PASTBEHAVIOR RESEARCH

Which Partners Are Worth Investing In?

What the historical record says about developing partners, and why partner investment usually follows the wrong signal

AUGUST 2026 SIXTEEN-CASE COHORT PLUS FALSIFICATION PASS

EVIDENCE LABELING

FACT A figure or event stated in a regulatory filing, a peer-reviewed experiment, or a company disclosure.

EVIDENCE-BACKED INFERENCE A reading the cited evidence supports but does not state.

REPORTED A dated attributable claim from an interested party or obtained through a secondary chain.

HYPOTHESIS Plausible and not established.

This study draws on a sixteen-case historical cohort of partner networks and a subsequent falsification pass that searched deliberately for evidence contradicting the cohort's conclusion. The second study found real counterexamples. That progression is preserved here rather than smoothed over, because the correction is the more useful half.

01

The Question

A company has built a partner network. It took years and it worked, in the sense that the agreements exist. Now the population inside it looks like this: a small group produces most of the revenue, a middle group produces occasionally, and a large group has never produced anything and shows no sign of starting.

The company has a finite amount of partner-manager time. It has a finite number of leads it can route, a finite incentive budget, a finite amount of margin it can give away, a limited quantity of enablement capacity, and a very small amount of executive attention. All of it has to go somewhere.

Where should it go?

Three answers are on the table in most companies, usually defended by three different people. Activate the dormant partners, because the roster represents capacity that has already been paid for. Develop the weaker producers, because they have shown they can transact and simply need help. Concentrate on the biggest producers, because that is where the revenue is and the rest is a distraction.

The historical record does not support any of the three as a standalone rule. It supports parts of each, under conditions that are more specific than any of them assume, and it points at a fourth consideration that none of them contains.

Prior work in this library covered what happens before this point. *Independent Distributor Activation* traced how a recruited network becomes a producing channel and found that economics come first and position inside an existing sales occasion comes second. *The Second Transaction* mapped the states an intermediary passes through and identified which transitions are expensive. *When Does Access Become Distribution?* established that reach predicts almost nothing about channel value on its own.

This study begins after all of that. The channel exists. The question is not how to build one. It is where to spend the next hour inside one that already exists.

02

What We Already Know

Two findings from earlier work are prerequisites rather than the argument, and are compressed accordingly.

Partner count is usually the wrong denominator. UWM discloses something almost nobody discloses, which is both the number of loan officers affiliated with its broker partners and the number who actually submitted a loan. `FACT:` in 2022, roughly 33,000 of over 45,000 affiliated loan officers submitted a loan. By 2024 the affiliated population had grown to over 55,000 and roughly 35,000 submitted. The roster grew about 22 percent. The producing population grew about 6 percent. The producing rate fell from 73 percent to 64 percent.

The exception is instructive because it is always the same exception. Primerica's per-representative productivity has held inside a band of roughly 0.17 to 0.24 policies per month for fifteen years, and the company's own 10-K states that sales volume therefore tracks the size of the sales force. `FACT.` But Primerica counts *life-licensed* representatives, and the funnel from recruit to licensed runs at roughly 10 to 14 percent. The licensing exam does the activation work before anyone enters the denominator. Counts predict production when the count already contains a gate.

Production concentrates, and more sharply than the familiar rule of thumb. `FACT:` at June 30, 2020, Datto had over 17,000 MSP partners, of whom just over 1,000 contributed $100,000 or more in annual run-rate revenue. Under six percent of the base accounted for 44 percent of ARR.

The more useful part of that disclosure is not the concentration. It is what the company put in its 10-K to explain what actually drives revenue: partner additions are a leading indicator of business health, but they **"do not immediately drive material revenue growth."** `FACT.` Growth came from net retention running between 111 and 119 percent, meaning existing partners buying more. A vendor with seventeen thousand partners told its investors, under securities liability, that adding more of them would not move the number.

03

The Case for Concentrating

The first cohort produced a conclusion that was uncomfortable and looked very well supported. Every documented improvement in producing-partner rate came from narrowing rather than from developing. Not one came from enabling the long tail more effectively.

The examples are not marginal.

eXp Realty ran a deliberate offboarding program beginning in late 2023. `FACT:` agent count fell across multiple consecutive quarters, from roughly 85,800 in March 2024 to 82,704 by mid-2025, while transactions per agent rose, with Q4 2024 showing a 12 percent increase in transactions per agent and a 23 percent increase in sales volume per agent on a smaller base. Management disclosed why the arithmetic worked so cleanly: 77 percent of agents who exited had between zero and two transactions in the

prior month. The company had been carrying platform cost, support cost, and management attention for a population contributing close to nothing.

SelectQuote did the same thing under harder circumstances. After a rapid agent-force expansion that ended in a lifetime-value correction and a program to remove over $250 million in annual cost, the company reversed. `FACT:` in the second quarter of fiscal 2025, a 22 percent reduction in agent headcount coincided with a 33 percent increase in productivity and close rates 24 percent higher, with Senior segment adjusted EBITDA margin rising from 32 percent to 39 percent. Management's stated explanation was overweighting tenured agents.

Herbalife runs the industrial version, continuously. `FACT,` from SEC filings: sales leaders must re-qualify annually. In 2010, 290,900 needed to re-qualify and 165,900 failed, a retention rate of 43.0 percent. By 2013, 379,600 needed to re-qualify and 182,800 failed, a retention rate of 51.8 percent. Roughly half the qualified producing population is removed every year and replaced. The company's own filing states that the average monthly purchase by a sales leader has remained relatively constant over time, so growth comes from the count rather than from the producer.

HubSpot has institutionalized the mechanism rather than discovering it under duress, requiring partners to reach gold tier within 24 months or face termination, and routing its own referrals preferentially toward partners already sourcing business. `FACT.` That second half will matter later.

Read together this is a strong case. Companies in unrelated structures improved partner-level production, and they did it by removing or de-resourcing non-producers rather than by teaching the long tail to sell.

The conclusion that follows is clean, memorable, and slightly too strong: you cannot create a productive partner, you can only find one and stay out of the way.

04

Where That Breaks

A second study was run specifically to falsify it, searching for controlled evidence that an existing population could be developed rather than culled. That evidence exists.

Bloom, Eifert, Mahajan, McKenzie and Roberts ran a management field experiment on large Indian textile firms, randomly assigning plants to five months of intensive on-site consulting from a major international firm, provided free. `FACT:` treated plants raised adoption of 38 measured management practices by 38 percentage points against 12 points in control. Productivity rose 17 percent in the first year through quality improvement, efficiency, and reduced inventory. Within three years, treated firms had opened more production plants.

The follow-up matters as much as the result. `FACT:` revisiting the plants eight to nine years later, the researchers found roughly half the adopted practices had been dropped, but a statistically significant 19.7 percentage point practice gap and a significant performance gap both remained. The most cited reasons for the reversal were managerial turnover and lack of director time.

These were independent firms rather than channel partners, and the study covered 17 firms. The transfer is an analogy and is treated as one. What it establishes without ambiguity is that an outside party can cause a durable change in what an independent business produces, under random assignment, and that roughly half the gain leaks away over a decade because the people who learned it leave.

Luo and colleagues ran randomized field experiments with two fintech companies, published in the Journal of Marketing. `FACT:` 429 sales agents were randomly assigned to on-the-job training with an AI coach or a human coach. The incremental benefit of the AI coach followed an inverted U across the performance distribution. Middle-ranked agents improved most. Bottom-ranked and top-ranked agents showed limited gains. The mechanism the authors identified for the bottom group was information overload.

The researchers then did the thing that makes this the most useful case in the study. They rebuilt the AI coach to give bottom-ranked agents **less** feedback, and ran a second

experiment on a separate sample of 100 bottom-ranked agents. FACT: performance improved substantially.

The bottom of the distribution was developable. The standard intervention failed on it. The correction was to make the intervention smaller.

These were internal sales agents rather than independent partners, and persistence past the experiment is unreported. Both limits are real and neither erases the finding. A third result points the same way with better durability evidence: **Frayne and Geringer** randomly assigned 30 insurance salespeople to self-management training with 30 controls and found performance improvement sustained across twelve months, then replicated it by training the control group afterward. Sixty people is a small study. The reversal design is unusually good protection against the selection explanation, and the content is worth noting, because the training was about self-management rather than about the product.

The simple selection thesis does not survive this. Development works.

What it does not do is work the way most partner programs assume. Each successful intervention was individualized rather than broadcast, ran for months rather than arriving as content, and aimed at changing behavior rather than transferring information. In the one case that reached the bottom of the distribution, the volume of intervention had to be reduced.

05

Four Populations Wearing One Label

The Second Transaction modeled intermediaries as passing through six states and identified State 2, signed and dormant, as where most of the population lives and where the entire problem sits. That model holds. The amendment is that State 2 is not one population.

NEVER PRODUCED, NEVER ENGAGED

The best measurement of this group is accidental. When Pierce, Rees-Jones and Blank ran a randomized experiment on a car manufacturer's dealer incentive program, they first had to invite dealers to participate. FACT: of 1,227 dealers in the program, 294 opted in, 336 explicitly opted out, and 597 failed to respond at all. Non-respondents averaged 31.1 monthly vehicle sales against 48.5 for participants, a statistically significant gap.

Nearly half of an active dealer network did not answer a message about changes to how their own bonus money would be paid, and the non-answerers were the smaller ones. Reachability comes before convertibility, and this group fails the first test.

NEVER PRODUCED BUT ENGAGED

No controlled evidence in either direction. That is the largest hole in the record and it should be named as a hole rather than filled by analogy to the group above or below it.

LOW-PRODUCING BUT OPERATING

Every positive causal result sits here. Bloom's plants were running. Luo's bottom-ranked agents were ranked, which means they had sales. Frayne's salespeople had books.

FORMERLY PRODUCTIVE AND DECAYED

Herbalife's annual demotions put this population between roughly 145,000 and 183,000 people a year inside a single company. Bloom's nine-year

follow-up supplies the mechanism most likely to matter, since practice reversal was attributed primarily to managerial turnover.

`EVIDENCE-BACKED INFERENCE:` a decayed partner is often not a decayed firm. It is a new individual inside an experienced firm, never trained because the firm was already onboarded. Treating that as re-engagement of a lapsed account will miss it entirely.

A fifth group exists and does not appear on that list, because it is not dormant at all. It is high-producing and unresponsive, and conventional partner metrics have no field for it. Section 07 is about that group.

One hard result should temper any enthusiasm about the first population. Three independent randomized studies found that units far from an attainable target do not respond to economic intervention. The Pierce distributional analysis found dealers far short of their threshold showed no response to the incentive change. A relative performance pay experiment in a retail chain found stores lagging far behind did not respond, with responsiveness rising as the gap narrowed. And `REPORTED,` from secondary coverage of Chung and Narayandas's field experiment at an Indian consumer durables manufacturer, a delayed bonus produced roughly a 10 percent lift concentrated among high performers, with the effect on low performers close to zero.

Money moves people who are already producing. There is no case in this record of money moving people who are not.

06

Why the Standard Enablement Stack Has So Little Behind It

Certification tracks, content libraries, partner academies, portals, and onboarding curricula absorb most of the discretionary budget in partner programs. The published evidence that any of them converts a non-producing partner into a sustained producer is close to nonexistent. *The Second Transaction* rated certification completion as unresolved, with plausible correlation and no causal evidence. The falsification study did not find the causal evidence. It did find a good explanation for the absence.

McKenzie and Woodruff reviewed the accumulated randomized evaluations of business training. A companion analysis examined five studies that measured business practices with a common instrument and found that effects on sales and profits were always consistent with the observed change in practices, and that most studies found small and insignificant effects on output **because they found small effects on practice adoption in the first place.** `FACT.` The authors are explicit that the correct conclusion is not that practices do not matter, but that most training programs are too weak to change what people actually do.

That reads as a criticism of training and is better understood as a specification. Bloom's intervention moved practice adoption 38 points, which is why output moved. A webinar and a certification quiz do not move practice adoption 38 points. Output follows behavior change, behavior change follows intensity and individualization, and the enablement stack that scales provides neither.

The most expensive negative result available belongs here too. The Pierce dealer experiment randomized $66 million in bonus payments across 294 dealerships representing over 15,000 vehicles and $600 million in monthly revenue. The intervention was prepaying the bonus with a clawback if the target was missed, a technique with a substantial behavioral economics literature behind it. `FACT:` treated dealers sold 2.31 fewer cars per dealer per month, or 3.8 percent fewer, and the authors state they can reject any positive effect larger than 0.47 percent of sales at the 5 percent level. Dealers protected the larger of two bonuses by neglecting the smaller one.

The same experiment's baseline data shows heavy bunching just above the bonus thresholds. Dealers respond to compensation with precision. Changing the framing of

identical compensation produced nothing positive and something negative.

07

The Turn: Production Level Is Not Responsiveness

Everything above is about whether partners can be developed. The more consequential finding is about something else, and it arrived from an industry that has been studying this question with better data than anyone in software.

Pharmaceutical companies send sales representatives to call on physicians. The industry has decades of individual-physician panel data linking calls received to prescriptions written. It has every reason and every input required to allocate that expensive human effort well.

Manchanda, Rossi and Chintagunta built a model designed to handle the fact that sales calls are not randomly assigned, and used it to ask how firms actually allocate them. `FACT,` in the authors' own words: **"physicians are not detailed optimally; high volume physicians are detailed to a greater extent than low volume physicians without regard to responsiveness to detailing. In fact, it appears that unresponsive but high volume physicians are detailed the most."**

A companion study by Manchanda and Chintagunta found that detailing does raise prescriptions written, with diminishing returns for most physicians, and showed that reallocating calls could increase revenue.

Sit with what the first finding says. The most data-rich sales organizations in the economy send their most expensive resource to the people who buy the most, without reference to whether those people buy more because of the visit. The single largest concentration of effort goes to physicians who are both high volume and unresponsive, which is the worst available combination for a marginal dollar.

This is not a criticism of concentration. It is a criticism of the variable concentration is measured on.

"Which partner produces the most?" and "which partner produces more because of what we do?" are different questions, and the answers come apart.

The second question has a name in the underlying literature. It is the slope of the response to effort. In ordinary language, it is the additional production you get from the next unit of investment in a given partner. It is not the partner's revenue. It is the difference between what the partner produces with your effort and what the partner would have produced without it.

Consider a company deciding where to put the next hour of a partner manager's week.

ILLUSTRATION, NOT RESEARCH DATA

PARTNER A		PARTNER B	
Annual production	$2,000,000	Annual production	$500,000
Without active support	≈ $1,950,000	Without active support	≈ $150,000
Value of the effort	**≈ $50,000**	**Value of the effort**	**≈ $350,000**

Partner A's relationship is embedded, the end customers renew, and the account largely runs itself. Partner B's support is what generates the deals. Partner A is four times larger. Partner B is seven times more responsive. Every conventional partner scorecard ranks A first, even though the next marginal hour may be considerably more valuable with B.

Nothing in the historical record establishes that this pattern is common in software partner programs specifically. The detailing evidence is from pharmaceuticals and I am not going to claim otherwise. What that evidence does establish is that the error is real and that it persists in an industry with far better measurement than most partner programs have. In the detailing evidence, the error runs consistently in one direction: effort follows volume, while volume is not responsiveness.

There is a second-order version worth naming because it runs against current fashion. The concentration strategy in section 03 is well evidenced as a way to raise average production per partner. It is not, on this evidence, established as a way to maximize return on the resources being concentrated. Removing non-producers and reallocating their support to top producers are two separate decisions. The first has good evidence behind it. The second contains an assumption nobody checks.

08

What a Company Would Actually Have to Know

Conventional partner systems answer who produces, who is active, who logged in, who completed certification, who registered a deal, and who sits in which tier. Those are measures of state, and state turns out to be a weak guide to where the next dollar should go.

The questions that would matter:

Who changes when we intervene?

Not who produces after we intervene. Who produces differently than they would have. That requires a comparison, and the comparison requires that some comparable partners did not receive the intervention.

Which intervention changes them?

Interventions are not interchangeable. Bloom's worked because it was intensive and individualized. Luo's worked on the bottom cohort only after it was made less intensive. A partner who responds to leads may not respond to coaching, and a partner who responds to margin may not respond to either.

How much does the change cost?

Five months of top-tier consulting per plant produced a 17 percent productivity gain. That is a real result and not a scalable one. Any responsiveness measure that ignores intervention cost will recommend spending unlimited money on the most responsive partner.

Does the effect persist?

Bloom's follow-up is the only long-run answer in the record and it says roughly half the gain reversed over nine years, primarily because people left. Luo's persistence is unmeasured, which caps how much weight the software finding can carry.

When does additional attention stop paying back?

The evidence supports three warning signs and no numerical threshold. Distance from an attainable target predicts non-response, in three independent studies. Non-engagement with communications about the partner's own economics predicts everything else, on the Pierce non-response data. Absence of observable behavior change after an intervention predicts absence of output change, on the McKenzie and Woodruff mechanism. What the evidence does not support is any claim that a specific number of dormant days or failed touches predicts non-conversion. Nobody has published that study.

The honest position is that measuring responsiveness is hard, for a reason that will not go away. It requires variation in the intervention. A company that gives every partner the same support cannot learn which partners respond to it, because there is nothing to compare against. Learning requires deliberately treating comparable partners differently and watching what happens, which is uncomfortable to propose, awkward to explain to the partners who get less, and the only way the question gets answered.

09

What This Implies About Software

The Luo experiments are the first credible randomized evidence in this record that a software-delivered intervention changed production. That is a genuine update.

The shape of the result matters more than its sign. The naive version of the AI coach failed precisely on the population with the most room to improve, because it delivered too much. The version that worked delivered less. The diagnosis that produced the fix came from a research team studying the failure, not from the software.

Most enablement technology does the opposite of what that implies. It reduces the marginal cost of delivering content, which means it delivers more content to more partners, which means it delivers the most to the partners who can absorb the least.

The roles that follow more naturally from the evidence are detection and measurement rather than delivery. Detecting who is producing against a real denominator rather than a roster. Detecting decay before a partner formally lapses, given that partners rarely resign and simply stop. Reducing the cost of running the variation that responsiveness measurement requires. Personalizing intensity rather than scaling volume. And helping humans decide where to go, since the human layer does not disappear as automation improves. It concentrates, and the detailing evidence says it currently concentrates on the wrong signal.

10

Returning to the Question

Which partners are worth investing in?

Not the dormant ones, on the available evidence, with an important qualification. There is no credible evidence that conventional enablement converts a partner who has never produced into a sustained producer, and there is direct evidence that units far from an attainable target do not respond to the economic levers that work on everyone else. But absence of evidence is not proof of impossibility. Failed reactivation programs are not written up, and a forty-year-old dealer development consulting sector has operated at scale without producing a single denominator-bearing outcome study I could locate.

The weaker producers, yes, and under narrower conditions than most programs assume. Under random assignment, intensive individualized intervention produced durable productivity gains in independent firms, and calibrated coaching improved the bottom of a sales distribution. The intervention had to be individualized, sustained,

aimed at behavior rather than information, and delivered to a unit already operating. In the one case that reached the bottom of the distribution, it had to be made smaller rather than larger.

The biggest producers, sometimes, and this is where the question gets harder than it looks. Concentrating on producers is well evidenced as a way to raise average production per partner. Whether it is the best use of the resources being concentrated is a separate question, and the one industry that has studied it with adequate data found that firms send their most expensive resource to their highest-volume accounts without reference to whether those accounts respond, with the least responsive high-volume accounts absorbing the most attention of all.

So the answer is not a list of partner types. It is that the question most companies are answering is not the question they think they are answering. Most companies can identify their biggest producers. That is a solved problem and the data has been sitting in the CRM for years. What it does not tell them is where the next dollar or the next hour of partner investment will create the most incremental production, and on the evidence available, those two things are not the same.

Methodology and Limitations

Construction. A sixteen-case historical cohort of partner networks spanning SaaS channels, insurance distribution, mortgage wholesale, real estate brokerage, veterinary referral, franchise systems, and one regulatory natural experiment, followed by a falsification pass that searched deliberately for controlled evidence contradicting the cohort's conclusion. The falsification pass excluded all cohort members and drew new cases from the randomized field-experiment literature, SEC filings, and peer-reviewed marketing science.

The strongest counterexample is not from a channel. Bloom's units were independent textile firms receiving free consulting, not partners choosing whether to sell for a vendor. Seventeen firms. The transfer is an analogy and is treated as one throughout.

Luo's agents were internal. Sales agents at fintech companies, not independent partners with their own P&L and competing priorities. An independent partner has an outside option an employee does not, which is the central difference in every partner problem. Persistence past the experiment is unreported, which caps the weight the software conclusion can bear.

Frayne and Geringer is small and old. Sixty insurance salespeople, published in 2000. The reversal design is unusually good protection against selection, and the sample is what it is.

The detailing evidence is not a partner program. Physicians are not resellers, prescriptions are not deals, and the compensation structure is entirely different. What transfers is the demonstration that a well-measured industry allocates on volume rather than responsiveness. Whether the same misallocation occurs at the same magnitude in B2B software channels is untested, and this is the most important untested claim in the study.

Some incentive evidence comes through secondary reporting. The Chung and Narayandas figures on high-versus-low performer response were obtained from trade coverage of the authors' remarks rather than from the paper. Labeled `REPORTED` and should not carry decisive weight.

Herbalife's retention improvement is unexplained. The re-qualification rate rose from 43.0 percent to 52.0 percent between 2010 and 2012 across roughly 300,000 producers. That is the largest sustained-production improvement in a partner-like network anywhere in this record and the company does not disclose what caused it. It could be a program change, a definitional change, a compensation change, or a market effect. Left open.

Publication bias runs both ways. Failed activation programs are not published, biasing the record against finding conversion. Successful interventions are

published preferentially, biasing it the other way. A meta-analysis of the loss-framing literature specifically found that laboratory studies show large positive effects while field studies show near-zero ones, attributable partly to publication bias and underpowered designs.

A structural gap in the incentive literature. Nearly every incentive experiment located here tests a change in framing, timing, or frequency while holding the amount of money constant. Almost none tests a change in the amount. The corpus can say framing does not work and cannot say whether paying more works. Any conclusion about compensation as a lever inherits that limitation.

Not pursued. Direct-selling reactivation programs with disclosed cohort follow-through, franchisor field consultant allocation data, and distributor development programs in industrial equipment. All three would strengthen the sections on development and allocation and none produced usable denominators within scope.