BOOK EVIDENCE REVIEW

Hooked review

A book by Nir Eyal with Ryan Hoover

A checklist is not a theory of habits

The Behavioral Scientist

3,750 words / 13 min read

Evidence assessed through 13 September 2026 · Book edition and source-copy details below

OVERALL RATING

50%

An editorial assessment, not a percentage of truth.

Scientific Accuracy	**50%**
Reference Accuracy	**50%**
Practical Value	**50%**

Read it this way: A useful product-design checklist, not a validated formula for creating habits.

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

Confidence in the overall appraisal: moderate. Specific evidence limits accompany the findings.

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. Read the methodology, book-specific disclosures, and full editorial notice with the review. Corrections and substantive responses are welcome.

In the neuroscience chapter of *Hooked*, an experiment involving mice in the 1940s helps introduce the power of reward. The stimulation site is described as the nucleus accumbens.

The paper in the book's own endnote is from 1954. It concerns rats. Its title identifies the septal area and other regions of the rat brain. The year, animals, and anatomical account need correction. [1]

That does not erase the underlying finding that brain stimulation can reinforce behavior. But it is a revealing place to begin a book review. Neuroscience is doing a lot of persuasive work in *Hooked*, and some of that work depends on claims the cited experiments did not establish.

The larger question is more important than the historical error: does Nir Eyal's four-stage model actually explain how to create a habit-forming product?

My verdict: it is more defensible as a design checklist than as a validated causal recipe. Triggers, actions, rewards, and investments are useful things to think about. Their arrangement in a memorable loop does not prove that the complete sequence is necessary, sufficient, or superior to a well-designed alternative.

A product team can benefit from the questions while remaining skeptical of the theory. In fact, that is the reading I would recommend. Use the framework to generate tests. Do not use the framework as the result of the tests you have not yet run.

What the model is trying to do

The Hook Model proposes a loop: **trigger, action, variable reward, investment**.

An external cue initially brings someone into the product. An easy action delivers something rewarding. Some uncertainty keeps the experience interesting. The user then contributes something—data, effort, content, reputation, or relationships—that makes a return more likely. Repeated cycles are supposed to build an association between a recurring need and the product. [2]

There is a sensible design vocabulary here. What brings someone back? What makes the next action difficult? What does the person receive? Does use improve the next experience? A team unable to answer those questions probably has important work to do.

Eyal also adds qualifications that should survive criticism. Variable rewards are not a cure for a product nobody wants. Excessive investment can put people off. Privacy and deceptive invitations matter. Product testing is necessary. There is an ethics chapter, not simply an instruction to maximize engagement at any cost. [2]

The strongest reading is therefore a practical one: make a recurring useful experience easier to discover, perform, and repeat.

The trouble comes when the vocabulary is treated as an explanation that has already won. A product can fit the four boxes after it succeeds without demonstrating that those boxes caused the success. A good checklist can organize your attention. A tested theory has to survive a more demanding comparison.

Three claims should stay separate

Repeated actions in recurring contexts can become more automatic. This foundation receives a 75% rating. It has meaningful support, though translating it into a particular app, frequency threshold, or retention metric requires additional work.

Variable rewards have a privileged role in repeat engagement and habit formation. This receives 50%. Reward schedules and uncertainty can matter, but that is weaker than a general instruction that variability is necessary or best.

Investment and the prescribed four-stage sequence reliably create durable, unprompted engagement. This receives 25%. Several component ideas are plausible, yet the complete package is not established by the evidence examined in the review. [3]

Those ratings produce a Scientific Accuracy score of 50%. They preserve the legitimate habit research without treating it as blanket validation of the branded loop.

A reader should keep one additional distinction in view: engagement, automaticity, and benefit are separate outcomes. A person can repeatedly use a valuable product deliberately. A person can use something automatically and regret doing so. A product can retain users because leaving is inconvenient.

A useful explanation must tell these possibilities apart, because they imply different design decisions.

Brain activity is not a dopamine measurement

The most important neuroscience correction is not the wrong decade or the wrong animal. It is the claim that researchers observed increased dopamine levels when the cited experiments measured an fMRI signal.

Berns and colleagues manipulated predictable and unpredictable sequences of juice and water and measured blood-oxygen-level-dependent activity, usually abbreviated BOLD. The study did not independently measure dopamine. The paper itself cautioned about the relationship between the imaging signal and dopaminergic transmission. [4]

You can describe activity in reward-related brain regions without claiming that a neurotransmitter was directly observed. Substituting “dopamine levels” for the actual measurement makes the account sound more mechanistically specific than the experiment allows.

Another cited paper, by Aharon and colleagues, examined the reward value of face categories through behavioral and imaging measures. Its title's phrase "variable reward value" refers to differences between stimuli. It does not mean the researchers demonstrated that randomly withholding rewards creates product habits. [5]

This is why neuroscience language deserves the same questions as any other evidence. What was manipulated? What was measured? What comparison supports the interpretation? A brain image does not get an exemption because it looks closer to the machinery of behavior.

There is genuine research directly recording dopamine-neuron responses while manipulating reward probability. The appendix distinguishes that work from the imaging studies. The point is to assign each study the claim it can actually support. [6]

Even there, the interpretation has been debated: a computational critique offered an alternative account of the uncertainty-related signal, and the original researchers replied with evidence they considered inconsistent with that account. This was an argument about how to interpret actual neural recordings, not an empirical replication failure. It provides no shortcut from dopamine to a proven product strategy. [6]

For a designer, the difference is practical. A neural response to uncertainty may motivate a hypothesis about an interface. It cannot tell you that the interface will create durable automaticity, improve the user's life, or outperform a more predictable experience. Those are product questions that still need product evidence.

Repetition has a foundation; visit counts are a shortcut

The broad idea that repetition in recurring contexts can strengthen automatic responding is the strongest part of the model.

The everyday habit-formation work and research on linking flossing to toothbrushing are relevant. Work on smartphone checking is more directly connected to digital behavior. These studies give the model a legitimate foundation, while testing much narrower questions than the four-stage package. [7]

Eyal also deserves credit for rejecting a universal habit-formation timetable. A fixed number of days makes a good headline and a poor general description of the learning process.

A later review of health-behavior studies reinforces that caution. Habit strength generally improved, but formation times varied and only a few studies estimated how long the process took. Much of the evidence also carried a high risk of bias. That supports keeping repetition and context in the explanation; it does not establish a new timetable or validate the four-stage loop. [7]

The measurement problem is that a frequent action need not be automatic. People can intentionally check a useful service often. They can also keep using a product because colleagues are there, because it has the best information, or because transferring their accumulated work would be painful.

Imagine someone who opens a project-management app every morning because it contains the team's assignments. Their behavior is repeated. Perhaps parts of it become automatic. But the app's practical value and the team's coordination requirements are already substantial explanations. Counting the opens does not tell you how much each mechanism contributes.

The same caution applies to internal triggers. Becoming less dependent on a company's marketing messages is different from becoming independent of external cues. A familiar place, the end of a meeting, or a preceding action can continue to prompt a learned response. There is no requirement that every lasting product habit become an invisible emotion. [8]

The useful product question is specific: which action becomes easier to initiate in which context? "Our retention improved" may be good news. It is not yet a complete answer about habit formation.

Randomness is not a substitute for value

The phrase *variable reward* bundles together several different experiences. Uncertainty about whether anything arrives, uncertainty about how much arrives, changing content, task difficulty, and another person's response are not the same manipulation. [9]

That matters because the evidence for one cannot automatically validate the rest. A person might revisit a weather service because the weather changes and the information is useful. The mere fact that the content varies does not identify unpredictability as the independent reason they return.

The book recognizes that rewards must fit a need. That qualification is important. The stronger language about craving and the special power of variability still asks for evidence that compares a variable experience with a useful predictable one.

One small randomized health-app study did make a relevant comparison. Over six weeks, 61 participants received different fixed or variable reward arrangements. It did not detect superior engagement under the variable schedules; one arrangement had lower costs. The study is too small and specific to settle the general question, but it illustrates why the claim should be tested rather than assumed. [10]

There is also an instructive counterpoint inside *Hooked*: the car-wash loyalty study. Both groups needed eight more washes, but one card displayed an apparent head start toward a larger total. Completion improved under a predictable reward contingency. The example is useful evidence about goal pursuit, even though it does not measure automatic habit. [11]

A design team should therefore ask what uncertainty adds. Does it improve the experience? Does it make an important action more likely? Is it merely making the result less dependable?

The word "reward" cannot answer those questions for the user. Neither can a diagram. Start with something worthwhile to receive, then test whether varying its delivery improves the outcomes that matter.

Investment can add value—or just make leaving harder

"Investment" is the most intuitively useful and scientifically slippery part of the model.

A product may improve as a person uses it. Preferences can make recommendations more relevant. Saved work can reduce future effort. Contributions can prompt useful responses from other people. These are good reasons to examine what users accumulate over time. [12]

But accumulation can produce several mechanisms. The service may become genuinely better. Moving elsewhere may become more costly. Users may value something more because they worked on it. A cue may become associated with an action. These explanations do not recommend the same product strategy.

The original IKEA-effect experiments found increased valuation of successfully self-made objects under particular conditions. Unfinished or dismantled creations did not show the same advantage in those experiments. A later independent study using loom-band objects also supported increased valuation from making something, but did not reproduce every proposed mechanism or completion contrast. The useful lesson is conditional, not simply "make people do more work so they will love the result." [13]

The computer-reciprocity example offers a useful finding, but it answers a different question. The full 1997 ACM report shows that participants completed nearly twice as many color-comparison tasks for the computer that had previously helped them as for a different computer. That measures how much help they provided, not a doubling of accuracy. It supports a narrow reciprocity effect; it does not compare investment before a reward with investment after a reward in a habit-forming product. [14]

Imagine two services that accumulate the same useful user data. One makes that information easy to export; the other makes leaving difficult. Both might show retention. The mechanisms and implications differ. A metric that treats them as the same success can reward friction rather than value.

I would use the investment idea to ask how each interaction improves the next one. Then check whether users experience that improvement. The strongest form of investment is something the person benefits from keeping, not merely something they would suffer to lose.

That is a design judgment about user value, not a result proven by the origami experiments. Keeping those levels separate makes the advice more useful.

A successful product is not a controlled test of the loop

A 2023 paper explicitly applied the Hook Model to Uber and Instagram. It analyzed features and use cases. That shows the framework can describe products, but the study did not randomly assign alternative designs and establish the model's incremental effect. [15]

The difference is straightforward. Once a product succeeds, it is often possible to find cues, actions, rewards, and accumulated value somewhere in the experience. The harder test is whether specifying the model beforehand helps a team build a better product than it would build using a credible alternative.

A complete test would need to identify what counts as each stage, what would count as failure, which outcomes matter, and how the comparison avoids quietly giving one condition more utility or attention. It would also need to examine the proposed sequence rather than merely finding that a product containing rewards gets used.

The Bible App case is useful in the descriptive sense. Portability, shorter reading tasks, audio, reminders, and saved annotations are plausible design lessons. The book also acknowledges religious communities, leaders, and timing, which are additional explanations for growth. [16]

A curious statistic in that chapter illustrates the need for restraint. The book says sixty-six thousand people open the app every second. Taken literally, that implies more than 5.7 billion opens per day. The review could not recover the underlying interview data or intended unit, so the figure remains unverified. It would be wrong to silently replace it with a guess about what the author probably meant. [16]

The practical response is to keep the inspectable design observations and reserve judgment about the unexplained number. Success stories can generate hypotheses. They cannot provide the missing counterfactual simply by being impressive.

“Twice as likely” is doing too much work

The book describes the “but you are free” technique: make a request while explicitly reminding the person that they can refuse. The historical synthesis behind the example reported an odds ratio of about 2.03, not a universal doubling of the probability of agreement. [17]

The distinction changes the result. With a starting probability of 50%, doubling the odds produces a probability of about 67%, not 100%. The baseline has to travel with the number.

The loose wording also appears in the source, so it would be unfair to portray Eyal as having independently invented the ambiguity. He also acknowledges the lack of direct product testing. Those are relevant qualifications.

Later evidence warrants more caution. A preregistered re-examination found a positive pooled effect, but the seven studies rated at low risk of bias produced a much smaller, uncertain estimate. That weakens a confident promise that the technique reliably delivers a large gain. It is later evidence, not a reason the original book should have known a future result. [18]

There is a useful distinction between the ethical and tactical reasons for the phrase. Giving a person meaningful freedom to refuse can be appropriate regardless of whether mentioning that freedom increases compliance. If the freedom is merely verbal while refusal is made costly, the product has

not solved the autonomy problem.

A designer should therefore ask two questions: does this wording change behavior, and does the actual choice remain free? A positive answer to the first cannot substitute for the second. The framework becomes stronger when respect for users is a requirement, not another technique justified only by its conversion rate.

The ethics chapter uses the wrong denominator

Eyal's ethics discussion deserves credit for existing and for taking the question of manipulation seriously. But a reassuring statistic in that discussion carries a major limitation.

The approximately 1% gambling-prevalence figure recovered from the cited source refers to the adult population. It is not an estimate for all slot-machine users, much less a risk estimate for people using any habit-forming technology. The book moves toward a broader reassurance about an "addicted 1 percent" that the denominator cannot support. [19]

A population statistic and a statistic among users answer different questions. Applying the first to a selected group can change the conclusion substantially. Extending it across products adds another unsupported step.

The review does not supply a replacement prevalence figure or claim to estimate the harm caused by the book's techniques. The criticism is narrower: this statistic cannot carry that reassurance.

There is also a practical limitation to judging a product mainly through a designer's intentions and willingness to use it. Designers and users may have different needs, vulnerabilities, and incentives. A sincere belief in the product's benefit is worth less than evidence about the experience it creates.

For a team, the better questions concern user outcomes and control. Can people understand the relevant choice? Can they stop reminders? Can they leave? Does the product help them accomplish what brought them there? What happens to people for whom the product becomes costly?

These questions should sit alongside engagement metrics, not underneath them. More use may be good, bad, or incidental. The judgment depends on what the product is doing for the person.

Several examples are worth keeping

A critical review should distinguish a flawed explanation from a wholly worthless example.

The car-wash head-start study is reasonably faithful at its core. The habit-formation timetable discussion correctly rejects a single universal duration. Parts of the model's treatment of action simplicity represent the cited framework fairly. These strengths matter to the overall assessment. [20] [21] [11]

Other examples retain a narrower lesson after correction. The wine-pricing experiment supports an effect of price information on reported pleasantness and related activity, but the book misdescribes parts of the procedure. The self-disclosure paper supports a rewarding aspect of disclosure, but it was original experimental research, not the "meta-analysis" described in the book. [22][23]

Later wine research also found that price information could change reported pleasantness. That is favorable evidence for the narrower effect, from the same research program rather than a fully independent replication. It neither repairs the book's description of the original procedure nor establishes why people keep returning to an app. [22]

That pattern is more instructive than an indiscriminate verdict of "all wrong." A real finding can survive while its method, mechanism, or application needs revision.

It also suggests a productive reading habit. When a story feels useful, ask which part is doing the useful work. Perhaps it is the practical observation that a task is easier. Perhaps it is evidence that an apparent head start encourages completion. The more elaborate neurological explanation may be unnecessary—and less well supported.

A product team does not need to wait for a perfect theory before trying a sensible improvement. It does need to stop treating every successful tweak as confirmation of the whole theory.

How I would use the book

Turn the four boxes into questions, not requirements.

Trigger: what situation makes the product relevant, and what cue helps someone notice it?

Action: what worthwhile task is the person trying to complete, and what avoidable difficulty gets in the way?

Reward: what useful or enjoyable result does the person actually receive? Would a predictable result serve them better?

Investment: does their participation improve the next experience, or merely make departure more expensive?

Then choose an outcome and a comparison. The comparison need not be an elaborate academic study to be informative. It must be credible enough to distinguish the improvement from the other changes occurring around it.

An illustrative product test might compare a proposed reminder with the existing version, measure completion of the intended task, and also examine opt-outs or complaints. If the theory concerns automaticity, include an appropriate measure rather than quietly substituting visits. If the objective is user benefit, measure something connected to that benefit.

Keep: the questions about relevance, simplicity, recurring context, and accumulated usefulness, plus the invitation to test.

Modify: variable-reward prescriptions, investment demands, and the assumption that retention reveals habit.

Reject: the inaccurate neuroscience descriptions, the unsupported risk reassurance, and any claim that the diagram itself proves the order of operations. [24]

This is an editorial translation of the book into a testing approach, not a validated substitute framework. Its purpose is to keep product decisions tied to outcomes instead of making the team defend a favored diagram.

The best result is not that every product perfectly fits the loop. It is that each important design choice earns its place.

Make the behavior precise before measuring success

Even a well-run comparison can answer the wrong question when the target is vague. "People used the app more" leaves several possibilities open: more people returned, the same people opened it more often, sessions became longer, or users needed extra attempts to finish the task.

Consider an illustrative scheduling service. A redesign that helps someone book an appointment in one visit might reduce session count while making the product better. A confusing redesign might produce more repeat visits because people cannot tell whether the booking worked. Engagement alone would give those outcomes the wrong ordering.

This is a logical example, not a claim about a tested product. It shows why the outcome should follow from the person's goal. For the scheduling service, successful bookings, avoidable effort, and the need to correct mistakes would be more informative than raw opens.

A habit-related metric requires the same care. Define the action and the opportunities to perform it. Then decide what evidence would distinguish automatic responding from deliberate use. The complete appendix discusses those measurement limits. [7]

Getting this right is more important than making every observed behavior fit the terminology of the loop. A framework should make the product easier to understand, not make inconvenient outcomes easier to relabel.

That is the purpose of a useful model.

Why the rating is 50%

Scientific Accuracy: 50%. The cue-and-repetition foundation is stronger than the claims about variability and the complete sequence.

Reference Accuracy: 50%. The source audit is genuinely mixed. Several illustrations are useful; others misidentify methods, measurements, or what the result can support.

Practical Value: 50%. The checklist can help teams inspect an experience. Its incremental effect as a complete method, its transfer across products, and its relationship with user benefit remain uncertain. [25]

The overall rating gives those categories equal weight. It is an appraisal under the published rubric, not a finding that half the book's sentences are false.

Hooked is useful when it gets a team asking better questions. It becomes less useful when the team treats its vocabulary as evidence that the questions have already been answered.

Build something worth returning to. Then study why people return.

> **Want to inspect the evidence?** The technical appendix contains the full claim register, chapter assessments, source audit, score calculations, limitations, and evidence assessment. The reader essay and appendix make the same numerical judgments; they serve different reading needs.

Evidence notes

Each note links to the relevant passage of the technical appendix, which identifies the relevant book passages, original research, and source-access limitations. The evidence assessment includes relevant replication findings, syntheses, corrections, and access checks. These notes do not certify that every underlying dataset was reproduced.

1. A08 · Three factual errors in the foundational brain-stimulation story. Detailed assessment, source references, and access limits in the technical appendix.
2. The Hook Model, faithfully reconstructed. Detailed assessment, source references, and access limits in the technical appendix.
3. Three central scientific claims. Detailed assessment, source references, and access limits in the technical appendix.
4. A10 · Berns measured an fMRI signal, not dopamine release. Detailed assessment, source references, and access limits in the technical appendix.
5. A11 · A title's word "variable" is not evidence of variable-ratio reinforcement. Detailed assessment, source references, and access limits in the technical appendix.
6. · 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. Detailed assessment, source references, and access limits in the technical appendix. Includes the computational critique and the original authors' reply; neither is an empirical failed replication.
7. Claim 1 · Frequent, cue-linked repetition can make product-related actions increasingly automatic. Detailed assessment, source references, and access limits in the technical appendix. Includes the 2024 health-behavior synthesis and its bias and duration limits.
8. A03 · Habit cues do not have to become invisible emotions. Detailed assessment, source references, and access limits in the technical appendix.

9. Claim 2 · Variable rewards are a privileged driver of repeat engagement and habit formation. Detailed assessment, source references, and access limits in the technical appendix.
10. · 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. Detailed assessment, source references, and access limits in the technical appendix.
11. A07 · Endowed progress is a reasonably faithful field example. Detailed assessment, source references, and access limits in the technical appendix.
12. Claim 3 · Investment and the prescribed four-stage sequence produce durable, unprompted engagement. Detailed assessment, source references, and access limits in the technical appendix.
13. A14 · Labor can increase valuation; completion is a tested boundary. Detailed assessment, source references, and access limits in the technical appendix. Includes the favorable independent conceptual replication and its mixed mechanism/completion findings.
14. A16 · Computer reciprocity cannot establish a mandatory order of stages. Detailed assessment, source references, and access limits in the technical appendix. Based on the full official ACM report; near-double task quantity is not near-double accuracy.
15. · 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). Detailed assessment, source references, and access limits in the technical appendix.
16. Chapter 7 · Case Study: The Bible App. Detailed assessment, source references, and access limits in the technical appendix.
17. A13 · The original “but you are free” finding was reported too loosely. Detailed assessment, source references, and access limits in the technical appendix.
18. · 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. Detailed assessment, source references, and access limits in the technical appendix.
19. A17 · The denominator changes the ethical conclusion. Detailed assessment, source references, and access limits in the technical appendix.
20. A01 · A formation timetable that varies rather than a universal rule. Detailed assessment, source references, and access limits in the technical appendix.
21. A04 · Fogg’s model is represented substantially faithfully, but “formula” overpromises. Detailed assessment, source references, and access limits in the technical appendix.
22. A06 · The wine effect survives; the account of the procedure does not. Detailed assessment, source references, and access limits in the technical appendix. Includes a favorable 2017 follow-up from the same research program.
23. A18 · The self-disclosure source was not a meta-analysis. Detailed assessment, source references, and access limits in the technical appendix.
24. Practical use, limitations, and safeguards. Detailed assessment, source references, and access limits in the technical appendix.
25. How the ratings were determined. Detailed assessment, source references, and access limits in the technical appendix.

Disclosures and corrections

How we research and score these reviews.

Relevant interests: Hreha wrote the commercially available behavior-change book *Real Change* and offers behavioral-science services. These are potentially competing commercial and intellectual interests. Additional relationships, funding, review copies, sponsorships, and affiliate arrangements have not been comprehensively established; no blanket absence-of-interests claim is made.

Additional interest: Jason Hreha is named in *Hooked*'s acknowledgments, and the site has previously published criticism of the Hook Model. The acknowledgment alone does not establish the extent of a contribution or relationship. Full disclosure.

Publisher: The Behavioral Scientist / Haystack Group LLC, identified in the site's terms. Editorial contact: Jason Hreha. This is not an official Red Pen Reviews review and is not affiliated with or endorsed by Red Pen Reviews.

Corrections: Send the passage, proposed correction, and sources through the contact form. Substantiated factual errors will be corrected; material changes after first publication will be dated. See the book-specific evidence assessment and read the full editorial notice for scope, score interpretation, professional-advice limitations, attribution, and correction policy.

Book reviewed: Hooked, Nir Eyal, with Ryan Hoover. 2014 Portfolio / Penguin ebook; ISBN 9780698190665. Evidence assessed through 13 September 2026.

Evidence assessment: This review includes the study-integrity assessment described in the technical appendix. The nine scoring inputs follow the published rubric.