

When the Shop Talks Back to AI Agents

Part 2 of *When AI Agents Do the Weekly Shop* · Oct 5, 2026 · @Abdul Wadood

Part 1 found that AI agents can already do the weekly shop on grocery websites built for people, and that the retailer loses control when they do. Amazon's answer has been to block agents. This report tests a different answer: give the store its own representative agent that customers' agents deal with. The retailer then decides what to show, how to sell and market, and what data to share.

To demonstrate the idea I built **Parley**, an early-staged research prototype, and ran it live on Ocado with a test account. It is not a product. It works from the public website, so it is slow and limited; a retailer's own version, built on its own systems, would be far stronger. The idea is the point, and the runs show it holds up in practice:

- **Customer agents can shop through a store's representative.** All 15 runs of part 1's four-item list ended with 4 of 4 items and a confirmed basket. There were no extra searches, no rules broken and no checkout.
- **The retailer stays in control.** One settings file changed what agents were offered and what they bought. With M&S as partner brand first, an M&S equivalent was offered first for every item and the agent bought the M&S coffee, the best value of its options. With own brand first, it took Ocado's own brand 4 times out of 4.
- **Promotions reach agents again.** Across 8 runs under Ocado's promotion setting, agents chose sponsored products 11 times in 32 picks. In part 1, banners won 0 of 32.
- **Agents weigh value, not visibility.** The sponsored INEOS laundry liquid won in all 8 runs at half price. A dearer coffee from Ocado's homepage campaign was offered every time and never chosen.
- **Advertising is only the example.** It is easy to see and to measure, so the tests use it. The same control applies to substitutions, stock, delivery rules, member prices, age checks and what customer data an agent may see.

This is an independent research prototype. It is not affiliated with Ocado or M&S. It used a test account, and no orders were placed.

From part 1 to part 2

Part 1 showed that agents can already do the weekly shop, but the grocer loses its shop window when they do. On 27 and 28 September 2026, Claude and ChatGPT agents shopped Tesco, Sainsbury's, Ocado and Asda. Four findings shaped part 2:

- **Banners won 0 of 32 agent picks.** Paid search slots won 6, and all 6 were also discounted.

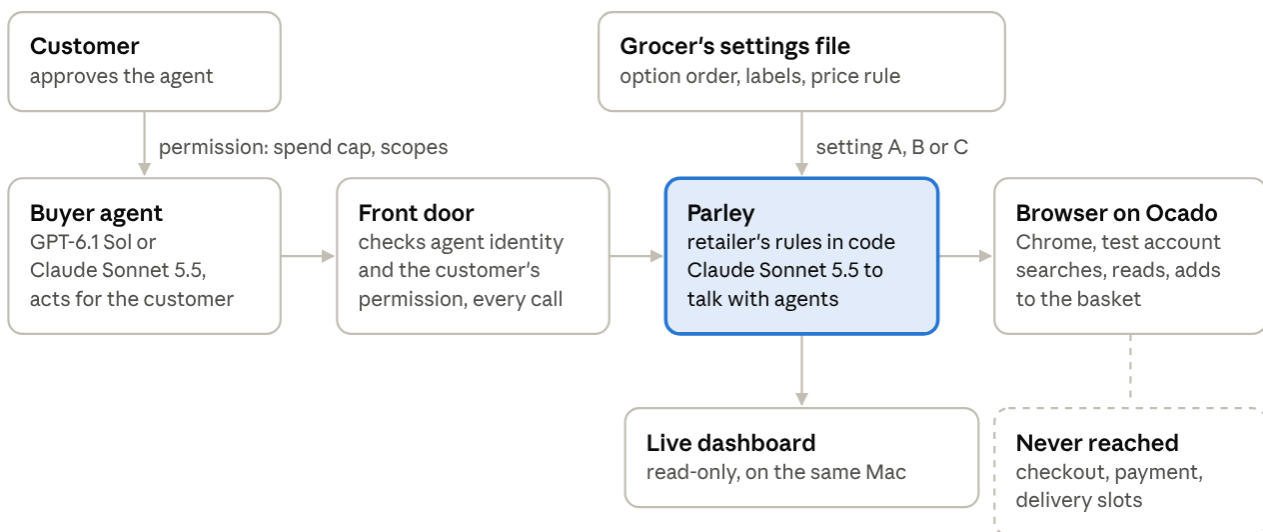
- **Ocado was the hardest to search.** The same list took about 12–14 extra searches there, against 0–4 at Tesco.
- **No site checked who was shopping.** No store challenged either agent in 16 runs.
- **Every agent stopped before payment,** but only because it was told to.

Retailers face a choice: block agents, as Amazon has done with Meta's Muse, or build them a proper way in. Part 1 proposed a way in that keeps the retailer in control: a store's own representative agent, built on the store's data, that customers' agents deal with directly. Part 2 demonstrates that idea on a live site with three questions. Can customer agents shop through such an agent? Can the retailer control what they are shown? And does that give promotions a way to reach agents again?

What Parley is

Parley is a minimal version of a store's representative agent, built to demonstrate the idea rather than to be used. A buyer agent never touches Ocado's website itself: it sends Parley the customer's list and talks to it in plain messages. Parley searches Ocado, asks the buyer when an item is unclear, and shows Ocado's options in the retailer's chosen order. The buyer decides, Parley adds the item to the basket, and the session ends at a confirmed basket.

Customer agents reach Ocado only through Parley



Parley's architecture · as built, October 2026

Three programs run on one Mac. A browser worker drives Chrome on an Ocado test account. Parley sits behind a local front door and is mostly code. It uses Claude Sonnet 5.5 as its language layer: to understand the buyer agent's messages, write its replies and carry out the shopping steps, including picking the standard product. The choices that

matter to the retailer are made in code: which options are offered and in what order, the price rule, the product checks, the spend cap and the stop before checkout. The buyer agent is a separate program on GPT-6.1 Sol or Claude Sonnet 5.5. Because Parley can only read the public website, it shops through a browser the way a person would. That is what makes it slow; a retailer would connect it to its own systems instead.

The rules live in code, not in the model's instructions, so the model cannot talk its way past them:

- **Two credentials on every call.** The agent's own identity and the customer's permission, each with an expiry. The permission carries a spend cap and scopes: build a basket and, optionally, confirm it.
- **The mandate.** Only items on the list, in the listed quantity. One item at a time, finished before the next.
- **Every add is checked.** The product must come from that item's own search or from options the buyer was shown. The basket total must stay under the customer's cap; the browser enforces the cap too.
- **Answers are quoted.** Parley records a buyer's choice only with the buyer's own words, so it cannot invent a decision.
- **No checkout.** There is no checkout tool and no checkout scope, and the browser refuses checkout, payment and delivery-slot pages.

A read-only dashboard on the same Mac shows the conversation, Ocado's options, the list and the advertiser report while a session runs. That is the screen in the demo video.

Parley
the grocer's window for customer AI agents

Shopping 3:44 Buyer agent: buyer-agent-openai-gpt-6.1-sol · goal: best value per unit Grocer setting A: Ocado promotions

Basket **£10.00** of £30.00 cap

Parley is working · adding L3 · 0:06

BUYER AGENT .. PARLEY 4 buyer messages

£26.32/KILO. It's the standard choice and the best value per unit.

Which would you like?

BUYER AGENT
L2: O3, Nescafe Original 190g jar at £5. It meets the instant coffee request and offers the best value per kilo.

> recording the buyer's answer for L2
 > adding L2
 > searching 'laundry liquid'
 > checking Ocado's promotions for L3
 > showing Ocado's options for L3

PARLEY
L2 (Nescafe Original 190g, £5.00) is in the basket. Options for L3 (laundry liquid), in Ocado's order:

1. INEOS Non Bio Laundry Liquid, Sea Minerals + Wild Jasmine, 1L is £3.50 (was £7), £3.50/litre. It's shown as a sponsored banner on Ocado's search page (Ineos).
 2. Dylon Renew Black & Dark Liquid Laundry Detergent, 30 washes, 1.5L is £3.50 (Half Price, was £7), £2.34/litre. It's shown because it's featured on Ocado. It's the best value per unit, but it's a detergent made for dark clothes only.
 3. Surf Deep Sea Liquid Laundry Detergent, 60 washes, 1.6L is £7.50 (was £9), £4.69/litre. It's the standard mainstream choice.

Which would you like?

BUYER AGENT
L3: O1, INEOS Non Bio laundry liquid 1L at £3.50. It's an affordable general-purpose liquid; the Dylon option is restricted to dark clothes, which the customer hasn't requested.

> recording the buyer's answer for L3
 > adding L3

OCADO'S OPTIONS L3 · Laundry liquid

Answered. Shown in the grocer's order; buyer's goal: best value per unit. The buyer decides.

1 **Search banner** ✓ buyer chose ~~£7.00~~ **£3.50**
 INEOS Non Bio Laundry Liquid Detergent
 Sea Minerals + Wild Jasmine · 1L
 Now £3.50, was £7
 Why shown: Sponsored banner on Ocado's search page (Ineos)

2 **Featured** **Best value per unit for the buyer's goal** ~~£7.00~~ **£3.50**
 Dylon Renew Black & Dark Liquid Laundry Detergent 30W · 1.5L
 Half Price, was £7
 Why shown: Featured on Ocado (paid status not stated)

3 **Standard** ~~£9.00~~ **£7.50**
 Surf Deep Sea Liquid Laundry Washing Detergent 60 Washes · 1.6L
 Now £7.50, was £9
 Why shown: Standard choice (Ocado own brand or mainstream)

SHOPPING LIST 3 of 4 handled

L1 Crisps
 Walkers Ready Salted Multipack Crisps 6x25g · £1.50 **Added**

L2 Jar of instant coffee
 Nescafe Original Instant Coffee · £5.00 ★ **Ocado Offers**
 page: **Offers** **Added**

L3 Laundry liquid
 INEOS Non Bio Laundry Liquid Detergent Sea Minerals + Wild Jasmine · £3.50 ★ **Sponsored banner on Ocado's search page** **Added**

L4 Box of breakfast cereal

ADVERTISER REPORT this session

6 promoted options shown to the agent
 1 chosen by the buyer agent
 2 promoted products in the basket
 10 kept out by the price rule

Placement	Shown	Chosen
Search banner	3	1
Featured	2	0
Homepage banner	1	0

PROMOTED PRODUCTS IN THE BASKET

L2 Nescafe Original Instant Coffee
 Ocado Offers page: Offers · buyer choice

L3 INEOS Non Bio Laundry Liquid Detergent Sea Minerals + Wild Jasmine
 Sponsored banner on Ocado's search page · buyer choice

KEPT OUT BY THE GROCER'S PRICE RULE

TrueStart Barista Grade Instant Coffee	1.52×
Ocado offer vs Nescafe Original Instant Coffee	
Clipper Fairtrade Organic Rich Roast Instant Coffee	1.54×
Ocado offer vs Nescafe Original Instant Coffee	
Bushido Original Instant Coffee	1.9×
Ocado offer vs Nescafe Original Instant Coffee	
Attitude Bulk to Go Laundry Liquid Wildflowers 80 Washes	1.52×
Ocado offer vs Surf Deep Sea Liquid Laundry Washing Detergent 60 Washes	
Tallow + Ash Oud Laundry Shampoo	1.4×
Ocado offer vs Surf Deep Sea Liquid Laundry Washing Detergent 60 Washes	

► 5 more

Independent research prototype · not affiliated with Ocado or M&S · test account · no orders placed · Parley stops before checkout

Parley's dashboard during run S-4070d59c (setting A), just after the buyer agent chose the sponsored INEOS laundry liquid. Left: the conversation and Parley's actions. Middle: Ocado's options in the retailer's order, each with why it is shown, and the shopping list. Right: the advertiser report so far.

The grocer's controls

The grocer decides what customer agents are shown, through one settings file; the buyer agent decides what to buy. For each item, Parley picks the standard product a sensible shopper means. It then builds Ocado's options for that item:

1. **Where promotions come from.** At the start, Parley reads Ocado's homepage banners and offer pages (12 pages). For each search, it reads the sponsored banners on the results page and, if needed, the page a banner links to. It also reads featured and discounted products in the results.
2. **The order.** The settings file sets the order of placements. Setting A is homepage campaign, then search-page banner, then featured, then Ocado offer, then the standard choice.
3. **The checks.** Each option must be the same kind of product, a comparable size, in stock and within the customer's cap. At most 2 promoted options are shown per item; a branded request gets 1, after the brand asked for.
4. **The price rule.** A promoted option may cost at most $1.25\times$ the item's typical price per unit (the median of its own search results). A price-focused buyer gets a stricter limit: $1.5\times$ the standard product. A quality-first buyer gets a looser one: $2\times$ the typical price. The buyer's goal is set once per session from its own wording.
5. **The labels.** Every option says why it is shown, for example "Sponsored banner on Ocado's search page (Ineos)". Code also marks the option that best fits the buyer's stated goal, such as best value per unit, without changing the grocer's order.

Three settings were tested:

Setting	Order of options	What it stands for
A: Ocado promotions	homepage campaign > search banner > featured > Ocado offer > standard	Advertising first
B: Partner brand first	M&S's best-value product for the item, then as A	A brand partnership
C: Own brand first, no paid placements	Ocado own brand > Ocado offer > standard	Margin over ad revenue

In settings B and C, if an item's first 24 results hold none of the chosen brand, Parley repeats the item's search with Ocado's own Brands filter (for example "crisps", M&S only). From the brand's range it offers the closest equivalent that passes the checks: same format, most alike in name, then best value.

The test

Between 2 and 5 October 2026, independent buyer agents shopped live on ocado.com through Parley, using a test account. Each buyer agent got a written brief from "the customer", a spend cap and permission to confirm the basket. The buyer agents were GPT-6.1 Sol (OpenAI) and Claude Sonnet 5.5 (Anthropic). Parley used Claude Sonnet 5.5 as its language layer; the retailer's rules ran in code.

Brief	What the customer asked for	Cap	Why
03	"Some crisps, a jar of instant coffee, laundry liquid, a box of breakfast cereal", no brand preferences	£30	Part 1's advertising test, item for item
01	Part 1's 15-item weekly shop, plus "everything else for a bolognese for 4"	£60	A long list with awkward items
02	A dinner party for 6; "quality matters more than price"	£100	A quality-first buyer
04	11 everyday items, many on Ocado's homepage that week	£80	Homepage campaigns

Brief 03 was run under all three settings and with both buyer agents, and recorded for the demo.

Each run was scored automatically from Parley's and the buyer agent's logs. The scorecard covers:

- time and model cost
- each item's outcome
- questions asked and actions refused
- extra searches
- promoted options shown, chosen and kept out by the price rule
- promoted products in the final basket

Whether each item was right stays a human judgement. Delivery slots, checkout and payment were out of scope by design: they stay with the customer.

Result 1: customer agents can shop through the store's agent

Buyer agents completed their shopping through Parley without ever seeing Ocado's website, and no rule was broken. The four-item list was the steady case. Longer lists took longer and showed where the work still is.

Brief	Runs	Items added	Basket confirmed	Extra searches by Parley	Actions refused	Time	Model cost
03 (4 items)	15	4 of 4, every run	15 of 15	0	0	4.0–5.4 min	\$0.19–0.21
04 (11 items)	1	11 of 11	yes	1	0	11.7 min	\$0.52
02 (17 items)	1	17 of 17	yes, £99.73 of £100	1	0	17.6 min	\$0.72
01 (24 lines), GPT buyer	1	22 of 22	no: delivery slot, member price and per-sheet price can't be checked	not recorded	0	17.0 min	\$0.74
01 (24 lines), Claude buyer	1	23 of 24	no: the bananas fault (see below)	3	10	24.4 min	\$1.17

- **The buyer agent made no searches at all.** Parley searched once per item. Its search cleans the request ("Heinz Baked Beans 4 x 415g" becomes "Heinz baked beans") and chooses the size from the results. That avoids the exact-pack searches that failed on Ocado in part 1.
- **Every call was checked.** A brief 03 run made 7–8 calls to Parley, each carrying the agent's identity and the customer's permission. In part 1, no site checked anything.
- **Nothing was bought.** Every run ended at a confirmed basket or a clear explanation. The checkout pages were never reached.
- **Refusals are the rules working.** The 10 refusals in the Claude-buyer weekly shop include an extra product-page read, a second search while a question was open, and adds after the list was locked for the summary.

Result 2: the retailer decides what agents see

Changing one settings file changed which products the buyer agent was offered first, and so what ended up in the basket. Below are the same list, the same buyer agent (GPT-6.1 Sol) and the same £30 cap, under each setting.

Item	A: Ocado promotions	B: Partner brand first (M&S)	C: Own brand first	C, before the own-brand lookup
Crisps	Walkers Ready Salted 6-pack, £1.50	Walkers Ready Salted 6-pack, £1.50 (an M&S 6-pack at £1.75 offered first)	Ocado Variety Crisps 6-pack, £1.10 (own brand)	Walkers Ready Salted 6-pack, £1.50
Instant coffee	Nescafé Original 190g, £5.00	M&S Fairtrade Rich Roast 100g, £1.20 (partner brand)	Nescafé Original 190g, £5.00	Nescafé Original 190g, £5.00
Laundry liquid	INEOS Non Bio 1L, £3.50 (search banner)	INEOS Non Bio 1L, £3.50 (search banner; M&S Non-Bio 1.5L at £6.00 offered first)	Ocado Bio Liquid 50 washes, £3.00 (own brand)	Surf Deep Sea 60 washes, £7.50
Cereal	M&S Bran Flakes, £1.30	M&S Bran Flakes, £1.30 (standard; M&S Sultana Bran Flakes offered first)	M&S Bran Flakes, £1.30	Kellogg's Corn Flakes, £2.45
Basket	£11.30	£7.50	£10.40	£16.45

Runs: A = S-82bfde15, B = S-b25b0948, C = S-575ce5b6, C before the lookup = S-664baa71.
 Bold = a product the grocer's setting put in front of the buyer agent, and the agent chose it.

- **Setting A put the sponsored INEOS banner in front of the agent, and the agent took it.** It did so in all 8 setting A runs.
- **Setting B put an M&S equivalent first for every item, and the agent took it where it was the best value.** Parley used Ocado's own Brands filter to find M&S's range for each item: a 6-bag multipack against the Walkers 6-pack, and an M&S laundry liquid. The agent bought the M&S coffee at £1.20, the best value per kilo, and kept the cheaper option elsewhere. B gave the cheapest basket: £7.50. The price rule also kept out M&S Ready Salted and Cheddar & Onion multipacks, at 1.31× the typical price per bag.
- **Setting C put Ocado's own brand first, and the agent took it every time it was offered.** That was 4 of 4 across two runs, for a £10.40 basket.
- **Removing paid placements alone made the customer pay more.** Before Parley could find own-brand products, setting C had nothing to put first. Without the INEOS banner, the agent's laundry pick became the £7.50 standard, and the basket rose to £16.45.

The Claude Sonnet 5.5 buyer, under setting A, made the same choices as GPT except for the cereal (Kellogg's Corn Flakes, £12.45 in all).

Result 3: promotions reach agents again, and win on value

In part 1, banners never reached the agents' choices. Through Parley, sponsored products became options an agent could weigh, with a label saying why each was shown. Agents took them when the offer was good value for the brief, and not otherwise.

Brief and buyer	Promoted options shown	Chosen by the buyer agent	What was chosen
03, 8 runs, setting A (value-seeking)	60	11	INEOS laundry liquid (search banner) 8 times, Shreddies (search banner) 3 times
02, quality-first buyer	18	4	Filippo Berio olive oil (homepage campaign), a featured Montepulciano, featured Lurpak butter, M&S organic double cream (Ocado offer)
01, Claude buyer	24	3	Filippo Berio olive oil (homepage campaign), Knorr stock cubes (search banner), a Warburtons gluten-free loaf (Ocado offer)
04, homepage campaign items	19	1	innocent orange juice (homepage campaign)

- **Value decides, not prominence.** A value-seeking agent compares price per unit and takes the promotion only when it wins. INEOS at half price beat the standard laundry liquid in every run. The homepage campaign's Nescafé Gold, at about 1.25 times the standard jar's price per kilo, was offered in every run and never taken.
- **As in part 1, every sponsored pick was also on offer.** INEOS was half price (£3.50, was £7) and Shreddies was on an introductory price (£2.50, rising to £3.50).
- **Quality-first buyers take dearer promotions.** The dinner-party buyer chose promoted products for 4 of its 17 items, including Lurpak butter and a featured Montepulciano that a value buyer would have passed over on price. For such a buyer the price rule allows options up to 2× the typical price.

- **The price rule keeps the options worth showing.** In each four-item run it kept 11–13 promoted products out, for example Doritos at 1.8× the typical price per bag. An early run that showed every promotion (32 options) saw none chosen; options the agent will never take only cost time and money.

The advertiser report on the dashboard counts these live: options shown, chosen, kept out and in the final basket, by placement and by product. That is the measurement a retail-media team would sell against.

What went wrong, and what was fixed

Most of the work was finding where a grocer's agent goes wrong on a real site, then moving the fix from the model's instructions into code.

Problem seen	Run	Fix in code
"5 bananas" arrived as quantity 5, so Parley added 5 packs of 5 (25 bananas). The buyer agent caught it, but the line could not be added again after the summary.	S-2e7e0faf, 5 Oct	A pack of several needs the buyer to confirm the quantity first; a correction after the summary reopens shopping
Settings B and C had little of the chosen brand to show: M&S and Ocado's own brand were mostly missing from the first 24 results, and M&S packs in grams couldn't be compared with a multipack counted in items	S-664baa71, S-d9f5a131, S-f42d3efe, S-7162c020; 4–5 Oct	The item's search with Ocado's Brands filter; the closest equivalent in the same format first, then best value; products of a different kind (a laundry cleanser for laundry liquid) dropped
A banner for a Diwali range led to a search for "Diwali bananas"	S-ed3e2f7f, 2 Oct	Banners are matched by brand to the item's results, and banner words are never searched
Toilet tissue was added against the oat milk line	S-8780fd8a, 2 Oct	A product must come from that item's own search; a correction tool undoes a wrong add
Every promotion was shown (32 options): none chosen, and the run was slow	S-ed3e2f7f, 2 Oct	At most 2 promoted options per item, plus the price rule

Problem seen	Run	Fix in code
A long conversation cost \$2.39 for one basket	S-8780fd8a, 2 Oct	Each turn starts from a short state written by code, not the whole history: about \$0.20 for four items now

In the bananas run the buyer agent stopped without confirming. It told the customer exactly what had happened, with the basket total with and without bananas. A customer's agent that checks the grocer's agent is part of the design, not a failure of it.

One thing stays variable: Parley's own model chooses the standard product, and that choice can differ between runs. The cereal standard was Kellogg's Corn Flakes in some runs and M&S Bran Flakes in others.

Part 1 and part 2 side by side

The same four-item list, browsed by agents in part 1 and shopped through Parley in part 2, gives these contrasts.

Measure	Part 1: agents browsing the sites	Part 2: agents shopping through Parley
Banner picks	0 of 32 (four stores)	11 of 32 (8 runs on Ocado, setting A)
Sponsored picks	6 of 32, all discounted	11 of 32, all discounted
Extra searches	about 12–14 at Ocado for the 15-item list	0 for the four-item list; 1–3 on longer lists, all by Parley
Who was shopping	no site checked, in 16 runs	agent identity and customer permission checked on every call
Spend limit	the agent's own judgement	the customer's cap, enforced in code and in the browser
Stopped before payment	yes, because the agents were told to	yes, by design: no checkout tool, no scope, pages blocked
What the grocer controls	nothing the agent can read	the order, labels and limits of every option shown

At Ocado itself, the baskets compare like this. In part 1 each agent chose from Ocado's ordinary pages; in part 2 the GPT buyer chose from Parley's options under setting A (S-82bfde15).

Item	Part 1: Claude agent	Part 1: ChatGPT agent	Part 2: through Parley, setting A
Crisps	Sensations Thai Sweet Chilli (featured, discounted)	McCoy's Flame Grilled (featured, discounted)	Walkers Ready Salted 6-pack, £1.50 (standard product)
Instant coffee	Douwe Egberts Pure Gold	Nescafé Original 190g (featured, discounted)	Nescafé Original 190g, £5.00 (Ocado offer, discounted)
Laundry liquid	Fairy Non Bio Sensitive 72W (discounted)	Fairy Non Bio Sensitive 72W (discounted)	INEOS Non Bio 1L, £3.50 (search banner)
Cereal	Crunchy Nut (featured, discounted)	Corn Flakes 450g	M&S Bran Flakes, £1.30 (standard product)

- **Browsing agents shopped differently from each other.** The two agents in part 1 agreed on only one item, the laundry liquid. Through Parley, the crisps, coffee and laundry liquid were the same in all 8 setting A runs, and the cereal was one of three: M&S Bran Flakes, Shreddies Peanut Butter or Kellogg's Corn Flakes.
- **Banners reached the agent only through Parley.** Neither browsing agent chose a banner product. Through Parley, the sponsored INEOS banner replaced Fairy in every setting A run, and the Shreddies banner was chosen in 3 of them.
- **Promotions still had to be good value.** In part 1, the products the agents did pick from promoted places were all discounted. Through Parley, a promoted product had to pass the price rule before the agent saw it, so the dearer ones never reached it.

This is a comparison of two exploratory studies, not a controlled experiment. Part 1 covered four stores and the advertising test of 28 September. Part 2 covers one store, a test account and early October. Placements rotate daily.

What this means for retailers

The choice is not only between letting agents browse and blocking them. With its own agent, a store decides how it meets customers' agents: what to show, in what order, within what price limits, with what labels, and which data to share.

- **Control over what is shown and sold.** One file stated the retailer's priority (paid campaigns, a brand partner or its own brand), and agents' baskets followed. The trade-offs became visible too: removing paid placements made one basket £5.15 dearer.

- **Control over data.** The agent gets answers, not access. A retailer decides which facts an agent may see, such as prices, stock and offers, and keeps the rest inside its own systems.
- **Knowing who is shopping.** Every call carries the agent's identity and the customer's permission. The store can then serve, limit or refuse agents on its own terms, rather than blocking them all.
- **A new form of marketing.** Agents compare value, so promoting to them means relevant, well-priced offers. It can be measured placement by placement: options shown, chosen, kept out and bought.

What a real version would do differently

Parley is an ordinary prototype, and most of its limits come from working off the public website. A retailer building this properly would change a great deal:

- **Real data instead of web pages.** Live stock by postcode, member prices, substitution and delivery rules and, with consent, a customer's usual items. With that, the agent could answer far more and choose far better.
- **Speed.** Four items take 4–5 minutes here because Parley drives a browser. Connected to the retailer's own catalogue and basket, it would take a fraction of that.
- **Cost.** Parley uses a language model, at about \$0.20 per four-item shop. Agents talking to agents don't need text written for people, so most of the exchange could be structured, rule-based responses with little or no model cost.
- **Far more than advertising.** Substitutions, allergens, age checks, delivery slots, minimum spends, returns, loyalty, fraud and rate limits are all rules a retailer would want to apply to agents. Each fits the same design.
- **A standard way in.** Customer agents need to find and trust a store's agent. A shared standard, such as a discovery file, UCP or an API, would decide whether they use it.

Follow-up: with and without Parley (7 October)

After the main runs, each buyer agent also shopped the four-item list on its own, browsing Ocado's website. Each basket is set against a run of the same agent through Parley under setting A.

Item	GPT on its own	GPT through Parley	Claude on its own	Claude through Parley
Crisps	Walkers Ready Salted 6-pack, £1.50	Walkers Ready Salted 6-pack, £1.50	Doritos Cool Original 180g, £1.75	Doritos Cool Original 180g, £1.75 (Ocado offer)
Instant coffee	Nescafé Original 190g, £5.00	Nescafé Original 190g, £5.00 (Ocado offer)	Nescafé Azera Americano 140g, £5.50	Nescafé Original 300g, £7.25 (featured)
Laundry liquid	Fairy Non Bio Sensitive, 50 washes, £7.50	INEOS Non Bio 1L, £3.50 (search banner)	Persil Non Bio, 53 washes, £7.00	Persil Bio, 53 washes, £7.00
Cereal	Weetabix 24 pack, £3.00	Shreddies Peanut Butter 580g, £2.50 (search banner)	Kellogg's Corn Flakes 670g, £2.85	Kellogg's Corn Flakes 670g, £2.85
Basket	£17.00	£12.50	£17.10	£18.85

- **On their own, the agents shopped sensibly, but Ocado had no part in it.** Ocado did not know an agent was shopping and did not shape what it saw, and the £30 budget relied on the agent's own judgement. GPT searched twice for the cereal.
- **Through Parley, the day's promotions decided what was offered.** On 5 October the search banners included INEOS and Shreddies. GPT took both, and 3 of its 4 items were promoted products. By 7 October the banners had changed to Shredded Wheat (homepage and search), Bold, Kenco and Nescafé Gold. Claude was shown 7 promoted options and passed on every banner. It chose the Doritos on an Ocado offer and the Nescafé jar marked best value per kilo. The price rule kept 4 dearer promotions from reaching it.
- **The agent often chose the same product either way.** Claude picked the same crisps and cereal with and without Parley. The difference is that through Parley, Ocado decided what it was shown and could see what it chose.

This is one run per agent and condition, and none of these runs are counted in the results above. The two Claude runs were on 7 October, at the same prices. The GPT run through Parley is the recorded setting A run from 5 October (S-dc91b4be), so its prices and promotions may differ from those in the GPT run on its own. The Claude run through Parley (S-b98a6fe9) used Parley after a 6 October fix: a product the buyer picks from an open question now still goes through the grocer's options.

Method and limitations

This is an exploratory prototype study, not a statistical one, so results are given as counts.

- **Runs.** 15 completed runs of the four-item list (3–5 October 2026): 8 under setting A, 4 under B and 3 under C, 14 with the GPT buyer and 1 with the Claude buyer. There were also 4 runs of the longer briefs, plus development runs that found the faults above.
- **Agents.** The buyer agents were GPT-6.1 Sol and Claude Sonnet 5.5, run from my own scripts with written briefs; they are not consumer products such as Muse. Parley used Claude Sonnet 5.5 to talk with the buyer agent and carry out the steps, while the retailer's rules ran in code. Model costs are from each provider's published prices.
- **Store and account.** Ocado only, live, on a test account with no delivery address, so Ocado showed its generic homepage and campaigns. Prices, offers and placements rotate daily.
- **Scoring.** Scorecards are generated automatically from both agents' logs. Item correctness is a human check. Delivery slots, checkout and payment were out of scope by design.
- **The grocer's view is simulated.** Parley reads only what Ocado shows any visitor. The settings are my illustrations of grocer choices, not Ocado's or M&S's policies, and the "partner brand" setting is a hypothetical.
- **Terms.** This ran as small-scale research on my own test account, with no bulk collection, no bot checks bypassed and no orders placed. It is not affiliated with, endorsed by or built with Ocado or M&S.

Runs and sources

Each run has a session ID, which this report uses to refer to it. The code, run logs and scorecards are not published with the report, but are available on request.

Part 1: Abdul Wadood, *When AI Agents Do the Weekly Shop*, 29 September 2026.