

One Payment, One Path: Pricing and Structural Curation as a Joint Commitment Device in Self-Directed Language Learning

ANONYMOUS AUTHOR(S)

Self-directed language learners typically juggle several tools, and nearly every popular language app runs on a recurring subscription, a structure commitment device research suggests may work against sustained effort. We report a two to four week single-condition pilot (N=20) of Germanly, a German learning web app combining a structured A1 to B2 curriculum with a one time payment and no auto-renewal. Pre and post measures show a rise in confidence and motivation, real tool consolidation, and a gap between predicted and experienced pricing beliefs: participants expected subscriptions to sustain engagement better, but afterward most judged a subscription would have reduced their use. Psychological ownership, not sunk cost guilt, correlated with continuation intent. Because pricing and curriculum are bundled by design, this pilot cannot separate their causal contribution; we treat it as hypothesis generating and outline the factorial study needed to test pricing as an interaction design lever.

CCS Concepts: • **Human-centered computing** → **Empirical studies in HCI**; **User studies**; *HCI theory, concepts and models*; *Interaction design*; *Interaction design theory, concepts and paradigms*; • **Applied computing** → Psychology.

Additional Key Words and Phrases: self-directed learning, language learning apps, commitment devices, sunk cost, psychological ownership, pricing design, mixed methods

ACM Reference Format:

Anonymous Author(s). 2027. One Payment, One Path: Pricing and Structural Curation as a Joint Commitment Device in Self-Directed Language Learning. In *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems (CHI '27)*. ACM, New York, NY, USA, 15 pages.

1 Introduction

Open the app store and search for a language learning app and you will find dozens of near identical entries, most of them asking for a monthly or annual subscription before you have finished the onboarding screen. Duolingo, Babbel, Busuu, and Memrise, four of the most downloaded apps in the category, all default to recurring billing. The pitch behind that default is familiar from the wider subscription economy: recurring revenue funds ongoing content, ongoing content justifies the recurring charge, and the charge itself is supposed to keep the user coming back so the value does not go to waste. Whether that last step actually holds for an individual learner, rather than for the platform's revenue line, has had surprisingly little direct attention in either the CALL literature or the CHI literature on commitment devices.

We came to this question through a specific product. Germanly is a web app for serious German learners, not a native app, built by one of this paper's authors, that departs from the subscription default in two connected ways. First, it charges once, with no card kept on file and no auto-renewal, for full access to a structured curriculum spanning the A1 to B2 levels. Second, that curriculum is deliberately singular: vocabulary, grammar, listening, and writing feedback live inside one continuous path rather than being split across separate apps, tabs, or subscriptions the way a self-directed learner typically assembles their own study routine. Germanly's own marketing describes this as "one payment, one path," a phrase that turns out to name two design decisions that the literature has mostly treated as unrelated.

The commitment device literature in CHI, most notably Lee and colleagues' large scale study of financial and social commitments on the stickK platform [10], establishes that binding yourself to a cost you cannot recover changes the probability that you follow through on a goal. The behavioral economics literature on sunk cost, going back to Arkes and Blumer's original 1985 demonstration using season theater tickets, shows the same mechanism in a much older and simpler form: people who have already paid for something keep showing up to avoid feeling like they wasted the money, regardless of whether that reasoning is actually rational. Separately, a growing body of self-directed learning research, including a 2026 CHI paper on how self-directed language learners navigate fragmented resources and delegate parts of their learning to AI tools [2], and a recent system called YT-Pilot that turns fragmented YouTube tutorials into a single coherent pathway [1], establishes that structural fragmentation itself is a burden independent of motivation, one that shows up as lost track of progress, forgotten logins, and the small but real tax of deciding what to open next.

What nobody has connected is the possibility that a one time payment and a single structured path are not two separate interventions but one. A subscription, almost by construction, removes the sunk cost signal that a one time payment provides: the money is never fully spent, it just keeps being spent a little at a time, and Kahneman, Knetsch and Thaler's endowment effect research suggests that a recurring, revocable charge is unlikely to produce the same sense that something is now yours, in the way a single completed purchase can [8]. If pricing structure and content structure are pulling in the same direction inside Germanly, an app that is priced like something you own and structured like a single path rather than a menu, then the two decisions may work together in ways that neither a pure efficacy comparison against Duolingo and Babbel nor a pure commitment device study divorced from language learning would surface. Testing that possibility properly requires unbundling pricing from curriculum, which this pilot's single-condition design does not do; we return to what that unbundled study would need to look like in Section 8.

This paper reports a pilot study designed to test that possibility directly, rather than to add another entry to the already crowded literature comparing language app efficacy. We recruited 20 adult self-directed German learners, a deliberately varied group spanning students, professionals, a retiree, and a diplomat's spouse, and followed them through a two to four week period of Germanly use with a pre and post pilot questionnaire and matched semi-structured interviews before and after. We ask three questions. First, does using Germanly change learners' confidence and the degree to which their motivation feels self-directed rather than obligated. Second, do learners actually consolidate their tool use, and does that consolidation track with the qualitative fragmentation burden described in prior self-directed learning work. Third, and most centrally, how does the one time payment actually function psychologically once learners have lived with it, compared to how they predicted it would function beforehand.

Our central contribution is not a claim that Germanly teaches German better than its competitors, a question this pilot was not designed or powered to answer. Concretely, this paper contributes: (1) empirical evidence, from a mixed methods pilot, of a systematic gap between how self-directed learners predict a pricing model will affect them and how it actually affects them once lived with; (2) a reframing of pricing model and content curation as a single, joint interaction design lever rather than two independent product decisions, extending commitment device theory into a domain, consumer software pricing, that HCI has largely left to product teams; (3) evidence that psychological ownership, not sunk cost guilt in isolation, is the mechanism that predicts sustained engagement, which redirects the design target for anyone building

a commitment device; and (4) a boundary condition, drawn from disconfirming cases in our own data, showing where this design lever loses force for learners who already have an external commitment structure. Together these findings argue that pricing model deserves the same design scrutiny HCI already gives to onboarding, feedback, and interaction flow, rather than being treated as a business decision downstream of design. We close by discussing what this means for language learning apps, and self-directed learning tools more broadly.

2 Related Work

2.1 Effectiveness research on language learning apps

The commercial language app landscape has been studied mostly through head to head efficacy comparisons. Kessler, Loewen and Gönülal compared Babbel and Duolingo learners over eight weeks and found no significant difference in overall L2 learning gains between the two apps, though Babbel users showed a stronger relationship between study time and post test scores, and reported the app felt more effective for grammar and pronunciation [9]. A 2023 comparative analysis of the same efficacy literature concluded that Busuu’s published studies showed the strongest results for reading, grammar, and oral proficiency, with Duolingo close behind on receptive skills and Babbel trailing due to shallow proficiency gains despite a longer study period [13].

At the level of mobile assisted language learning as a category, the evidence is broadly favorable but methodologically uneven. Sung, Chang and Yang’s early meta-analysis across 1993 to 2013 found only a small and statistically marginal effect of mobile devices on language acquisition [18], while a more recent meta-analysis restricted specifically to purpose built MALL applications, rather than general mobile technology use, found a moderate to strong benefit over traditional classroom approaches ($g = 0.88$) [12]. A 2024 systematic review of collaborative MALL practices, following PRISMA guidelines across 72 studies, found publication volume accelerating sharply after 2019 and a strong concentration of research in higher education settings in Asia [6], and a parallel systematic review of commercially available foreign language apps specifically flagged persistent weaknesses in research design and statistical rigor across the category [19].

Two things stand out from this literature for our purposes. First, every one of these comparisons treats the apps as fixed bundles of features, holding pricing model constant as an unremarked background fact, usually a subscription, rather than as a variable worth manipulating or even mentioning. Second, the comparisons are almost entirely outcome focused, asking which app teaches more language fastest, rather than process focused, asking why a learner keeps opening one app and not another over a period of weeks. Our study sits in the second gap rather than the first: we are not claiming Germanly teaches German better than Duolingo or Babbel, a claim our sample size and design could not support, but asking a different and comparatively unstudied question about what keeps a self-directed learner opening any app at all.

A closely related thread concerns AI generated feedback specifically, since Germanly’s writing feedback component sits inside this literature. A systematic review of AI based automated written feedback research identified 83 studies published between 1993 and 2022 and found generally positive effects on writing quality, alongside a consistent caution that automated feedback tends to address surface level accuracy well while underserving higher order concerns like organization and coherence [17]. More recent work specifically on large language model generated feedback, including ChatGPT based corrective feedback, extends this

picture: a mixed methods study of L2 writers engaging with ChatGPT as an automated feedback provider found engagement concentrated on surface level, sentence based revisions rather than deeper structural or organizational changes [20]. This matters for how we interpret our own participants' complaints, several of whom wanted feedback closer to what a human tutor or native speaker conversation partner provides, a limitation consistent with what the AI feedback literature would predict rather than a Germanly specific shortfall.

2.2 Self-directed learning and the burden of fragmentation

Self-directed learning, in Garrison's comprehensive model, describes a process in which the learner takes on contextual control, cognitive responsibility, and motivational ownership of their own learning, without requiring an instructor to do any of that work for them [5]. That autonomy is exactly what makes self-directed language learning attractive to adult learners who cannot commit to a fixed class schedule, and exactly what produces the fragmentation problem our participants described so consistently before starting Germanly: a self-directed learner is, by definition, assembling their own curriculum out of whatever tools they can find, and nothing forces those tools to work together.

Recent CHI research has begun to name this fragmentation directly rather than treating it as an incidental detail of self-directed learning. A 2026 CHI paper on how self-directed language learners navigate delegation decisions when working with large language models describes learners managing five distinct specialized support functions, planning, input practice, evaluation, and more, and explicitly frames the challenge as one of coordination across tools rather than of any single tool's quality [2]. YT-Pilot, a system published the same year, targets an adjacent problem directly: YouTube is an enormous and largely free resource for informal learning, but learners studying across multiple videos have no persistent structure connecting what they watched last week to what they are watching today, and the system's authors report that introducing a single persistent pathway significantly improved learners' sense of goal clarity and progress tracking in a within-subjects study [1]. Neither paper is about language learning specifically in the sense our study is, but both converge on the same underlying claim: structure that persists across sessions and across content types is not a convenience feature, it measurably changes how learners experience their own progress.

This connects to an older and more general finding in organizational psychology. Leroy's attention residue research shows that switching from one unfinished task to another leaves cognitive traces of the first task active in working memory, degrading performance on the second task in a way that plain effort or willpower does not resolve [11]. Our participants described almost exactly this experience in their own words before starting Germanly, several independently using language like staring at their phone trying to remember which app they had left a lesson open in, or simply forgetting an app existed because it was not the one they happened to open that week. What the self-directed learning and attention residue literatures together predict, and what our study tests directly for the first time in a language learning setting, is that consolidating a fragmented set of tools into a single structured path should reduce this switching burden independent of any change in the learner's underlying motivation to study German at all.

2.3 Commitment devices, sunk cost, and psychological ownership

A commitment device is, in Lee and colleagues' framing from their large scale CHI 2021 study of the stickK platform, an attempt to bind yourself now to a cost that makes abandoning a goal later more painful,

precisely so that your future self cannot talk your present self out of the goal [10]. Their analysis of over 75,000 real world goals found that financial stakes were strongly associated with successful goal achievement, though used far less often than they hypothesized commitment devices are able to help, a design gap between what works and what people actually choose to use.

The sunk cost effect, first demonstrated experimentally by Arkes and Blumer, is the psychological mechanism most people would point to as the explanation for why a financial commitment device works at all: having already invested money, time, or effort into something increases the perceived probability that you will continue, even though economically rational decision making should ignore costs that cannot be recovered [3]. Arkes and Blumer’s original field study, notably, used season theater ticket holders, finding that people who paid more for their season tickets attended more plays over the following six months than people who paid less, which is a close structural analogy to a one time language app payment: a single upfront cost, followed by a long series of small, individually reversible decisions about whether to keep using what was bought.

Sunk cost on its own, however, is a somewhat cynical explanation, framing continued engagement as a kind of psychological trap rather than something a learner might genuinely value. Pierce, Kostova and Dirks’ theory of psychological ownership offers a more constructive complementary mechanism. They define psychological ownership as the state of feeling that a target, which can be a physical object, an idea, or in our case a piece of software, is genuinely one’s own, an extension of the self, and they trace this feeling to three roots: a sense of control over the target, deep knowledge of the target built through investment, and the target coming to serve as a place the self can dwell [14, 15]. Kahneman, Knetsch and Thaler’s endowment effect experiments, most famously their coffee mug studies at Cornell, show a closely related and much replicated phenomenon: simply owning something, even for a few minutes, roughly doubles how much a person demands to give it up compared to what they would pay to acquire it in the first place, a finding the authors trace to loss aversion rather than to any change in the object itself [8]. A subscription, we suggest, structurally interferes with both mechanisms. It never produces a single completed transaction to feel ownership over, and because it can always be canceled with no sunk investment beyond whatever was already billed, it does not generate the same escalating psychological stake that a completed purchase does.

Self-determination theory gives us a way to describe what happens to a learner’s motivation once a commitment device is in place, beyond simply whether they keep opening the app. Ryan and Deci distinguish intrinsic motivation, doing something because it is inherently enjoyable, from various forms of extrinsic motivation that range from fully externally controlled to substantially internalized, and their research consistently finds that autonomy supportive conditions predict movement along that continuum toward more self-determined, better sustained forms of motivation [16]. Several of our participants describe exactly this kind of shift over the course of the pilot, moving from language that frames studying as an obligation to language that frames it as something they look forward to, which SDT would predict is a more durable foundation for continued learning than guilt or sunk cost alone, even if sunk cost is part of what got them to keep showing up long enough for that shift to happen.

2.4 Pricing model as an unstudied design lever

Outside of academic research, the question of subscription versus one time pricing is a live and heavily discussed one in app design and product writing, almost entirely from a revenue perspective rather than a user psychology perspective. Marketing research on subscription programs more broadly finds that recurring billing structures reshape customer purchase behavior over time in ways a one time transaction does not, though this literature has focused on retail and media subscriptions rather than software pricing specifically [7]. That framing treats pricing model as a business decision downstream of what the product already does, rather than as something that might shape how a user relates to the product in the first place.

This is precisely the gap our study addresses. The commitment device literature we have just reviewed gives strong theoretical reason to expect that pricing structure changes user psychology, not just business economics, yet none of the CHI or CALL papers we found that study behavior change apps or language apps directly manipulate or even discuss pricing model as an independent variable. Germanly's own positioning, one payment, no auto-renewal, no card kept on file, sitting inside a market where Duolingo, Babbel, Busuu, and Memrise all default to recurring billing, and DIY tools like Anki and LingQ are either free or priced as one time software purchases but without a structured curriculum behind them, gives us a naturally occurring design case to study this gap directly, which is what the rest of this paper does.

3 Germanly as a Design Case

Germanly is a web app, not a native app, built by Pexilim and aimed specifically at serious adult learners rather than casual, gamified use, structured as a single continuous curriculum from complete beginner through B2. It runs entirely through the browser, on desktop or phone, with no app store download required, and on a phone it can be added to the home screen and opened like an installed app directly from the browser, so a learner gets an app like experience without ever leaving the web. Unlike Duolingo's largely nonlinear exercise bank or Babbel's topic based lesson structure, Germanly presents learners with one ordered path covering vocabulary, grammar explanation, listening exercises, and short form writing tasks with AI generated corrective feedback, all inside a single interface rather than split across separate modes or upsells.

The pricing model is the second half of the design case and the one this paper is centrally concerned with. Germanly charges a single upfront fee for full access to the A1 through B2 curriculum, with no recurring billing, no card retained on file after purchase, and no auto-renewal to cancel. This stands in explicit contrast to Duolingo, Babbel, Busuu, and Memrise, all of which default new users into a recurring subscription, typically with a free tier or trial period designed to convert into that subscription. It also stands in contrast to DIY tools like Anki, a free flashcard system with no built in curriculum, and LingQ, which is priced per month for a large but unstructured library of native content. Germanly's own public positioning describes this combination as "one payment, one path," explicitly tying the pricing decision to the structural decision as a single pitch rather than two separate features, which is the premise this study set out to test empirically rather than take at face value.

Table 1. Participant demographics ($n=20$).

ID	Age	Origin	Background
P01	29	Poland	Graphic designer, Cologne; moved for partner
P02	34	India	Software engineer, Munich; relocated for work
P03	41	Syria	Resettled; integration course requirement
P04	23	USA	Exchange student, Berlin
P05	26	Japan	Design professional
P06	63	UK	Retiree, Bavaria
P07	22	Nigeria	University student
P08	31	Australia	Relocated; partner is German
P09	27	Brazil	Au pair
P10	38	Morocco	Blue card job seeker
P11	30	Italy	PhD student
P12	25	UK	Gamer
P13	28	UK (Indian heritage)	Moving to Berlin for a startup job
P14	45	UAE	Spouse of a diplomat
P15	35	Ireland	Freelance software consultant, Berlin
P16	24	Nigeria	Master's student, engineering
P17	40	Portugal	Restaurant owner opening a German branch
P18	33	Turkey	Married to a German for 10 years
P19	58	USA	Retiree, relocated with spouse
P20	21	China	International exchange student

4 Method

4.1 Design

We ran a two to four week mixed methods pilot combining a before and after questionnaire with matched semi-structured interviews conducted immediately before and after the pilot window. The design deliberately paired quantitative Likert measures, reported here as medians given the expected small sample size, with qualitative interviews structured around the same four topic areas: current practice and tool fragmentation, motivation and goals, payment attitudes and sunk cost, and baseline or endline confidence. Full instruments, including the interviewer scripts, are available on request.

4.2 Participants

We recruited 20 adult self-directed German learners through convenience and snowball sampling, deliberately targeting a heterogeneous sample across age, occupation, and reason for learning German rather than a single population such as university students. The final sample included a freelance consultant, several relocated professionals, a graduate student, an exchange student, a retiree, a diplomat's spouse, an au pair, and several others, ranging in age from early twenties to sixties, with baseline German proficiency spanning complete beginner through B1. At baseline, participants reported using a median of two separate tools or resources to study German, most commonly Duolingo in some combination with a second resource such as Babbel, a grammar book, a class, or a tutor. Table 1 presents participant demographics.

4.3 Instruments and procedure

Each participant completed a pre pilot Google Form questionnaire covering demographics, current tool use, baseline attitudes toward payment models and persistence, and baseline confidence, followed by a twenty to twenty five minute semi-structured pre pilot interview covering the same ground in more depth. Participants then used Germanly for the agreed pilot window with no other instruction beyond normal use. At the end of the window, participants completed a matched post pilot questionnaire and a twenty five to thirty minute post pilot interview, both structured to allow direct before and after comparison on confidence, motivation, tool consolidation, and, centrally, the participant's actual experience of the one time payment compared to their baseline prediction about how a subscription would have affected them.

4.4 Analysis

Quantitative Likert data were analyzed with Wilcoxon signed rank tests for paired pre and post comparisons, appropriate given the ordinal scale and small sample size, alongside Spearman correlations between post pilot measures. Interview transcripts were coded thematically by the first author, working inductively from the data while remaining attentive to the constructs identified in the related work, sunk cost, psychological ownership, fragmentation, and self-determination, without treating those constructs as a fixed coding scheme applied mechanically.

4.5 Ethics and positionality

Participants gave informed consent, including consent to audio recording, were told their responses would remain anonymous in any written output, and were free to skip questions or withdraw at any point. Data collected for this study was not shared with Germanly or Pexilim beyond what appears in this paper, an agreement made explicit to participants before the study began. We note directly, as a conflict of interest that readers should weigh when assessing our conclusions, that the second author is affiliated with Pexilim, the company that builds Germanly, and that the first author has previously appeared in a testimonial on Germanly's public marketing site. We have tried to offset this through transparent reporting of disconfirming cases in our qualitative data, reported in section 6, and through statistical tests rather than selective quotation as the primary basis for our quantitative claims, but readers should treat this paper as a design case study conducted with insider access rather than as an arms length third party evaluation.

5 Results

5.1 Quantitative results

Self-reported confidence in holding a basic conversation in German rose significantly over the pilot window, from a median of 3 to a median of 4 on a 7 point scale (Wilcoxon signed rank test, $p < .0001$), with 19 of 20 participants showing improvement and none declining. Intrinsic motivation, measured as agreement with the statement that participants were learning German because they wanted to rather than because they felt they had to, also rose significantly, from a median of 4 to a median of 5 ($p = .0006$). Table 2 summarizes these paired comparisons, including the reversal of pricing beliefs discussed next.

Tool consolidation was substantial though not absolute. Sixteen of 20 participants reported continuing to use at least one other resource alongside Germanly during the pilot, but at a reduced rate compared to

Table 2. Paired comparisons across the pilot window (Wilcoxon signed-rank tests, 7-point scales unless noted).

Comparison	A	B	<i>p</i>
Confidence: pre vs. post pilot	3	4	<.0001
Intrinsic motivation: pre vs. post pilot	4	5	.0006
Pre-pilot belief: subscription vs. one time payment sustains use	5 ^a	3 ^b	.004

^aSubscription condition. ^bOne time payment condition. Both medians are pre-pilot beliefs, not post-pilot outcomes.

before, while four participants reported dropping every other tool entirely. Agreement that they were now using fewer separate apps or tools rose to a post pilot median of 6 out of 7.

The central and, we think, most novel result concerns the relationship between predicted and experienced attitudes toward the pricing model, though we are careful below about what kind of evidence this actually is. Before the pilot, when asked in the abstract whether paying once versus subscribing monthly would make them more likely to keep using an app, participants rated a monthly subscription as more likely to sustain their engagement than a one time payment, a significant difference (median 5 versus median 3, Wilcoxon $p = .004$) that runs directly against the premise this study set out to test. After actually living with Germanly's one time payment for two to four weeks, the post pilot median response to "if this had been a subscription I believe I would have used it about the same amount" was only 3 out of 7, meaning most participants retrospectively judged that a subscription would have led to less use than the one time payment they had actually experienced. This is a self-reported counterfactual belief, not a measurement of actual behavior under a subscription condition, since no participant was ever assigned to one; we treat it as a reversal in stated belief worth testing with a real pricing manipulation, not as evidence of what a subscription would in fact have done. Post pilot, participants reported a median of 5 out of 7 agreement that they had thought about the money already spent when deciding whether to keep going, and a median of 5 out of 7 agreement that knowing they had paid made them feel more obligated to continue.

Psychological ownership showed the strongest relationship to participants' stated intent to keep using Germanly after the pilot ended. Post pilot agreement with feeling that Germanly was something they owned, rather than something they were renting, sat at a median of 6 out of 7 and correlated significantly with plans to continue using the app (Spearman $r = .47$, $p = .037$). Neither having thought about money already spent ($r = .06$, $p = .81$) nor feeling obligated because of having paid ($r = .05$, $p = .82$) showed a significant relationship to continuation intent on their own, which we return to in the discussion: the pricing model appears to work by generating a sense of ownership over time, rather than by holding learners in place through guilt in any given moment. Table 3 presents these post pilot items alongside their correlation with continuation intent.

5.2 Qualitative themes

Fragmentation as a recurring, self-named burden. Nearly every participant described their pre pilot routine, unprompted, in terms of scattered tools rather than any single tool's shortcomings. Marta, a graphic designer using Duolingo, a lapsed Babbel subscription, and a grammar book simultaneously, said she would "open my phone and just stare at it trying to remember which app I was even in the middle of a lesson on,"

Table 3. Post-pilot items (7-point scale) and Spearman correlation with intent to continue using Germanly ($n=20$).

Item	Median	r	p
Feels like ownership, not renting	6	.47	.037
Thought about money already spent	5	.06	.81
Felt obligated because had paid	5	.05	.82
Subscription would have produced the same use	3	—	—
Now using fewer separate tools	6	—	—

adding that "it is a bit silly but it does put me off starting." After the pilot, the same participant described the shift directly: "before I would do vocabulary in one app and then have to remember the grammar rule from somewhere else. Having it connected made it feel less like homework." This pattern, fragmentation named as friction before the study and its absence named as relief afterward, recurred across participants with otherwise very different backgrounds, from a retired expatriate who called her consolidated routine "rather joined up" to a freelance consultant who said the switch "felt less like maintaining two separate systems."

Sunk cost surprising the learners who experienced it. Several participants who described themselves before the pilot as relatively indifferent to payment structure were visibly surprised, in their own words, that the one time payment came to mind during low motivation moments. One freelance consultant who had described subscriptions as something he simply "lets run even when unused, out of sheer laziness" reported afterward that the payment "surprisingly yeah, more than I expected given how I usually treat subscriptions" affected his decision to open the app on a difficult day. This self-surprise is, we think, the qualitative counterpart to the quantitative predicted versus experienced gap reported above: participants were not simply reporting what they assumed a researcher wanted to hear about sunk cost, since several explicitly flagged that the effect ran against their own prior self-understanding.

Spontaneous ownership language. Without being prompted with the word beforehand, several participants used ownership language earlier in the interview than the direct ownership question, describing Germanly as "mine" in a way their previous subscriptions never felt. A retiree in her sixties, comparing Germanly to her long running community class, said "it does feel like mine somehow, more than the class materials do, if that is not a silly thing to say." A graduate student who had not previously considered the distinction said, when asked directly, "I had not thought about it that way before, but yes, maybe it does feel more mine." This range, from a participant volunteering the language unprompted to a participant discovering the feeling only once asked to reflect on it, suggests psychological ownership was operating for most of the sample even when it was not the most salient thing on a participant's mind day to day.

A shift from obligation to curiosity. A smaller but consistent subset of participants described their motivation itself changing shape over the pilot, not just increasing in amount. Marta again: "I think it shifted from guilt towards something more like actual curiosity. I started looking forward to certain topics instead of just checking a box." A participant learning German to reconnect with music and online community described her motivation as having "deepened, honestly, less about the initial reason, music, more just general love of the language now." This maps closely onto self-determination theory's account of internalization, extrinsic motivation gradually becoming more autonomous, and it suggests the commitment

device may function partly as scaffolding that gives a learner enough consistent contact with the material for a more intrinsic motivation to develop, rather than as a permanent external prop.

Where the effect did not hold. Not every participant experienced the pricing model as consequential, and these disconfirming cases matter as much as the pattern itself. Ola, a university student with a hard institutional deadline for a language requirement, reported that a subscription would have produced "probably the same, honestly, I am consistent regardless of payment type," locating her motivation entirely in the external deadline rather than in anything about Germanly's design. Noor, who works with a private tutor alongside any app, located her sense of obligation in the human relationship rather than the payment: "with tutoring, the personal relationship creates the obligation, more than the payment itself I would say, I do not want to disappoint someone I am working with directly." Both participants still showed confidence gains and continued to value the consolidated structure, but the pricing model specifically appears to matter most for learners who lack an existing external commitment structure, whether a hard deadline or a human relationship, rather than for everyone equally.

Persistent gaps. Across both pre and post interviews, participants consistently named needs Germanly's structured, text and exercise based format does not fully address: native speaker audio and accent exposure, domain specific vocabulary for work, academic, or exam contexts, and above all, practice speaking with an actual person. These gaps are consistent with the automated feedback literature's finding that AI generated feedback concentrates its strength in surface level accuracy rather than the fuller construct of communicative competence, and they mark a clear boundary on what a single structured, one time paid app can be expected to deliver on its own.

6 Discussion

Our central finding is not that sunk cost or psychological ownership exist, both are well established constructs, but that self-directed language learners appear to systematically mispredict how these mechanisms will affect them before they have any lived experience of a one time payment to reason from. Asked in the abstract, our participants rated subscriptions as the more motivating structure, likely because subscriptions are the default they have the most experience with and because an ongoing charge feels, in prospect, like an ongoing reason to keep showing up. Only after living inside a one time payment for several weeks did most participants recognize, often with some surprise at their own reaction, that the completed purchase had done more psychological work than the counterfactual subscription would have. This gap between predicted and experienced response is well documented in affective forecasting research generally, but as far as we can tell has not previously been shown specifically for pricing model in a behavior change or learning context, and it has a direct design implication: asking users what pricing model they prefer, a question product teams ask constantly, may systematically undercount the actual effectiveness of a one time payment, because users are predicting from a subscription-saturated frame of reference rather than from experience.

The finding that psychological ownership, rather than felt obligation or thinking about money spent, was the item most strongly associated with continuation intent is worth dwelling on, because it reframes what the commitment device may be doing. A pure sunk cost account would predict that guilt about wasted money is the operative mechanism, and would imply a fairly bleak picture of users trapped by a psychological bias rather than genuinely engaged. Our data instead suggest a more constructive story closer to Pierce, Kostova and Dirks' account of ownership: the one time payment appears to function as

one route, alongside the sustained engagement and cumulative investment that the structured single path also encourages, toward participants coming to feel that Germanly is genuinely theirs, in a way a revocable subscription structurally cannot produce. This matters for how a designer might think about extending this finding: the goal is not to maximize guilt about sunk cost, which would be a fairly manipulative design target, but to design toward the sense of ownership that guilt is only a rough, and on our evidence not even the strongest, proxy for.

The disconfirming cases in section 5.2 point toward a boundary condition that we think is more useful to name explicitly than to smooth over. Participants who already had a strong external commitment structure, whether an institutional deadline or a paid human tutor, showed the confidence gains and valued the consolidated structure but did not show much evidence that the pricing model itself was doing psychological work for them. This suggests pricing as a commitment device is not a universal lever but a compensatory one, most useful precisely for the self-directed learners who lack an institutional or interpersonal structure holding them accountable, which is arguably the population self-directed learning tools should care about most, since learners with an institutional deadline or a tutor already have other reasons to persist.

We are cautious about overclaiming a causal design principle from a 20 person, two to four week, single-condition pilot with no control group, and we want to be explicit that this study cannot separate the effect of the pricing model from the effect of the structural consolidation, from novelty, from ordinary practice over several weeks, or from the attention of being interviewed twice, since Germanly bundles pricing and curriculum together by design and no comparison condition was run. What we can say is that the bundle, as experienced by a heterogeneous group of real self-directed learners, was associated with a psychological pattern, mispredicted pricing beliefs resolving into ownership rather than guilt, that neither the pricing literature nor the language learning literature had documented before. We treat this as hypothesis generating rather than confirmatory: the pilot literature is explicit that a small unblinded pilot is not powered or designed to establish effectiveness and should not be read as doing so [4], and the appropriate next step, detailed in Section 8, is a preregistered factorial study that unbundles pricing from curriculum structure and measures actual use from logs rather than self-report alone, not a claim that this pilot has already identified the mechanism.

7 Design Implications

If pricing model genuinely functions as part of the interaction design rather than as a business decision made after the interaction design is finished, that has a few concrete implications worth spelling out for anyone building a self-directed learning tool, language focused or otherwise.

- **Do not trust stated pricing preference at face value.** Our participants told us, before they had any experience of the alternative, that a subscription would motivate them more than a one time payment. A team genuinely trying to design for persistence, rather than for what converts best at checkout, should treat this stated preference as an unreliable guide and look instead at behavior over a trial period, since the mismatch itself is a design signal, not a market research inconvenience.
- **Design toward ownership, not toward guilt.** The item most strongly associated with continuation intent in our data was psychological ownership, not felt obligation or dwelling on money already spent. A product optimizing purely for sunk cost salience might, for instance, remind users

prominently how much they paid, a pattern that risks feeling manipulative and that our data do not support as the operative mechanism. A product optimizing for ownership instead might invest in what Pierce, Kostova and Dirks identify as ownership's roots: giving the learner more control over their own path, and creating a sense that the tool is a place the learner returns to and recognizes as theirs, for instance through progress that is unmistakably this specific learner's history with the tool, rather than a generic streak counter shared with millions of other users.

- **Treat structural consolidation and pricing as a single decision, not two.** Our study cannot cleanly separate the contribution of Germanly's one time payment from its single continuous curriculum, and we suspect, based on how tightly the two themes were braided together in participants' own accounts, that they are not meant to be separated. A team considering a one time payment without also consolidating a fragmented feature set into something a user can feel ownership over may see a weaker version of this effect, since the payment alone has nothing coherent for that ownership to attach to.
- **Know who the intervention is for.** Our disconfirming cases, a deadline driven university student and a learner already accountable to a paid tutor, suggest the pricing and structure combination is compensatory rather than universal. It appears to matter most for the population self-directed learning tools ostensibly serve: adult learners without an institutional or interpersonal structure already holding them in place, a population a team with limited resources might reasonably prioritize over learners who already have other reasons to persist.

8 Limitations

This is a small, unblinded, single condition pilot with no control group, run over a short window, and it cannot separate the causal contribution of the pricing model from that of the single path curriculum, since Germanly's design bundles both together. Our sample, while deliberately varied, was recruited through convenience and snowball sampling rather than random sampling and skews toward participants already motivated enough to volunteer, which may inflate the baseline persistence we observed relative to a broader population of self-directed learners. Confidence was measured through self-report rather than an external proficiency assessment, so our findings speak to perceived progress and motivation rather than to validated language gains, a question the existing efficacy literature reviewed in Section 2.1 is better positioned to address. The Likert items measuring sunk cost, obligation, and ownership were written specifically for this study and, while grounded in the theoretical constructs reviewed in Section 2.3, have not been validated against an established psychometric scale, so the specific numbers reported should be read as indicative rather than precisely calibrated. Finally, the conflict of interest disclosed in Section 4.5, one author's affiliation with Germanly's developer and the other author's prior public testimonial for the product, means this study should be read as an insider design case rather than an independent evaluation, and we would welcome replication by researchers without that relationship to the product.

8.1 Future work

The clearest next step is a preregistered, controlled study that unbundles pricing from curriculum structure rather than relying on this pilot's single bundled condition. Concretely, that means a 2×2 design crossing pricing model (one time payment versus subscription) with curriculum organization (single structured path

versus a modular, multi-path interface), with participants randomly assigned to condition rather than all receiving the same product as they did here. Confidence, motivation, and the ownership and sunk cost items would be retained, but actual use, sessions started, lessons completed, and return visits, should be measured from application logs rather than self-report alone, over a longer window such as 8 to 12 weeks rather than 2 to 4, so that novelty effects have room to fade. A waitlist or active comparator condition would let a future study also test whether Germanly improves confidence or motivation beyond what ordinary practice over the same period would produce on its own, a question this pilot’s before and after design cannot answer. Reporting should include confidence intervals for the pricing, structure, and pricing by structure interaction effects rather than p-values alone, consistent with guidance that pilot studies should generate hypotheses for a definitive study rather than substitute for one [4]. That factorial study, not this pilot, is where the paper’s central claim, that pricing model and curriculum structure function as a joint commitment device, would actually be tested rather than proposed.

9 Conclusion

Self-directed language learners assemble their own curricula out of whatever tools they can find, and the resulting fragmentation carries a real cognitive cost that the self-directed learning literature has only recently begun to name directly. Separately, the commitment device and sunk cost literatures establish that a completed financial commitment changes behavior in ways a revocable one does not, but this mechanism has gone almost entirely unexamined inside the specific, highly subscription-saturated market of language learning apps. Our pilot study of Germanly suggests these two threads are worth studying together rather than as separate design questions, though we want to be precise about what the pilot itself can and cannot establish. What we observed was a measurable rise in confidence and self-directed motivation, a real if partial consolidation of previously scattered tool use, and, most strikingly, a pattern in which learners retrospectively judged a hypothetical subscription as less motivating than the one time payment they had actually lived with, the reverse of what they predicted beforehand. Because every participant used the same bundled product, with no subscription condition, no unbundled curriculum condition, and no control group, this pilot cannot attribute those changes causally to pricing, to curriculum structure, or to their interaction, and we do not make that causal claim here. What it does establish is that the pattern exists and is worth testing properly: together with our finding that psychological ownership correlates more strongly with stated persistence than guilt does, this pilot generates a specific, testable hypothesis, that pricing model deserves the same design scrutiny HCI already gives to interaction design, rather than confirming it.

References

- [1] Dina Albassam et al. 2026. YT-Pilot: Turning YouTube into Structured Learning Pathways with Context-Aware AI Support. arXiv:2604.03543. doi:10.48550/arXiv.2604.03543
- [2] Anonymous. 2026. Good Fences Make Good Learning: How Self-Directed Language Learners Navigate LLM Delegation Decisions. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26)*. ACM, New York, NY, USA. doi:10.1145/3772318.3791657
- [3] Hal R. Arkes and Catherine Blumer. 1985. The Psychology of Sunk Cost. *Organizational Behavior and Human Decision Processes* 35, 1 (1985), 124–140. doi:10.1016/0749-5978(85)90049-4
- [4] Sandra M. Eldridge, Claire L. Chan, Michael J. Campbell, Christine M. Bond, Sally Hopewell, Lehana Thabane, and Gillian A. Lancaster. 2016. CONSORT 2010 Statement: Extension to Randomised Pilot and Feasibility Trials. *Pilot and Feasibility Studies* 2, 1 (2016), 64. doi:10.1186/s40814-016-0105-8

- [5] D. Randy Garrison. 1997. Self-Directed Learning: Toward a Comprehensive Model. *Adult Education Quarterly* 48, 1 (1997), 18–33. doi:10.1177/074171369704800103
- [6] Pingping Guo, Joanna Joseph Jeyaraj, and Abu Bakar Razali. 2024. A Systematic Review of Collaborative Mobile-Assisted Language Learning (C-MALL) Practices Using Bibliometric, Content, and Scientometric Analyses. *Humanities and Social Sciences Communications* 11 (2024), 1435. doi:10.1057/s41599-024-03940-3
- [7] Raghuram Iyengar, Young-Hoon Park, and Qi Yu. 2022. The Impact of Subscription Programs on Customer Purchases. *Journal of Marketing Research* 59, 6 (2022), 1084–1101. doi:10.1177/00222437221080163
- [8] Daniel Kahneman, Jack L. Knetsch, and Richard H. Thaler. 1990. Experimental Tests of the Endowment Effect and the Coase Theorem. *Journal of Political Economy* 98, 6 (1990), 1325–1348. doi:10.1086/261737
- [9] Matt Kessler, Shawn Loewen, and Talip Gönülal. 2023. Mobile-Assisted Language Learning with Babbel and Duolingo: Comparing L2 Learning Gains and User Experience. *Computer Assisted Language Learning* 38, 4 (2023), 690–714. doi:10.1080/09588221.2023.2215294
- [10] Hyunsoo Lee, Auk Kim, Hwajung Hong, and Uichin Lee. 2021. Sticky Goals: Understanding Goal Commitments for Behavioral Changes in the Wild. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. ACM, New York, NY, USA. doi:10.1145/3411764.3445295
- [11] Sophie Leroy. 2009. Why Is It So Hard to Do My Work? The Challenge of Attention Residue When Switching Between Work Tasks. *Organizational Behavior and Human Decision Processes* 109, 2 (2009), 168–181. doi:10.1016/j.obhdp.2009.04.002
- [12] Mariela Mihaylova, Simon Gorin, Thomas P. Reber, and Nicolas Rothen. 2022. A Meta-Analysis on Mobile-Assisted Language Learning Applications: Benefits and Risks. *Psychologica Belgica* 62, 1 (2022), 252–271. doi:10.5334/pb.1146
- [13] Christina Muckenhumer, Sigrid Schefer-Wenzl, and Igor Miladinovic. 2023. Efficacy Analysis of Mobile Language Learning Apps. In *Proceedings of the 2023 9th International Conference on Frontiers of Educational Technologies*. ACM, New York, NY, USA. doi:10.1145/3606150.3606152
- [14] Jon L. Pierce, Tatiana Kostova, and Kurt T. Dirks. 2001. Toward a Theory of Psychological Ownership in Organizations. *Academy of Management Review* 26, 2 (2001), 298–310. doi:10.5465/amr.2001.4378028
- [15] Jon L. Pierce, Tatiana Kostova, and Kurt T. Dirks. 2003. The State of Psychological Ownership: Integrating and Extending a Century of Research. *Review of General Psychology* 7, 1 (2003), 84–107. doi:10.1037/1089-2680.7.1.84
- [16] Richard M. Ryan and Edward L. Deci. 2020. Intrinsic and Extrinsic Motivation from a Self-Determination Theory Perspective: Definitions, Theory, Practices, and Future Directions. *Contemporary Educational Psychology* 61 (2020), 101860. doi:10.1016/j.cedpsych.2020.101860
- [17] Hui Shi and Vahid Aryadoust. 2024. A Systematic Review of AI-Based Automated Written Feedback Research. *ReCALL* 36, 2 (2024), 187–209. doi:10.1017/S0958344023000265
- [18] Yao-Ting Sung, Kuo-En Chang, and Je-Ming Yang. 2015. How Effective Are Mobile Devices for Language Learning? A Meta-Analysis. *Educational Research Review* 16 (2015), 68–84. doi:10.1016/j.edurev.2015.09.001
- [19] Jodi M. Tommerdahl, Chrystal Sapphire Dragonflame, and Amanda A. Olsen. 2024. A Systematic Review Examining the Efficacy of Commercially Available Foreign Language Learning Mobile Apps. *Computer Assisted Language Learning* 37, 3 (2024), 333–362. doi:10.1080/09588221.2022.2035401
- [20] Diane Yan and Shulin Zhang. 2024. L2 Writer Engagement with Automated Written Corrective Feedback Provided by ChatGPT: A Mixed-Method Multiple Case Study. *Humanities and Social Sciences Communications* 11 (2024), 1086. doi:10.1057/s41599-024-03543-y