hypotheses work without testing. That is the right orientation.

The measurement problem is that frequency-based devotion does not establish a habit, and a common path among loyal users does not show that forcing everyone down that path will increase loyalty. "Followed thirty people" could partly measure preexisting interest, social opportunity or a willingness to invest. A randomized test is needed to estimate the effect of encouraging the path.

The 5% threshold is expressly a rule of thumb, not presented as a research-derived constant. It should therefore be treated as a heuristic rather than counted as a fabricated scientific statistic. **Keep cohort analysis and iteration; add held-out prediction checks, randomized intervention tests and a habit-specific measurement plan.** [17]

06 Claim, number, and source audit

Eighteen checks against the cited evidence

This audit examines a claim together with its attached source, not just whether a bibliography entry exists. Judgments concern the scoped account below. **Selection was targeted**: these sources anchor the major mechanisms or present consequential reporting questions. The mix includes sound citations, overextensions and concrete errors. Later studies are discussed separately so that a 2014 author is not blamed for failing to cite future evidence.

ID	Source / focal issue	Book section	Source relationship
A01	Lally: variable formation times	ch. 1	Faithful at the stated scope
Legacy targeted citation-fidelity subtotal — excluded from current scores (Section 09)			44 / 72

ID	Source / focal issue	Book section	Source relationship
A02	Judah: repetition and attitude	ch. 1; ch. 5	Mostly faithful; qualification needed
A03	Wood–Neal: cues versus internal-trigger necessity	Introduction; ch. 2	Partial or overextended support
A04	Fogg: a conceptual model, not a calibrated formula	ch. 3	Mostly faithful; qualification needed
A05	Worchel: scarcity and valuation	ch. 3	Faithful at the stated scope
A06	Plassmann: wine experiment design	ch. 3	Mostly faithful; qualification needed
A07	Nunes–Drèze: endowed progress	ch. 3	Faithful at the stated scope
A08	Olds–Milner: date, species, stimulation site	ch. 4	Substantial support problem
A09	Knutson: anticipatory activity and risk taking	ch. 4	Partial or overextended support
A10	Berns: BOLD is not a dopamine measurement	ch. 4	Substantial support problem
A11	Aharon: reward value is not reward scheduling	ch. 4	Substantial support problem
A12	Brevers–Noël: clinical theory extrapolated	Introduction	Substantial support problem
A13	Carpenter: odds versus probability; later uncertainty	ch. 4	Mostly faithful; qualification needed
A14	Norton: completion condition and auction procedure	ch. 5	Mostly faithful; qualification needed
A15	Freedman–Fraser: compliance versus product habit	ch. 5	Mostly faithful; qualification needed
A16	Fogg–Nass: reciprocity versus sequence validation	ch. 5	Partial or overextended support
A17	Stewart / AGA: the “1 percent” denominator	ch. 6	Substantial support problem
A18	Tamir–Mitchell: original experiments, not meta-analysis	ch. 7	Mostly faithful; qualification needed
Legacy targeted citation-fidelity subtotal — excluded from current scores (Section 09)			44 / 72

A01 · A formation timetable that varies rather than a universal rule

Book and source: ch. 1; chapter 1, note 23; Lally et al. [3]

The book’s main takeaway—that different behaviors and people have different formation trajectories—is a fair use of this study. An important precision issue is that the widely cited long endpoint was **model-estimated**, not observed over a 254-day follow-up. Participants were followed over 12 weeks, and model fitting was possible for subsets of the sample. This makes the paper a poor basis for a universal deadline or a guaranteed number of repetitions.

Eyal does not impose such a deadline. The source also reports that one missed opportunity did not materially disrupt the modeled formation process. **Faithful at the stated scope for the scoped claim of variable formation times; not a certification of the Habit Zone curve.**

A02 · Flossing supports repetition, not a complete product formula

Book and source: chapters 1 and 5; chapter 1, note 16; Judah et al. [4]

The exploratory study involved 50 participants, a motivational intervention and instructions linking flossing to toothbrushing. Its results connect repetition, attitudes and prior behavior with reported habit development. This is relevant support for the importance of performing a behavior repeatedly and associating it with an existing routine. The book goes further when it presents attitude change as a general movement along its utility axis. Statistical predictors in this setting do not identify the causal contribution of every product-design mechanism that might alter perceived value. The published abstract supports the main qualitative assessment here; this review does not claim an independent reanalysis of the regression or the book's exact ranking of predictors. **Mostly faithful; qualification needed.**

A03 · Habit cues do not have to become invisible emotions

Book and source: Introduction; ch. 2; introduction, note 8; Wood and Neal. [2]

The cited theory supports learned associations between responses and recurring features of their performance contexts. Those features include physical surroundings and preceding actions. That supports the importance of cues, but not a universal transition from visible environmental triggers to private emotional ones. The book partly mitigates this by including existing routines among internal triggers. The remaining problem is conceptual: a learned association is internal to the person, but the cue activating it can remain external. A designer cannot infer the cue's location from the fact that the association was learned. **Partial or overextended support for the stronger transition claim, not for cue-based habit learning itself.**

A04 · Fogg's model is represented substantially faithfully, but "formula" overpromises

Book and source: ch. 3; chapter 3, notes 1 and 6; Fogg's model. [7]

The motivation–ability–trigger structure is a fair presentation of Fogg's conceptual proposal. Fogg's 2009 paper explicitly says its diagram has no units and is conceptual rather than a representation of precise values. The expression **B = MAT** should therefore not be read as a numerical multiplication model with independently estimated parameters. The further instruction to "always" begin by improving ability is not established as a universally optimal investment rule by that conceptual structure. The book deserves credit for recognizing context and privacy barriers, but its design priority is better read as a default heuristic than a demonstrated law. **Mostly faithful; qualification needed.**

A05 · The scarcity result is real, but conditional

Book and source: ch. 3; chapter 3, note 8; Worchel et al. [24]

The source reports higher desirability ratings for scarce cookies and differences depending on changes in availability and the explanation for scarcity. The broad direction described in the book matches the published account. However, this is a study of valuation under specific manipulations—not evidence that an inventory warning on any website increases purchases, improves retention or forms a habit. Scarcity due to demand and scarcity due to an accident are not equivalent conditions even within the original research. **Faithful at the stated scope for the scoped valuation finding.** The broader product application still needs testing, but that is not a reason to discount this substantially faithful account of the original result. Assessment is based on the original article's detailed abstract, not an independent verification of its raw data.

Broader synthesis — 2023: Ladeira and colleagues' product-scarcity meta-analysis distinguishes scarcity types and context-dependent consumer outcomes. The publisher abstract was inspected, not the full quantitative analysis. It is later synthesis relevant to scope, not an exact cookie-experiment replication, a retraction, or evidence that a scarcity message produces automaticity or durable app retention. The original narrow source-fidelity assessment is retained. [34]

A06 · The wine effect survives; the account of the procedure does not

Book and source: ch. 3; chapter 3, note 10; Plassmann et al. [22]

The actual experiment used **three wines presented under five price labels**, not one identical wine at every tasting. Two wines were each presented at both a lower and a higher stated price; a third served as a separate condition. Presentation was randomized, rather than a simple ascent from the cheapest sample to the most expensive. Twenty participants contributed to the analysis. The reported result—price information affected experienced pleasantness ratings and related brain activity—remains directionally consistent with the book's lesson. These are nonetheless substantive procedural inaccuracies, not mere typographical errors. The imaging supplies a neural correlate, not a second omniscient judgment of what someone truly felt. **Mostly faithful; qualification needed.**

Favorable follow-up — 2017: Schmidt and colleagues tested price cues and wine pleasantness in 30 participants and again found higher pleasantness ratings with higher price information. This is new participant evidence from the same research program, not a fully independent replication. Low-price decreases contributed importantly to the price-informed versus blind comparison. The study supports the narrow price/experience effect, not the book's mistaken account of the original procedure, a uniquely established neural mechanism, or product retention. [31]

A07 · Endowed progress is a reasonably faithful field example

Book and source: ch. 3; chapter 3, note 11; Nunes and Drèze. [23]

Both groups needed eight further car washes, but one card displayed two prior stamps toward a ten-stamp target. The source reports higher redemption with the apparent head start: **34% versus 19%**, using its rounded percentages. That is a 15-percentage-point difference and approximately a 79% relative increase. The book's 82% should **not** be labeled a demonstrated

error solely from those rounded figures; rounding or underlying counts can affect the ratio. "About 80% higher completion in this experiment" is safer than implying an exact portable effect. The study tests completion of a reward program, not automaticity or continued use after the prize. **Faithful at the stated scope for the core experimental illustration.**

A08 · Three factual errors in the foundational brain-stimulation story

Book and source: ch. 4; chapter 4, note 1; Olds and Milner. [11]

The narrative places the work in the 1940s, describes mice, and identifies the nucleus accumbens as the stimulation site. The cited paper is from **1954**, concerns **rats**, and identifies the **septal area and other rat-brain regions**. Those corrections are established even by the title and bibliographic record of the paper—and are visible in the book's own note 1 to chapter 4. This review does not deny the fundamental finding that electrical stimulation can reinforce behavior. It rejects the historical and anatomical account given for this particular source. **Substantial support problem.** The title and record are sufficient for these limited corrections; they do not verify every dramatic deprivation anecdote in the narrative.

A09 · Anticipatory reward activity does not establish that pleasure never motivates action

Book and source: ch. 4; chapter 4, note 2; Knutson et al. [10]

The cited 2008 experiment involved 15 heterosexual men and examined whether cues predicting rewarding images influenced subsequent financial risk taking, with nucleus accumbens activity partially mediating the relationship. It supplies evidence about anticipatory processing and choice in that task. It does not establish that this brain region never responds when rewards are received, or that all reward-seeking is exclusively an attempt to relieve craving rather than to obtain enjoyment. The book's general direction—anticipation matters—is defensible; the categorical inference is not. **Partial or overextended support.** A narrower explanation would distinguish anticipatory activity, the experience of an outcome and the reasons a person acts.

A10 · Berns measured an fMRI signal, not dopamine release

Book and source: ch. 4; chapter 4, note 6; Berns et al. [8]

The experimental manipulation was relevant: predictable versus unpredictable sequences of juice and water. The measured outcome was **blood-oxygen-level-dependent activity**, or BOLD, in brain imaging. The paper itself cautions that the link between dopaminergic transmission and BOLD was uncertain and that the study lacked an independent dopamine measure. It is therefore inaccurate to describe this as researchers observing dopamine levels rise in the nucleus accumbens. One can reasonably discuss a relationship to reward circuitry without claiming a neurotransmitter was measured. **Substantial support problem for the dopamine-measurement claim.** This is not a finding that unpredictability had no effect on brain activity.

A11 · A title's word "variable" is not evidence of variable-ratio reinforcement

Book and source: ch. 4; chapter 4, note 7; Aharon et al. [9]

The paper compared the reward value of face categories and separated attractiveness ratings from effort to view images. Its imaging component measured fMRI activity, not dopamine concentrations. Moreover, the title's "variable reward value" refers to differences in value across stimuli, not a demonstration that randomly withholding rewards builds habits. This makes the source a reasonable example of dissociable aesthetic and motivational responses, but a poor citation for the precise claim attached to it in the book. **Substantial support problem.** The error is in the translation from the experiment to a mechanism, not in acknowledging that attractive images can be rewarding.

A12 · A clinical neurocognitive account is not a typical-user experiment

Book and source: Introduction; introduction, note 12; Brevers and Noël. [12]

This paper reviews pathological gambling and a theoretical imbalance among systems involved in impulses, reflection and craving. It is relevant to hypotheses about problematic gambling. It does not directly manipulate ordinary app-content variability and establish that this suppresses the brain systems responsible for judgment and reason in the general population. The book's sentence presents a broad causal effect more confidently than this clinical review warrants. **Substantial support problem.** The correction is to narrow the population, evidence type and mechanism claim, not to imply that product design can never influence attention or self-control.

A13 · The original "but you are free" finding was reported too loosely

Book and source: ch. 4; chapter 4, note 24; Carpenter. [25]

Carpenter's meta-analysis included 42 independent effect estimates and 22,333 participants. It reported a weighted odds ratio of **2.03**, not a universal doubling of the probability of agreement. Those differ: at a baseline probability of 50%, an odds ratio of 2 would imply approximately 67%, not 100%. The original paper itself uses loose "twice as likely" language, so the book inherits rather than invents this ambiguity. Eyal correctly acknowledges the absence of direct product testing. **Mostly faithful; qualification needed for the historical citation.**

Later evidence, not a retrospective citation penalty: Fillon et al.'s preregistered re-examination reported a positive pooled effect but a much smaller, uncertain estimate in seven low-risk-of-bias studies: **g = 0.11, 95% CI −0.18 to 0.40**. That weakens a confident promise of efficacy; it does not prove an exactly zero effect. [26]

A14 · Labor can increase valuation; completion is a tested boundary

Book and source: ch. 5; chapter 5, note 4; Norton, Mochon and Ariely. [19]

The origami comparison is approximately right: creators' mean valuation was about $0.23, versus $0.05 from others and $0.27 for expert creations. The consequential omission is that the researchers explicitly tested **successful completion**: unfinished or dismantled creations did not show the same advantage. That limits the slogan that more effort simply means more love. The book also describes winning participants as paying their own bid; the reported valuation procedure used a randomly drawn price conditional on the bid meeting it. **Mostly faithful; qualification needed:** the core result is recognizable, but the missing boundary condition matters to product advice. Citing the 2011 working paper is not an error merely because the final journal publication appeared in 2012.

Favorable independent conceptual replication — 2017: Sarstedt, Neubert, and Barth used loom-band objects and found a self-made valuation advantage and greater psychological ownership. The proposed pride mediation was not supported. In their empowerment/incompletion comparison, the completion difference was not statistically significant; dismantling was associated with lower valuation. A nonsignificant completion contrast is not equivalence and does not demonstrate that successful completion never matters. Different objects and procedures also prevent treating the report as an exact replication of every original condition. Retain the original completion result as a tested boundary, not a universal necessary condition. The published paper was inspected through primary text excerpts in the author's university-hosted thesis; raw data were not reanalyzed. [30]

A15 · Foot-in-the-door evidence is not evidence of habitual product use

Book and source: ch. 5; chapter 5, note 5; Freedman and Fraser. [20]

The original experiments support the proposition that agreeing to a small request can increase later compliance with a larger one under particular conditions. The book's safe-driving-sign example captures the central comparison. Its application to small product investments is a hypothesis about transfer, not something the field experiment itself tested. The experiment does not uniquely identify a desire for consistency as the mechanism. Neither a second act of compliance nor consistency with a prior public commitment establishes automaticity or long-term retention. The result is useful as a possible mechanism to investigate, not an expected conversion multiplier for onboarding. **Mostly faithful; qualification needed.** The review does not treat one historical experiment as a comprehensive contemporary estimate of foot-in-the-door effects.

Later perspective: Dillard, Hunter, and Burgoon's synthesis describes a small pooled foot-in-the-door effect (approximately $r = 0.17$). Burger's multiple-process review documents condition-dependent success, failure, and competing mechanisms. Their primary abstracts and relevant mechanism passages were inspected; the pooled datasets were not reproduced. Neither source is an exact replication of the safe-driving-sign procedure or evidence that a small request necessarily creates an automatic product habit. [32; 33]

A16 · Computer reciprocity cannot establish a mandatory order of stages

Book and source: ch. 5; chapter 5, note 7; Fogg and Nass. [21]

The full official ACM report, inspected on 13 September 2026, describes 76 participants randomly assigned in a balanced 2 × 2 design: high versus mediocre help from a computer, followed by helping the same versus a different but identical computer. In the high-help comparison, participants completed 11.0 versus 5.9 color comparisons—nearly twice the quantity of help. The time means were 104.4 versus 60.3 seconds; the reported number-correct means were 4.47 versus 3.19, not nearly twice the accuracy. This is original-report verification, not independent replication or execution of the raw analyses. The effect supports a narrow reciprocity illustration but does not compare investment-before-reward with investment-after-reward in a digital habit loop. Nor does it establish reciprocity and accumulated useful data as one mechanism. Partial or overextended support for the stage-order claim; the source assessment is based on the full report. [21]

A17 · The denominator changes the ethical conclusion

Book and source: ch. 6; chapter 6, note 9; Stewart for the American Gaming Association. [28]

The recovered text of this industry white paper gives its approximately 1% prevalence figure for the **adult population**, not for all slot-machine users and certainly not for users of any habit-forming technology. The book moves toward the broader reassurance that all but an "addicted 1 percent" can bear ultimate responsibility for their use. That denominator and cross-product generalization are not supported by the cited figure. **Substantial support problem.** This does not supply a replacement prevalence estimate. The originally cited hosting link was not recovered as an accessible original-host document; the denominator check uses a reproduced copy of the primary white paper, a retrieval limitation that remains material.

A18 · The self-disclosure source was not a meta-analysis

Book and source: ch. 7; chapter 7, note 5; Tamir and Mitchell. [29]

The paper is a set of **original experiments**, not a meta-analysis. It reports evidence connecting self-disclosure with reward-related activity and willingness to forgo money for opportunities to disclose. That supports the possibility that disclosure is rewarding. It does not show that Bible-verse sharing is driven by humblebragging, nor validate the complete explanation offered for the Bible App's growth. The book signals some uncertainty about those motives, which should be preserved. **Mostly faithful; qualification needed:** the main reported finding has support, but the study type is wrong and the application remains indirect.

What this audit establishes

The problem is not an empty bibliography: many relevant sources are real, and several illustrations are reasonably faithful. The recurring failure is **a change in evidentiary level**—from BOLD to dopamine, from clinical theory to ordinary users, from valuation to automaticity, from description to causation, or from population prevalence to product-user risk. Correcting those shifts leaves useful design ideas intact while removing unwarranted confidence.

07 Practical use, limitations, and safeguards

Practical Value · 50%

The intended user is a product builder. Commercial retention can be a legitimate objective, but Practical Value also considers the people affected by the design. The score does not equate high engagement with addiction or assume that a frequently used product is harmful.

Practical Value is 50%: there are credible components and useful design questions, but package-level effectiveness, durable transfer, and sufficient governance of engagement-versus-benefit trade-offs remain unresolved. Criterion-by-criterion rationale.

Ease of application — not included in the score: Easy to understand; meaningful implementation needs product research and testing.

The implementation suggestions that follow are the reviewer's synthesis, not evidence that the book already supplies every safeguard or that this review has validated a replacement program. The score evaluates the book itself, not the reviewer's improved version of its advice.

Good intentions are not an outcome measure

The ethical critique is not that all persuasion is wrong, or that habitual use is necessarily harmful. It is that **a company's engagement objective and a user's objective can diverge**, and the book does not provide enough machinery for identifying and governing that divergence.

Give the book credit for its actual position

Eyal says products should help people do things they already want to do. He distinguishes habits from destructive dependencies, condemns deceptive invitations, addresses privacy and reactance, and assigns companies responsibility for detecting and helping users with unhealthy attachments. Those passages are inconsistent with the caricature that the book simply tells designers to addict people without limits. (B, Introduction; ch. 1; ch. 2; ch. 4; ch. 6.)

The problem is the proposed sufficient condition for a clear conscience: the designer uses the product, believes it improves lives, and has procedures for the addicted minority. That framework leaves the definition of benefit largely with an interested party. It also concentrates attention on severe dependency, leaving ordinary disappointment, distraction, overspending or loss of control outside its central test. Those outcomes need not meet a diagnostic threshold to matter to a user. This is a normative judgment about what should count, not a claim that every product causes those harms.

Three distinctions a responsible product review needs

User-endorsed value versus revealed behavior. Clicking is evidence that a person clicked. It is not, by itself, evidence that the person would endorse the experience after considering its time, money and opportunity costs. Behavior and reflective preferences can both inform

evaluation; neither should be replaced by the other without an explicit rationale.

Supportive persistence versus unwanted persistence. Helping someone return to a chosen learning goal differs from making it difficult to stop a low-value session. The same reminder can have different effects depending on timing, expectations and control. The relevant question is not simply whether it increased engagement, but whether it helped this user pursue an endorsed goal.

Beneficial investment versus manufactured exit cost. Saving a user's work is useful. Preventing its export is a separate design decision. A personalized service can create real value while still giving users meaningful deletion, portability or reset options. Treating every obstacle to switching as evidence of a better product hides this distinction.

A stronger governance standard

The following are **reviewer-recommended design requirements**, not empirically proven guarantees of safety: define the user benefit independently of usage; measure burdens as well as gains; make reminders and social exposure controllable; make stopping and restarting straightforward; and set product-specific thresholds that trigger review when retention rises while user outcomes worsen.

An effective review should also examine distributions, not only averages. An average gain can conceal a subgroup that experiences persistent regret or disproportionately high costs. Such signals are not a diagnosis, and high usage alone is not a diagnostic label. They are reasons to investigate with users rather than an excuse to classify them as a fixed "addicted 1 percent."

The red line

A design should not be justified by arguing that it bypasses judgment while simultaneously assuming that subsequent use proves informed user benefit. Those two claims do not fit together. **Build repeat value first. Preserve the person's ability to choose, stop, leave and return.** Where the goals conflict, the report's ethical recommendation is to prioritize the user-endorsed outcome over the company's raw engagement metric.

A stronger implementation protocol

This section is a proposed operating approach derived from the review. It is not a validated replacement theory. Its purpose is to preserve the useful questions in *Hooked* while making claims falsifiable and reducing the risk of optimizing the wrong outcome.

1 · Define the job, action, opportunity and benefit separately

Specify what the user is trying to accomplish, the smallest behavior that plausibly helps, when there is an actual opportunity to perform it, and how success will be recognized outside the app. "Become a daily active user" is a company metric, not a user job.

For a learning tool, useful outcomes might involve completion and retention of material. For a recommendation service, they might involve finding a suitable option with less search effort. For a workflow tool, they might involve finishing work accurately and on time. Those

examples illustrate possible definitions; the correct endpoint must come from the particular product and user research.

Document the opportunity denominator. Ten helpful uses in ten relevant situations differs from ten uses generated by a hundred interruptions. Likewise, a weekly task should not fail a daily-engagement test simply because its natural cadence is weekly.

2 · Diagnose the constraint before choosing a technique

Observe representative users attempting the task. Determine whether they fail to notice an opportunity, cannot perform the action, see insufficient value, distrust the service, or encounter a competing priority. The diagnosis should guide whether to change the cue, friction, outcome, proposition or context.

Use interviews to understand experience and observation to check behavior. Neither a five-whys narrative nor a heatmap establishes an unobserved motive by itself. Write the inferred mechanism as a hypothesis: "Users abandon this step because the benefit is unclear," not "Their true emotional trigger is fear." Include at least one plausible alternative explanation and a result that would change the team's view.

3 · Test the useful baseline before adding uncertainty

Establish that the service can reliably deliver its core benefit. Then compare any variable-reward feature against a predictable alternative. Keep expected reward value, required effort, opportunity and notification exposure as comparable as the design permits. Otherwise a purported variability effect could instead be a payment difference, novelty effect or difference in how often users are contacted.

Do not deliberately make the core outcome unreliable to create suspense. There is a difference between not knowing which interesting recommendation will appear and not knowing whether a tool will fulfill its promise. The former may merit testing; the latter may simply degrade the service. This is a product judgment to be evaluated through outcomes, not a neuroscientific rule.

4 · Make investment earn its cost

For each requested contribution, state how it improves the next experience. A preference rating might improve recommendations; an annotation might make future retrieval easier. A meaningless setup task has no such direct justification.

Separate at least three possibilities: the contribution improves the service, it predicts that the user was already motivated, or performing it changes later commitment. An observational association between contribution and return cannot discriminate among them. Where feasible, randomize the invitation or timing and evaluate everyone assigned to each condition, not just those who completed the contribution.

Measure burden too: time to initial value, abandonment, errors, privacy concern and willingness to repeat. Successful investment is not simply the amount of data collected or the number of steps a user endured.

5 · Use a measurement stack, not a single “habit score”

Layer	What to record	What it cannot establish alone
User benefit	A product-specific outcome and an explicit user evaluation	The psychological mechanism causing it
Behavior	Completion, latency, return and relevant opportunities	Automaticity or welfare
Context	The cues and situations around the specific action	Causal influence without comparison
Automaticity	A carefully adapted, tested measure for that action	An objective diagnosis of habit
Agency and burden	Control, regret, stopping, opt-outs and relevant costs	Clinical addiction or universal safety

The four-item Self-Report Behavioural Automaticity Index offers an established starting point, but its validity should not be assumed unchanged for every new wording, behavior or population. Self-report is useful evidence, not a window that directly reveals unconscious processing. [17]

Outcome-devaluation paradigms provide another approach to action control, but they are not an effortless gold standard. Gera et al. demonstrate why naturalistic task designs can be informative. A separate 2026 methodological paper shows why an ineffective devaluation manipulation can make goal-directed behavior look habitual. That later paper is **not treated here as a direct replication of Gera’s smartphone study**. The practical lesson is to verify comprehension and actual changes in outcome value before interpreting persistent responding. [15,18]

6 · Predefine the comparison, duration and decision

Choose one primary user-relevant endpoint, a minimum worthwhile effect, a measurement window justified by the behavior’s cadence, and guardrails for burden or loss of control. Determine sample requirements before looking for significance. Where results are uncertain, report the interval of plausible effects rather than declaring success or failure from a threshold alone.

Use concurrent randomized comparison when feasible. Account for repeated observations from the same person; account for group-level assignment or spillovers when social features affect other users. Avoid concluding that a feature works merely because users who adopted it later had better retention. That is a selection problem unless the analysis credibly addresses it.

Keep a separate follow-up question: does the effect last after the novelty wears off or reminders are reduced? A short-term increase while incentives are active is not evidence of a durable automatic routine. Changing reminders should itself be studied carefully, because reminder removal changes the cue environment and may reduce use for reasons other than the absence of a habit.

7 · Ship based on net user value, not a completed four-box diagram

A feature can be worth shipping even if it does not form a habit. A feature can also form a habit and still be a poor choice. The decision rule should therefore be whether the intervention improves the intended outcome enough to justify its costs while preserving acceptable user control.

Illustrative application: a recommendation service. Show a useful initial recommendation, then invite a small amount of preference feedback. Randomize whether the feedback invitation appears now or later. Test whether resulting recommendations actually improve and whether the burden reduces initial success. Interesting variation may be inherent in discovering new options; an artificial random-prize mechanism is not required merely because the framework contains a variable-reward stage.

Illustrative application: a knowledge-capture tool. Help users save a note at the moment they have something worth preserving, then make retrieval reliable. The stored note creates a rational reason to return. Repeat use may become automatic in a recurring context, but that is a separate hypothesis to measure. An efficient search session that ends quickly should not be counted as a retention problem by default.

These examples retain trigger, action, outcome and useful contribution as design questions. They reject the idea that every successful interaction must pass through a rigid sequence or maximize time spent.

08 What changes with time—and what should change in the book

What an updated edition should change

The following replacement formulations are proposed editorial language, **not quotations from Eyal**. They identify the smallest repairs needed to preserve the useful claim without overstating the evidence.

Passage or idea	Proposed replacement
Introduction: predictable loops do not create desire; variability suppresses reason (Introduction)	Predictable outcomes can motivate repeat use. Relevant uncertainty can sometimes increase attention or persistence, but its effects depend on the task, reward and user. The cited evidence does not establish a general suppression of judgment in ordinary app users.
Habit Zone as a threshold (ch. 1)	Frequency and perceived utility are useful planning dimensions. This diagram is conceptual, not calibrated; it cannot identify a numerical threshold for habit formation.
Depression study methods (ch. 2)	The study associated depressive-symptom survey scores with a month of measured campus Internet use. It did not establish that Internet use relieved symptoms or caused them.

Passage or idea	Proposed replacement
“Always” improve ability first (ch. 3)	Reducing friction is often worth testing, especially when users already want the outcome. First identify whether difficulty, value, trust or another factor is constraining action.
Wine study (ch. 3)	Participants tasted three wines under five price labels, with two wines each presented at different stated prices. Higher stated prices altered reported pleasantness and related fMRI activity.
Olds–Milner experiment (ch. 4)	The cited 1954 experiment studied electrical stimulation of the septal area and other regions of rat brains. It is part of the history of reinforcement research, not a direct model of ordinary app use.
Dopamine claims attached to Berns and Aharon (ch. 4)	These studies measured fMRI activity, not dopamine release. They examined predictability of liquid delivery and differences in the reward value of face categories, respectively.
“But you are free” (ch. 4)	An earlier meta-analysis found higher odds of compliance. A later quality-focused analysis raises uncertainty about the size and reliability of the effect. Neither establishes a universal effect on product retention.
Investment as necessary (ch. 4; ch. 5)	Contributions that improve future value or create relevant follow-up opportunities can encourage return. A separate post-reward investment step has not been established here as necessary or optimal for every product.
Labor leads to love (ch. 5)	In the cited experiments, successful construction increased valuation. The effect did not extend in the same way to unfinished or dismantled creations. Do not add arbitrary effort and expect appreciation.
The “addicted 1 percent” (ch. 6)	The cited figure concerns an adult-population estimate of pathological gambling. It does not establish the risk among users of a specific technology or delimit all potential harms.
Bible App neuroscience (ch. 7)	The cited paper reports original self-disclosure experiments, not a meta-analysis. Its relevance to sharing scripture is indirect and does not establish the app’s growth mechanism.
Habit Testing (ch. 8)	Use engagement paths to generate hypotheses, then test their causal effects. Measure the automaticity of a specified action separately from use frequency and user benefit.

The most valuable editorial addition would be a visible label distinguishing **established finding**, **tentative mechanism**, **business illustration** and **design hypothesis**. That change would do more for scientific clarity than adding further neuroscience anecdotes. A revised bibliography should also identify whether each empirical claim comes from an original study, a review, a company report, a secondary article or an interview.

Publication-time accuracy versus later evidence

Temporal fairness

Errors demonstrable from sources already available by 2014—such as the rat-brain citation, the wine procedure and the distinction between BOLD and dopamine—are edition-level accuracy problems. Later findings, including the 2023 compliance re-examination and the 2026 measurement critique, are updates to the present evidentiary picture. They are not used to accuse the author of failing to know future research. A 2011 working paper is not misdated merely because its final journal publication appeared in 2012.

09 How the ratings were determined

The headline scores summarize three different editorial judgments. **Scientific Accuracy** evaluates three central propositions; **Reference Accuracy** evaluates traceability, description and inference in the examined evidence; **Practical Value** evaluates likely benefits, applicability and net value. Ease of application is reported separately and is not included in the average.

The source audits are not identical probability samples. Reference Accuracy therefore uses the same three anchored criteria in every review—not a percentage of references that passed an audit. Both faithful and problematic source use inform the assessment, and inaccessible passages are not counted as errors. See Sections 02, 06 and 10 for scope and coverage.

Overall rating: 50%. The three categories receive equal weight. This is an editorial convention, not an empirically validated estimate of overall truth or benefit. The percentage does not override the recommendation or the consequential caution on the cover.

Scientific Accuracy · 50%

Do the book's main claims hold up?

Frequent, cue-linked repetition can make specific product actions more automatic.

The underlying cue–repetition principle is reasonably supported. Retention and frequency remain imperfect proxies for automaticity, and no universal usage threshold follows. [Section 04]

Variable rewards are a privileged driver of repeat engagement and habit formation.

Reward schedules and uncertainty can matter. The evidence does not establish variability as necessary or generally superior across products, rewards, and outcomes. [Section 04]

Investment and the complete Hook sequence reliably create durable product habits.

The full sequence is not validated by component experiments or retrospective company cases. Investment may create value or switching costs without demonstrating the proposed habit mechanism. [Section 04]

Reference Accuracy · 50%

Do the cited sources support what the book says?

Traceability

For the audited examples, the notes generally make the intended paper or source identifiable; access and correctness remain separate questions. [Section 06]

Accurate description

Several high-salience neuroscience and prevalence illustrations materially misdescribe a measurement, result, or denominator. [Section 06]

Appropriate inference

The narrative repeatedly extends laboratory associations, clinical samples, and valuation effects to stronger product-habit mechanisms. [Section 06]

Practical Value · 50%

Is the advice likely to be worthwhile for its intended reader?

The intended user is a product builder. Commercial retention can be a legitimate objective, but Practical Value also considers the people affected by the design. The score does not equate high engagement with addiction or assume that a frequently used product is harmful.

Evidence of intended benefit

Cue-linked repetition, reducing friction and delivering recurring value have relevant support, and the checklist can help organize design work. The complete trigger–action–variable-reward–investment sequence is not directly validated by the evidence reviewed. A descriptive case study of successful products and a component reward-schedule trial do not establish the package's incremental benefit. [2; 3; 14; 16]

Applicability and durability

The book acknowledges that not every product needs frequent use and that variable rewards cannot rescue an unwanted service. These are important qualifications. Generalization across products remains uncertain, however, and engagement, automaticity and durable user outcomes need different measures. The book's retrospective Habit Testing approach does not by itself remove selection or novelty effects. [book; 6; 17]

Benefits relative to burdens and risks

The text explicitly favors beneficial products and discusses privacy, deceptive invitations and unhealthy attachment. It therefore deserves more credit than a blanket charge of endorsing manipulation. Nevertheless, self-reported good intentions and the designer's own use are insufficient safeguards against burdens, unwanted persistence or manufactured exit costs; these need user-centered outcome checks. [book; Section 07]

Practical Value is 50%: there are credible components and useful design questions, but package-level effectiveness, durable transfer, and sufficient governance of engagement-versus-benefit trade-offs remain unresolved.

Calculation and scoring anchors

The full audit uses five anchored grades, from 0 to 4. These are the inputs behind the percentages, not an additional reader-facing rating system. At a given criterion, 0 denotes a demonstrated fundamental failure or contradiction; 1 major limitations; 2 partial or mixed support; 3 substantial support with qualifications; and 4 strong support at the defined scope. The series methodology gives category-specific anchors and missing-data rules.

Scientific Accuracy: (3 + 2 + 1) ÷ 12 × 100 = 50.000% ➜ **50%**.

Reference Accuracy: (4 + 1 + 1) ÷ 12 × 100 = 50.000% ➜ **50%**.

Practical Value: (2 + 2 + 2) ÷ 12 × 100 = 50.000% ➜ **50%**.

Overall: (6 + 6 + 6) ÷ 36 × 100 = 50.000% ➜ **50%**. This is exactly the average of the three *unrounded* category scores. Only the final displayed values are rounded.

Uncertainty. A one-grade disagreement on a single criterion changes its category by about 8.3 percentage points and the overall rating by about 2.8 points before rounding. This is a sensitivity calculation, not a confidence interval. Independent duplicate scoring and measured inter-rater reliability are not documented. Small differences between books should not be interpreted as precise scientific rankings.

Missing evidence. No category in this edition is withheld as unrateable, but some individual source passages remain unresolved. Those gaps retain their explicit access labels; they are not zeroes or hidden failed checks. If a whole required criterion could not responsibly be judged, the category and overall score would be withheld rather than silently changing the denominator.

Score rationale

The nine inputs follow the stated anchors. Scientific Accuracy is (3, 2, 1): cue-linked repetition has substantial support; variability is mixed; and the complete sequence has major limitations. Reference Accuracy is (4, 1, 1): the full ACM paper verifies task quantity, but the book's other measurement, procedure, denominator, and inference problems meet these anchors. Practical Value is (2, 2, 2): favorable component findings do not establish package-level effectiveness, durable transfer, or net user benefit. Neither the IKEA and wine follow-ups nor the broader habit synthesis provides enough support for a higher criterion grade. Repository availability is not successful reproduction; a computational disagreement is not an empirical replication failure.

10 Verification record and remaining limits

Research stages and access. The source-access labels describe what was retrieved and inspected at each stated date. The first ledger below records targeted source checks on 8 September 2026. "Primary text" means the stated material was retrieved, not that an entire article, its supplements or its dataset was independently audited. A repeat metadata or abstract check does not upgrade access to full text. The dated checks below identify the research performed and its limits; not every search or retrieval was independently repeated.

Study-integrity assessment — 13 September 2026

The preparation record documents checks of the 30 initial source-table rows and 18 registered claims for relevant replication, reproducibility, correction, retraction, and expression-of-concern issues; additional named studies were traced where identifiable. It records source-specific searches and the series' exact-DOI Crossref/Retraction Watch screen. Not every search or retrieval was independently repeated. No retraction or expression of concern was confirmed for the exact Hooked studies examined; that is not a guarantee that none exists. The reader review, appendix, and relevant book endnote pages were inspected. The full official ACM report supplies the methods and results for the reciprocity assessment. Sources 30–44 document favorable and mixed IKEA, wine, habit, compliance, and clinical evidence, the dopamine interpretation exchange, and dataset access. Identification of clinical sources does not mean every underlying medical trial was individually audited.

Exact-title, author, and DOI searches were paired with "replication," "retraction," "correction," "expression of concern," "data," "code," and relevant synthesis terms. The retained local research register records exact queries, primary URLs, publication/version identity, actual access, and unresolved items. This is a targeted critical narrative review, not preregistered or exhaustive database coverage. No participant-level dataset, imaging analysis, or code was independently executed. An original-report access upgrade is not replication success; a nonsignificant contrast is not equivalence; and a computational critique is not an empirical failed replication.

Check and source	Finding at the checked scope	Remaining limitation
K01 Lally et al.: habit-formation trajectories Primary abstract and text retrieved	The study supports heterogeneous trajectories. Its long endpoint is model-estimated beyond the 84-day observation period; it is not a universal timetable. The book deserves credit for rejecting one universal duration. Chapter 1 / audit A01.	No re-fitting of individual automaticity curves.
K02 Olds & Milner (1954): brain stimulation Primary bibliographic record retrieved	The title and record identify 1954, rats, and the septal area and other regions, rather than the book's 1940s/mice/anatomical account. This does not dispute that brain stimulation can reinforce behavior. The relevant passage in chapter 4 and its note 1 were inspected; the passage was also checked visually.	Metadata establishes only the narrow historical, species and source-identity correction; not every deprivation anecdote.
K03 Nuijten et al. (2021): reinforcement schedules Primary abstract and text retrieved	The six-week, 61-participant study reports similar engagement across three schedules, with a cost advantage for one variable schedule. It is not evidence that variable rewards invariably outperform fixed rewards for durable habits. Audit and later-evidence discussion.	A nonsignificant comparison is not proof of equivalence; the trial is small and context specific.
K04 Lukyanchikova et al. (2023): Hook Model cases Primary abstract and article text retrieved	Descriptive mapping of Uber and Instagram is not a randomized comparison of the Hook Model with alternative designs. Central claim 3.	The check does not rule out other trials or the practical usefulness of the model.
K05 Fillon et al. (2023): but-you-are-free meta-analysis Primary journal abstract retrieved	The overall effect was g = 0.44; the seven low-risk studies gave g = 0.11 with a confidence interval spanning zero. A robust universal compliance guarantee is not supported by that contrast. Audit and later-evidence discussion.	The low-risk interval is compatible with more than one effect size; it is not proof of a zero effect.
K06 Vázquez-Millán et al. (2026): devaluation measurement Primary abstract and article text retrieved	The preregistered conceptual replication shows how deficient devaluation can produce apparently habitual responding. It cautions against equating repeated or persistent responding with established habit. Later methodological context.	A 2026 measurement study is not evidence of what a 2014 author knew, and does not show that habits do not exist.
K07 Book acknowledgments Book acknowledgments visually inspected	The acknowledgments name Jason Hreha. This is disclosed, without inferring the extent of his contribution or relationship. Acknowledgments.	Does not establish endorsement, compensation, collaboration details or a current relationship.

Edition identification and source limits

Edition details are taken from the book's title and copyright pages. They do not authenticate the reviewed copy against a publisher-controlled master or establish correspondence with every commercially distributed edition.

What has not been independently certified

This review is not independent human fact-checking, expert peer review, legal clearance, an exhaustive quotation audit, a complete retraction/correction search, or participant-level reanalysis. Historical anecdotes, private interviews, unpublished data, blocked sources, and any unresolved numerical claims retain their original limits. Repeated AI review does not create independent corroboration.

HTML is the canonical report text; the PDF is generated from that same text without abridgment. The package's validation record documents the executed structural, scoring, citation-target, text-coverage, and rendering checks and their results. A functioning reference link is not proof that the linked study supports an assertion. Nor does a clean PDF layout validate the science.

Substantive access limits

Access limits

Some sources were available as full articles or author manuscripts; others only as publisher abstracts, metadata or substantial extracted passages. The bibliography labels this distinction. The AGA denominator check relied on recovered text of a reproduced industry white paper because the original hosting link was not recovered as accessible. The Bible App's 66,000-per-second statement remains an unresolved unit claim. These limitations are not silently filled with plausible substitutes.

The report does not certify every endnote, historical market figure, quoted anecdote, clinical statement or financial comparison in the book. It does not independently reproduce statistical analyses, audit datasets, establish freedom from publication bias, or perform a complete retraction-status audit of every reference. It also does not evaluate the benefits or harms of every named app. The selective audit is substantially more informative than accepting a bibliography on trust, but it is not universal verification.

Temporal fairness

Errors demonstrable from sources already available by 2014—such as the rat-brain citation, the wine procedure and the distinction between BOLD and dopamine—are edition-level accuracy problems. Later findings, including the 2023 compliance re-examination and the 2026 measurement critique, are updates to the present evidentiary picture. They are not used to accuse the author of failing to know future research. A 2011 working paper is not misdated merely because its final journal publication appeared in 2012.

Additional source check · 8 September 2026

Additional source check · Lukyanchikova et al. (2023), A case study on applications of the Hook Model. Primary publisher abstract and methods inspected on 8 September 2026. The paper reverse-engineers use cases in two successful products. It is a comparative descriptive study, not a randomized test that the complete Hook sequence improves habit-specific or welfare outcomes.

Limit: No new experiment or reanalysis; descriptive coverage is not an estimate of causal effect. This check does not constitute a new comprehensive literature search.

Production checks cover score arithmetic, internal links, preservation of the substantive analysis, agreement of HTML and PDF text, and representative browser and PDF rendering. The accompanying production record reports the actual results; these checks do not establish

scientific or legal clearance.

11 References and source access

These are the sources cited in the evidence dossiers. Their dated access notes state what was checked; an abstract or metadata record does not imply full-text verification. Source checks and unsuccessful access attempts are distinguished in Section 10. Identifying a source, including a review article or institutional guidance, is not itself evidence that it supports every claim associated with it.

Book citations use chapters and named sections rather than edition-dependent page numbers. Reference numbers are local to this report. Codes such as T/full text, A/abstract, M/metadata, or U/unresolved are defined in the local source ledger; a source can have accessible metadata while a particular passage remains unverified. "Full text" means the stated article or manuscript content was inspected, not that its data were reproduced. "Extracts" means substantial source passages were retrieved but complete coverage was not certified. Source-access labels state checks on 8 September and 13 September 2026; the date and access level are specified for each check. External links require internet access; the review text and internal navigation work offline.

Sources and access ledger

Book locators use **B**. Numbered sources identify the particular research or methodological source supporting the adjacent text. "Full text" means the article or manuscript content was inspected, not that its data were reproduced. "Extracts" means substantial source passages were retrieved but complete coverage was not certified. The ledger records online sources consulted on 8 September 2026. Further source checks are separately dated 13 September 2026.

Row	Source identity, bibliography, links, and access record
01	**B · Reviewed book.** Eyal, N., with Hoover, R. (2014). *Hooked: How to Build Habit-Forming Products.* Portfolio / Penguin. Ebook ISBN 978-0-698-19066-5. Chapters, figures, acknowledgments and endnotes inspected. Citations identify chapters and named sections.
02	**1 · Review method.** Red Pen Reviews. *Our Process*; full scoring method, version 2.2. Official process page and rubric inspected. Official process · Rubric PDF.
03	**2 · Habit theory.** Wood, W., & Neal, D. T. (2007). A new look at habits and the habit–goal interface. *Psychological Review, 114*(4), 843–863. DOI: 10.1037/0033-295X.114.4.843. Access: author-hosted manuscript, substantial text inspected.

Row	Source identity, bibliography, links, and access record
04	**3 · Formation trajectories.** 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. DOI: 10.1002/ejsp.674. Access: publisher abstract and study/method extracts; not a data reanalysis.
05	**4 · Flossing.** Judah, G., Gardner, B., & Aunger, R. (2013). Forming a flossing habit: An exploratory study of the psychological determinants of habit formation. *British Journal of Health Psychology, 18*, 338–353. DOI: 10.1111/j.2044-8287.2012.02086.x. Access: publisher abstract, including design and results summary.
06	**5 · Broader habit synthesis.** Wood, W., & Rünger, D. (2016). Psychology of habit. *Annual Review of Psychology, 67*, 289–314. DOI: 10.1146/annurev-psych-122414-033417. Access: publisher abstract; used for broad framing, not fine-grained quantitative claims.
07	**6 · Smartphone checking.** Oulasvirta, A., Rattenbury, T., Ma, L., & Raita, E. (2012; online 2011). Habits make smartphone use more pervasive. *Personal and Ubiquitous Computing, 16*, 105–114. DOI: 10.1007/s00779-011-0412-2. Access: University of Helsinki research record and original abstract.
08	**7 · Fogg's conceptual model.** Fogg, B. J. (2009). A behavior model for persuasive design. *Proceedings of Persuasive '09.* DOI: 10.1145/1541948.1541999. Access: original conference-paper text in an accessible reproduction. A corroborating primary articulation of the model referenced by the book's website notes.
09	**8 · Predictability and fMRI.** Berns, G. S., McClure, S. M., Pagnoni, G., & Montague, P. R. (2001). Predictability modulates human brain response to reward. *Journal of Neuroscience, 21*(8), 2793–2798. DOI: 10.1523/JNEUROSCI.21-08-02793.2001. Access: journal full text.
10	**9 · Face reward value.** Aharon, I., Etcoff, N., Ariely, D., Chabris, C. F., O'Connor, E., & Breiter, H. C. (2001). Beautiful faces have variable reward value: fMRI and behavioral evidence. *Neuron, 32*(3), 537–551. DOI: 10.1016/S0896-6273(01)00491-3. Access: journal and author-hosted full-text extracts, including methods.
11	**10 · Reward cues and risk.** Knutson, B., Wimmer, G. E., Kuhnen, C. M., & Winkielman, P. (2008). Nucleus accumbens activation mediates the influence of reward cues on financial risk taking. *NeuroReport, 19*(5), 509–513. DOI: 10.1097/WNR.0b013e3282f85c01. Access: author-hosted paper, methods and results inspected.

Row	Source identity, bibliography, links, and access record
12	**11 · Electrical reinforcement.** Olds, J., & Milner, P. (1954). Positive reinforcement produced by electrical stimulation of septal area and other regions of rat brain. *Journal of Comparative and Physiological Psychology, 47*, 419–427. PubMed 13233369. Access: bibliographic record and the book's own endnote; sufficient for date, species and named-site corrections, not a complete experiment audit.
13	**12 · Pathological gambling.** Brevers, D., & Noël, X. (2013). Pathological gambling and the loss of willpower: A neurocognitive perspective. *Socioaffective Neuroscience & Psychology, 3*, 21592. DOI: 10.3402/snp.v3i0.21592. Access: original abstract and open article record; used to establish clinical population and review/theoretical scope.
14	**13 · Dopamine and uncertainty.** Fiorillo, C. D., Tobler, P. N., & Schultz, W. (2003). Discrete coding of reward probability and uncertainty by dopamine neurons. *Science, 299*, 1898–1902. DOI: 10.1126/science.1077349. Access: publisher abstract; used for the narrow probability/uncertainty finding.
15	**14 · Direct discussion of the Hook Model.** Lukyanchikova, E., Askarbekuly, N., Aslam, H., & Mazzara, M. (2023). A case study on applications of the Hook Model in software products. *Software, 2*(2). DOI: 10.3390/software2020014. Access: substantial original-paper text, including comparative-case methodology; publisher retrieval was intermittently blocked.
16	**15 · Naturalistic habit induction.** Gera, R., Barak, S., & Schonberg, T. (2024; online 2023). A novel free-operant framework enables experimental habit induction in humans. *Behavior Research Methods, 56*, 3937–3958. DOI: 10.3758/s13428-023-02263-6. Access: publisher full text; not an independent reproduction of its statistical models.
17	**16 · Reward-schedule app trial.** Nuijten, R., Van Gorp, P., Khanshan, A., Le Blanc, P., Kemperman, A., van den Berg, P., & Simons, M. (2021). Health promotion through monetary incentives: Evaluating the impact of different reinforcement schedules on engagement levels with a mHealth app. *Electronics, 10*(23), 2935. DOI: 10.3390/electronics10232935. Access: publisher PDF text, including design, conditions and results. 13 September 2026 source check: The original data statement identifies the public Figshare dataset at reference 38. It was not downloaded or independently reanalyzed. The small trial's nonsignificant comparison is not equivalence.
18	**17 · Automaticity measurement.** Gardner, B., Abraham, C., Lally, P., & de Bruijn, G.-J. (2012). Towards parsimony in habit measurement: Testing the convergent and predictive validity of an automaticity subscale of the Self-Report Habit Index. *International Journal of Behavioral Nutrition and Physical Activity, 9*, 102. DOI: 10.1186/1479-5868-9-102. Access: original-paper text and methods/results extracts.

Row	Source identity, bibliography, links, and access record
19	**18 · Measurement caution, 2026.** Vázquez-Millán, A., Martínez-López, P., Rueda, M., León, J. J., & Luque, D. (2026). The evaluation of devaluation: Deficient outcome devaluation leads to wrongly considering goal-directed actions as habits. *Behavior Research Methods, 58*, article 224. DOI: 10.3758/s13428-026-03099-6. Access: publisher full text; published 7 July 2026. Used as a separate methodological warning, not as a direct replication of reference 15.
20	**19 · IKEA effect.** Norton, M. I., Mochon, D., & Ariely, D. (2012; working paper 2011). The IKEA effect: When labor leads to love. *Journal of Consumer Psychology, 22*, 453–460. DOI: 10.1016/j.jcps.2011.08.002. Access: Harvard working paper and journal-version text, including completion conditions and valuation procedure. 13 September 2026 source check: Independent conceptual replication at reference 30 is favorable for valuation and ownership but mixed for the proposed mechanism and completion comparison. The original completion condition is not treated as universally necessary.
21	**20 · Foot-in-the-door.** Freedman, J. L., & Fraser, S. C. (1966). Compliance without pressure: The foot-in-the-door technique. *Journal of Personality and Social Psychology, 4*(2), 195–202. DOI: 10.1037/h0023552. Access: original-paper reproduction and substantial text. The review uses the conditional compliance finding, not a contemporary pooled effect estimate.
22	**21 · Computer reciprocity.** Fogg, B. J., & Nass, C. (1997). How users reciprocate to computers: An experiment that demonstrates behavior change. *CHI '97 Extended Abstracts on Human Factors in Computing Systems*, 331–332. DOI: 10.1145/1120212.1120419. Earlier 8 September 2026 access: ACM abstract only; exact effect magnitude and full procedure not independently verified. 13 September 2026 source check: The source assessment uses the complete official CHI '97 electronic report: full methods, results, tables, and conclusions inspected. Near-double color-comparison quantity is confirmed; neither original-data reproduction nor stage-order validation is claimed.
23	**22 · Price and wine pleasantness.** Plassmann, H., O'Doherty, J., Shiv, B., & Rangel, A. (2008). Marketing actions can modulate neural representations of experienced pleasantness. *Proceedings of the National Academy of Sciences, 105*(3), 1050–1054. DOI: 10.1073/pnas.0706929105. Access: full paper via PubMed Central, including experimental conditions. 13 September 2026 source check: The favorable 2017 same-program wine study is documented at reference 31. It does not repair the original book's procedure description.
24	**23 · Endowed progress.** Nunes, J. C., & Drèze, X. (2006). The endowed progress effect: How artificial advancement increases effort. *Journal of Consumer Research, 32*(4), 504–512. DOI: 10.1086/500480. Access: author-hosted manuscript; study 1 text and relevant PDF page inspected.
25	**24 · Scarcity and valuation.** Worchel, S., Lee, J., & Adewole, A. (1975). Effects of supply and demand on ratings of object value. *Journal of Personality and Social Psychology, 32*(5), 906–914. DOI: 10.1037/0022-3514.32.5.906. Access: detailed original abstract; raw results not reanalyzed.

Row	Source identity, bibliography, links, and access record
26	**25 · Original compliance meta-analysis.** Carpenter, C. J. (2013). A meta-analysis of the effectiveness of the “But You Are Free” compliance-gaining technique. *Communication Studies, 64*(1), 6–17. DOI: 10.1080/10510974.2012.727941. Access: original article in a reproduced full-text copy; sample and odds-ratio results inspected.
27	**26 · Later compliance re-examination.** Fillon, A., Souchet, L., Pascual, A., & Girandola, F. (2023). The effectiveness of the “But-you-are-free” technique: Meta-analysis and re-examination of the technique. *Meta-Psychology, 7.* DOI: 10.15626/MP.2020.2640. Access: journal abstract and full article; low-risk-of-bias estimate checked against the published account.
28	**27 · Internet use and depressive symptoms.** Kotikalapudi, R., Chellappan, S., Montgomery, F., Wunsch, D., & Lutzen, K. (2012). Associating Internet usage with depressive behavior among college students. *IEEE Technology and Society Magazine, 31*(4), 73–80. DOI: 10.1109/MTS.2012.2225462. Access: journal record and original-paper abstract, including one-month design, CES-D and network-data methods.
29	**28 · Industry prevalence source.** Stewart, D. (2010). *Demystifying Slot Machines and Their Impact in the United States.* American Gaming Association. Access: reproduced primary white-paper text, including its adult-population denominator; original hosting link not recovered as accessible. Reproduced document. This is an industry source, not an independent prevalence estimate endorsed by this report.
30	**29 · Self-disclosure.** Tamir, D. I., & Mitchell, J. P. (2012). Disclosing information about the self is intrinsically rewarding. *Proceedings of the National Academy of Sciences, 109*(21), 8038–8043. DOI: 10.1073/pnas.1202129109. Access: original journal abstract; sufficient to establish experimental rather than meta-analytic design and the reported qualitative findings.
31	30 · Sarstedt, M., Neubert, D., & Barth, K. (2017). *The IKEA Effect. A Conceptual Replication.* DOI: 10.1561/107.00000039. Published 13 April 2017. Primary source. Primary published paper embedded in the author’s university thesis: indexed methods/results/discussion excerpts inspected. Direct publisher downloads failed; raw data not reanalyzed. Checked 13 September 2026.
32	31 · Schmidt, L., Skvortsova, V., Kullen, C., Weber, B., & Plassmann, H. (2017). *How context alters value: The brain’s valuation and affective regulation system link price cues to experienced taste pleasantness.* DOI: 10.1038/s41598-017-08080-0. Published 14 August 2017. Primary source. Primary publisher abstract, design, results, and discussion inspected. Same research program; no independent execution of imaging/mediation analyses. Checked 13 September 2026.

Row	Source identity, bibliography, links, and access record
33	32 · Burger, J. M. (1999). *The foot-in-the-door compliance procedure: A multiple-process analysis and review.* DOI: 10.1207/s15327957pspr0304_2. Primary source. Primary abstract and mechanism-discussion extracts inspected; synthesis, not a fresh exact safe-driving-sign replication or habit study. Checked 13 September 2026.
34	33 · Dillard, J. P., Hunter, J. E., & Burgoon, M. (1984). *Sequential-request persuasive strategies: Meta-analysis of foot-in-the-door and door-in-the-face.* DOI: 10.1111/j.1468-2958.1984.tb00028.x. Primary source. Primary publisher abstract inspected. Small pooled compliance effects are not an expected onboarding multiplier; full pooled data not reproduced. Checked 13 September 2026.
35	34 · Ladeira, W. J., et al. (2023). *A meta-analysis on the effects of product scarcity.* DOI: 10.1002/mar.21816. Primary source. Primary publisher abstract inspected; full methods/results not accessed. Different scarcity types and consumer outcomes must not be collapsed into an app-habit estimate. Checked 13 September 2026.
36	35 · Singh, B., Murphy, A., Maher, C., & Smith, A. E. (2024). *Time to Form a Habit: A Systematic Review and Meta-Analysis of Health Behaviour Habit Formation and Its Determinants.* DOI: 10.3390/healthcare12232488. Primary source. Primary abstract, methods, results, and bias/discussion extracts inspected. High-risk studies and limited duration evidence retained; pre/post estimates are not randomized package effects. Checked 13 September 2026.
37	36 · Niv, Y., Duff, M. O., & Dayan, P. (2005). *Dopamine, uncertainty and TD learning.* DOI: 10.1186/1744-9081-1-6. Primary source. Primary abstract and analytical interpretation excerpts inspected; direct PMC access was intermittent. Computational alternative, not an empirical failed replication; original neural data not executed. Checked 13 September 2026.
38	37 · Fiorillo, C. D., Tobler, P. N., & Schultz, W. (2005). *Evidence that the delay-period activity of dopamine neurons corresponds to reward uncertainty rather than backpropagating TD errors.* DOI: 10.1186/1744-9081-1-7. Published 15 June 2005. Primary source. Primary abstract and reply excerpts inspected; direct PMC access was intermittent. Authors' interpretive response retained alongside the critique; no raw spike-data reproduction. Checked 13 September 2026.

Row	Source identity, bibliography, links, and access record
39	38 · Nuijten et al. (2021). *Monetary-incentive mHealth trial dataset.* Figshare. DOI: 10.6084/m9.figshare.16692271. Primary source. Dataset identity identified in the original MDPI data-availability statement. Dataset not downloaded or reanalyzed; public availability is not successful reproduction. Checked 13 September 2026.
40	39 · Bouton, M. E. (2004). *Context and Behavioral Processes in Extinction.* Learning & Memory, 11(5), 485–494. DOI: 10.1101/lm.78804. Primary source. Exact book endnote recovered; official abstract and primary theoretical discussion inspected. Context-dependent learning review, not a longitudinal medical test of every human habit. Checked 13 September 2026.
41	40 · Kirshenbaum, A. P., Olsen, D. M., & Bickel, W. K. (2009). *A quantitative review of the ubiquitous relapse curve.* DOI: 10.1016/j.jsat.2008.04.001. Primary source. Exact book endnote recovered; primary abstract, methods, results, and discussion inspected. Selected alcohol/tobacco curves and digitized graphs; no participant-level reproduction or universal clinical mechanism. Checked 13 September 2026.
42	41 · Zywiak, W. H., Kenna, G. A., & Westerberg, V. S. (2011). *Beyond the Ubiquitous Relapse Curve: A Data-Informed Approach.* DOI: 10.3389/fpsyt.2011.00012. Primary source. Complete official PubMed abstract inspected. Broad first-drink curve supported, with hidden recovery/drinking transitions; full original daily-data analyses not reexecuted. Checked 13 September 2026.
43	42 · Jeffery, R. W., et al. (2000). *Long-term maintenance of weight loss: Current status.* DOI: 10.1037/0278-6133.19.suppl1.5. Primary source. Exact book endnote and complete official abstract inspected. Original full synthesis and every embedded clinical trial not freshly audited; no universal individual regain prediction. Checked 13 September 2026.
44	43 · Hartmann-Boyce, J., et al. (2021). *Association between characteristics of behavioural weight loss programmes and weight change after programme end: systematic review and meta-analysis.* BMJ, 374, n1840. DOI: 10.1136/bmj.n1840. Primary source. Primary BMJ abstract, method/bias, and discussion extracts inspected. Comparator/time variation and sparse longer follow-up retained; across-trial features are not randomized causal mechanisms. Checked 13 September 2026.

Row	Source identity, bibliography, links, and access record
45	44 · Wood, W., Quinn, J. M., & Kashy, D. A. (2002). *Habits in everyday life: Thought, emotion, and action.* DOI: 10.1037/0022-3514.83.6.1281. Primary source. Exact book endnote and official original abstract inspected. Two diary studies; frequency/context self-report is not an objective automaticity gold standard. Raw diaries not reproduced. Checked 13 September 2026.

12 Editorial Review Notice, Disclosures, and Corrections

Purpose and scope.

This report is published for criticism, commentary, education, and discussion of matters of public interest. It evaluates the particular work, edition, claims, and evidence identified in the report. Its conclusions should not be extended to editions, claims, publications, or professional activities that it does not examine. It is not a comprehensive audit of the author's work or a certification of scientific accuracy.

Factual reporting and editorial judgment.

This review contains both factual reporting and editorial analysis. Quotations, descriptions of source material, and reported research findings are presented as factual information. Assessments of evidentiary strength, methodological quality, interpretation, balance, and practical value are the reviewer's evaluative conclusions, based on the sources, reasoning, and criteria identified in the report. Readers are encouraged to examine those materials and assess the conclusions for themselves. Evaluative judgments are not intended to imply possession of undisclosed facts concerning anyone's conduct or motives.

Meaning of critical assessments.

Descriptions such as "unsupported," "overstated," "misleading," or "inconsistent with the evidence" concern the specific claim, presentation, or evidentiary relationship discussed, for the reasons explained in the accompanying analysis. They do not, by themselves, assert that an author or other person knowingly made a false statement, intended to deceive, fabricated evidence, or engaged in professional misconduct. An assessment that the available evidence does not adequately support a claim is distinct from a finding that the claim is necessarily false.

Interpretation of scores.

Scores, grades, rankings, and recommendation categories summarize the reviewer's application of the stated rubric. They are not measurements of the author's honesty, character, intentions, or overall professional competence. They are not estimates of the

percentage of a book that is true or false, probabilities that an author is correct, or representations of scientific consensus. The accompanying analysis explains the basis and limitations of each assessment.

Evidence, timing, and verification limits.

Conclusions reflect the materials examined and the state of the evidence assessed through the literature-search cutoff stated in this report. Relevant limitations—including unavailable sources, incomplete access to underlying data, and unresolved verification questions—are identified in the methodology or accompanying analysis. An inability to locate or independently verify a source does not, by itself, establish that the source does not exist or that a claim was fabricated. Evidence published after the reviewed work is distinguished from evidence available when it was written; later developments do not, by themselves, establish what an author knew or should have known at the time. Independent replication of studies, reanalysis of original datasets, and expert peer review are not claimed unless expressly documented.

Preparation and review disclosure.

This paired reader review and technical appendix was prepared with ChatGPT at Jason Hreha's request. AI tools materially assisted research retrieval, comparison, targeted source checking, analysis, drafting, scoring and production. The report presents the evidence, study-integrity findings and source-access limits described in Section 10, with all nine scoring inputs evaluated under the stated method. The evidence cutoff is 13 September 2026; this is not a publication date.

The publisher supplied an additional critique, *The Audit, Audited*, reporting approximately 110 checks of the review series. That document informed preparation; its author, tool identity and independent human-review status were not established in the supplied material. Its reported checks are not relabeled as independent human fact-checking or complete verification. Source and production checks do not constitute a comprehensive literature search or independent replication. The targeted study-integrity screen is described in Section 10.

No independent human subject-matter sign-off or media-law clearance is documented. Jason Hreha is not represented as having personally verified every assertion. No response from a reviewed author or publisher was solicited for this review. Source-target, arithmetic, content-preservation and rendering tests are production checks, not scientific peer review. The extent and limits of source access are stated in Section 10 and the reference records.

Relevant interests and relationships.

Jason Hreha leads The Behavioral Scientist, is the author of the commercially available behavior-change book *Real Change*, and offers behavioral-science consulting and related services. These are relevant commercial and intellectual interests, including potentially competing books, explanations or products. The publication's book page and services/contact page were checked on 8 September 2026.

Jason Hreha is named in *Hooked*'s acknowledgments. That fact alone does not establish the nature or extent of any contribution, endorsement, financial relationship or current relationship. The site also carries an earlier critical article about the Hook Model; a pre-existing editorial position is not independent corroborating evidence for this review.

The preparation record does not independently establish the presence or absence of additional relationships with the reviewed author or publisher, funding, compensation, review-copy arrangements, sponsorships or affiliate arrangements. No blanket absence-of-interests statement is made. The links included in this report are source and navigation links, not newly created affiliate links.

Affiliation and attribution.

References to individuals, organizations, institutions, publications, and trademarks identify the subjects or sources of discussion and do not, merely by their inclusion, imply sponsorship, approval, or endorsement. This is not an official Red Pen Reviews review and is not affiliated with or endorsed by Red Pen Reviews. Any use or adaptation of its publicly described approach is identified in the methodology; the ratings in this report are not ratings issued by Red Pen Reviews. Copyright in quoted or reproduced third-party material, where applicable, remains with the respective rights holders. Its inclusion in this report does not grant permission for separate reuse.

General information, not individualized advice.

This report does not provide individualized medical, psychological, legal, financial, or other professional advice, and reading it does not establish a professional-client relationship. Discussion of a recommendation's evidentiary support does not establish its suitability for any particular person or circumstance.

Corrections and substantive responses.

The publisher welcomes factual corrections, relevant additional evidence, and substantive responses from authors, publishers, researchers, and readers. Please send submissions through **The Behavioral Scientist contact form**, identifying the passage at issue, the proposed correction or clarification, and supporting sources. The publisher will assess submissions and correct substantiated factual errors. After first public publication, material corrections will be dated and explained in a correction record; changes resulting from new evidence or revised judgments will be identified as updates. A person's failure to respond to a request for comment is not treated as agreement with the review.

Publication details

Reviewer / preparation	ChatGPT-assisted critical analysis, commissioned by Jason Hreha. The extent of AI and human review is disclosed above.
Publisher and responsible editorial contact	The Behavioral Scientist / Haystack Group LLC, as identified in the website's terms. Responsible editorial contact: Jason Hreha.

Book and edition reviewed	Hooked — Nir Eyal, with Ryan Hoover. 2014 Portfolio / Penguin ebook; ISBN 9780698190665.
Literature-search cutoff	13 September 2026. Targeted critical narrative review; not a guarantee of exhaustive coverage.
Publication date	The first public publication date is recorded on the hosting page when the review is published. The evidence cutoff and source-access dates describe research, not publication.
Evidence date	13 September 2026. Dated source-access records state the actual checks and their limits. This is not a publication date.
Current evidence assessment and contact	The current evidence assessment below summarizes the basis and limits of this review. For the current public version or to submit a correction, use The Behavioral Scientist contact form and identify the book and report ID.

Current evidence assessment

Evidence basis and limits. The full official ACM methods/results confirm nearly double task quantity, not accuracy. The assessment includes favorable partial IKEA replication and bounded completion wording; favorable same-program wine evidence; later scarcity, foot-in-the-door, and habit syntheses; both sides of the dopamine interpretation exchange; app-dataset availability without claimed execution; and exact clinical source identities with bounded later evidence. The 2014 source-fidelity judgments are not penalized once per later result. The nine scoring inputs support overall and category ratings of 50%. The evidence cutoff is 13 September 2026, not a publication date. No independent human sign-off is claimed.

Report format. The reader review is accompanied by this technical appendix, with shared policy on the series pages, a claim-navigation register and a source table. The evidence qualifications and numerical scoring inputs accompany the analysis. The additional audit supplied by the publisher is acknowledged as a supplied critique, not authenticated human peer review.

Scoring method. The report uses whole-number category percentages and an equally weighted overall rating. Practical Value is assessed from the documented evidence; it is not a relabeled usability score. Scientific Accuracy and Reference Accuracy use the common criteria. Difficulty is described separately from the scored categories. The substantive claim, chapter and source audits explain the judgments; the numerical scoring inputs and arithmetic are shown in Section 09.

Source and response limits. Source-specific findings and access limits are documented in Sections 08–10. No author or publisher response is recorded in the preparation material.