

When Relevance Becomes Work: Cognitive Load and Trust in the AI Personalization–Engagement Relationship

Tran Nguyen Gia Thinh¹, Nguyen Viet Trung Nhan², Nguyen Manh Ha^{3*}

¹HCMC University of Technology, Ho Chi Minh City, Vietnam

Email: giathinh4876@gmail.com

²High School for Gifted Students, Hanoi University of Science, Hanoi, Vietnam

Email: trungnhannguenviet@gmail.com

³Truong Thanh Viet Nam Group, Vietnam

Email: ha.nm1589@gmail.com

*Corresponding author

Abstract: Personalization is the most widely deployed AI consumer application, built on the logic that closer matching increases engagement. However, evidence reveals a downturn at high intensities. This paper proposes that personalization imposes a processing cost alongside relevance, and that trust moderates whether this cost drives withdrawal, positioning trust on the cost path rather than as a benefit-side mediator. Across an experiment manipulating personalization intensity ($N = 360$) and a two-wave observational panel ($N = 421$), engagement exhibited an inverted U-shape. Crucially, engagement fell at high intensity while perceived usefulness continued rising, locating the downturn on the cost side rather than a failure of relevance. This processing cost is highly asymmetric, driven overwhelmingly by presentation-specific extraneous load, while a transparency control moved no load indicators. Furthermore, cognitive load translated into disengagement significantly less among higher-trust respondents (+.112 experimentally; +.188 in the panel), attenuating from $-.436$ at low trust to $-.061$ at high trust. The position of the engagement optimum did not move with trust. Trust therefore operates as a defensive shield against processing costs rather than extending users' tolerance for dense personalization. The paper notes the load path accounts for a minority of the downturn as its principal limitation.

Keywords: AI personalization; Cognitive load; Consumer engagement; Trust in algorithms; Two-wave design; Experiment

1. Introduction

Personalization driven by artificial intelligence is the default configuration of consumer platforms, and systematic reviews identify it as the most widely deployed application of the technology (Peltier et al., 2024). The operating logic is monotone: closer matching produces more engagement. Evidence against that logic has accumulated. Personalization intensity exerts an inverted U-shaped effect on consumer attitudes (Sun et al., 2025); algorithmic disclosure beyond a moderate level reduces adoption (Ngo, 2025); and choice research has long documented the same nonlinearity for assortment size (Iyengar & Lepper, 2000). What none of this work supplies is a mechanism for the downturn that is estimated against the benefit path inside one model.

Two gaps follow. First, where the cost of personalization has been specified it has been specified as a privacy cost, so the processing cost of a denser stream has not been measured for AI personalization specifically. Second, trust has been positioned almost uniformly on the benefit side, as an antecedent or a mediator of acceptance, and not as a condition on whether a cost already incurred is translated into withdrawal.

Cognitive load is proposed here as the counterbalancing path. A denser stream of matched items presents more material to filter, compare and discard, and that filtering is work. Because the instrument separates intrinsic from extraneous load, the design can locate where any cost falls: on the difficulty of the decision, or on the presentation of the stream.

Trust is where the paper departs from existing specifications. In the algorithmic response literature the behavioural question is whether a recommendation is accepted or re-verified (Dietvorst et al., 2015, 2018; Logg et al., 2019), and re-verification is what consumes resources. Reading trust as willingness to accept without checking therefore places it not on the benefit path, where the literature has generally placed it, but as a condition on the cost path: the same load should cost a trusting user less.

The paper contributes in three respects. First, and principally, it specifies trust as a moderator of the cost path rather than as a benefit-side antecedent or mediator, and tests that specification in an experiment and a two-wave panel. Second, it manipulates personalization intensity directly and measures cognitive load with three instruments, which localises the cost to the presentation component rather than to the decision. Third, it crosses a transparency control with intensity, which separates the cost of matched density from algorithmic salience in general.

2. Theoretical background and hypothesis development

The stimulus–organism–response framework (Mehrabian & Russell, 1974; Vieira, 2013) supplies the structure: personalization intensity as stimulus, two internal states of opposite valence as organism, and engagement as response, measured on the three-dimensional scale of cognitive processing, affection and activation (Hollebeek et al., 2014). The benefit path is not in dispute and is included only as the comparison against which the cost path is estimated: perceived usefulness is an established mediator of technology response (Davis, 1989; Venkatesh et al., 2003), and more closely matched content is more useful. The cost path is the contribution. Cognitive load comprises perceived task difficulty and invested mental effort (Paas, 1992) and separates into intrinsic,

extraneous and germane components (Sweller, 1988; Sweller et al., 2019; Leppink et al., 2013), a distinction that matters here because it localises where any cost falls.

H1. Personalization intensity raises consumer engagement indirectly through perceived usefulness.

2.1. The cost path

Where the cost of personalization has been examined, it has almost always been framed as data-related risk. Covert collection reverses the effect of personalization on click-through intentions through feelings of vulnerability (Aguirre et al., 2015), trust conditions whether personalization is effective at all (Bleier & Eisenbeiss, 2015), and recent work treats perceived intrusiveness as the operative cost (Chen & Cheng, 2026). A processing cost is a different object. A denser stream of matched items presents more material that must be filtered, compared and discarded, and that filtering is work whether or not the user regards the matching as intrusive. Evidence that the two are distinct is available in adjacent settings: interface complexity raises measured load in electronic commerce (Schmutz et al., 2009), and a recent review identifies the metacognitive demand generative systems impose on their users (Tankelevitch et al., 2024). What has not been established is whether AI personalization intensity itself imposes such a cost, how large it is beside the benefit it delivers, and which component of load carries it.

H2. Personalization intensity raises cognitive load, which in turn lowers consumer engagement.

H3. The total effect of personalization intensity on engagement is nonlinear, rising at low to moderate intensity where the benefit path dominates and falling at high intensity where the cost path dominates.

2.2. Trust as the condition on the cost path

The hypothesis that distinguishes this study concerns the position of trust. In the algorithmic response literature the behavioural question is whether a recommendation is accepted or re-verified: users abandon an algorithm after observing it err (Dietvorst et al., 2015), rely on it more when permitted to adjust its output (Dietvorst et al., 2018), and prefer algorithmic to human judgement under some conditions (Logg et al., 2019). Re-verification is what consumes resources. Reading trust as willingness to accept without checking (Mayer et al., 1995; McKnight et al., 2011) therefore places it not on the benefit path but as a condition on the cost path: the same load should cost a trusting user less, because less of it is spent confirming what the system proposes.

This is a testable departure from existing specifications, in which trust is the dependent variable (Li et al., 2024), a mediator (Ngo, 2025; Teepapal, 2025, 2026), or a moderator of an earlier link than the one specified here (Chen & Cheng, 2026), while the moderator is something other than trust altogether (Dai et al., 2025; Hassan et al., 2025). Trust is measured rather than manipulated in both studies, so what the design establishes is a conditional association: the load-engagement slope differs across levels of trust. The hypotheses are worded accordingly.

Trust is measured rather than manipulated in both studies, so what the design establishes is a conditional association: the load-engagement slope differs across levels of trust. The hypotheses are worded accordingly, and Section 5.3 records the limitation.

H4. The association between cognitive load and engagement is conditional on trust in the recommendation algorithm, such that the negative association is weaker at higher levels of trust.

H5. Consequently, the downturn in H3 is located at a higher level of personalization intensity among users reporting higher trust in the algorithm. H5 is stated because it follows from H4 under a particular reading of how the moderation operates, and it is reported below as not supported; the distinction between conditioning the price of load and displacing the optimum is developed in Section 5.1.

Both studies are conducted with users of Vietnamese commerce and social platforms. The context functions as a boundary condition on magnitudes rather than as a contribution, and it is described with the samples in Sections 3.4 and 4.2.

3. Study 1: Experimental manipulation of personalization intensity

3.1. Design, ethics and procedure

Study 1 was a between-subjects experiment with a three by two design. Personalization intensity was manipulated at three levels, low, moderate and high, and algorithmic transparency at two, absent and present. Participants were assigned to conditions at random. Three levels of the focal factor were required rather than two because H3 specifies a nonlinear total effect, and a two-level manipulation cannot distinguish a downturn from a smaller positive effect.

Ethical approval was granted before any data collection began. Informed consent was obtained electronically from every participant before the preference-profile task, and no personally identifying information was retained.

3.2. Stimulus and manipulation

Participants browsed a mock feed on a commerce platform interface. Personalization intensity was manipulated by the proportion of items presented as matched to a preference profile participants supplied at the outset. Transparency was manipulated by the presence or absence of an explanation of why each item was selected. The match-signalling label was held constant across intensity conditions, so only the proportion of matched items varied with intensity and only the transparency factor varied whether an explanation accompanied that label.

Manipulation checks were administered for both factors, the personalization check across the three intensity levels and the transparency check between the absent and present conditions. Participants failing either check were retained in the primary analysis, and the analysis excluding those participants appears in Supplementary Table S3 as a robustness exercise.

Transparency was crossed with intensity as a design control rather than as a focal variable. It locates the design against the two nearest experimental studies, both of which manipulate transparency rather than intensity (Li et al., 2024; Ngo, 2025), and it holds

constant a second source of algorithmic salience that would otherwise be confounded with intensity. No hypothesis is stated for it, every analysis pools the two conditions, and its coefficients are reported once, in Section 3.6.

3.3. Measures

Cognitive load was measured by three instruments rather than one, because self-report alone diverges from other indicators in interactive media (Skulmowski & Rey, 2020; Brünken et al., 2003): the raw workload index (Hart, 2006; Hart & Staveland, 1988), the differentiated instrument separating intrinsic, extraneous and germane load (Leppink et al., 2013; Klepsch et al., 2017), and response latency on each choice.

The five resulting indicators are reported in Table 1. Perceived usefulness followed Davis (1989), trust followed McKnight et al. (2011) and engagement the three-dimensional scale of Hollebeek et al. (2014).

3.4. Sample and power

Participants were recruited from an online panel with screening and attention checks, with quotas set against population margins for age, sex and region; the achieved margins are reported in Supplementary Table S1. The sample was 50.6 per cent female and 48.9 per cent male, with 0.6 per cent declining to state; 37.2 per cent were aged 25 to 34, 21.4 per cent 35 to 44, 20.6 per cent 18 to 24, 15.0 per cent 45 to 54 and 5.8 per cent 55 or over; and 46.7 per cent were resident in the south, 35.3 per cent in the north and 18.1 per cent in the centre. Because interactions are estimated with substantially lower power than main effects (Aguinis et al., 2005; McClelland & Judd, 1993), and because the parameters of the original *a priori* calculation cannot be recovered, a sensitivity analysis is reported in its place. With $N = 360$, $\alpha = .05$ and five predictors, the design attains 80 per cent power against an interaction of $f^2 = .022$, corresponding to a standardised coefficient of approximately .14. The interaction actually observed was smaller than that ($\beta = .112$, $f^2 = .014$), so the achieved power for H4 in Study 1 was .62, and an experiment of this design would require $N \approx 555$ for 80 per cent power against an effect of the observed size. Study 1 is therefore reported as underpowered for H4 considered alone; Study 2, with $N = 421$ and a larger effect, attains power above .99 for the same test, and the two studies are read together for that reason.

3.5. Analysis

Hypotheses concerning the manipulated factor were tested by one-way analysis of variance on personalization intensity with planned contrasts, the two adjacent contrasts tested separately since H3 predicts opposite signs, with Welch tests alongside because the cell variances are unequal. Effect sizes are partial eta-squared and standardised contrasts on the pooled within-cell standard deviation. Mediation through cognitive load and through perceived usefulness was estimated in parallel with percentile bootstrap intervals over 5,000 resamples, and repeated separately on each adjacent segment since a linear coding cannot represent a curve that turns. The H4 moderation was estimated on standardised composites with perceived usefulness controlled. Analyses used Python 3.11 with statsmodels.

The turning point in H5 is more fragile than the contrasts that surround it, since a quadratic fitted to three cell means is exactly identified and its vertex is sensitive to sampling error in any one mean. It is therefore treated as an interpolated summary of three levels rather than as a parameter of a continuous dose-response function, bootstrap intervals over 2,000 resamples are reported, and the two vertices are model-implied quantities at plus and minus one standard deviation of trust rather than vertices of two subsamples.

3.6. Results

Achieved cell sizes ranged from 59 to 61 against a target of 60, giving 360 participants against the 348 required by the power calculation on the H4 interaction. Both manipulations moved their own check and nothing else: the personalization check rose from 2.33 to 4.30 to 5.84 across intensity levels, $F(2, 357) = 1003.86$, $p < .001$, and the transparency check moved from 2.07–2.24 to 5.75–5.82, $F(1, 358) = 3926.71$, $p < .001$, with no drift across intensity levels in either case. Check failure rates were 6.4 and 5.0 per cent with no concentration in any cell. Quotas, cell sizes and check detail appear in Supplementary Table S1.

The cost of personalization is asymmetric across the components of load, and the asymmetry is the result (Table 1). Extraneous load, the component attributable to how material is presented rather than to the decision itself, rose from 3.05 to 5.00 across the manipulated range, $F(2, 357) = 98.18$, $p < .001$, $\eta^2 = .355$, with steps of $d = 0.57$ and 1.20. The raw workload index tracked it, rising by approximately 1.6 pooled standard deviations, $F(2, 357) = 79.21$, $p < .001$, $\eta^2 = .307$, and response latency, the one indicator that is not self-reported, rose from 4.68 to 7.26 seconds, $F(2, 357) = 48.09$, $p < .001$, $\eta^2 = .212$. Intrinsic load, the component attributable to the difficulty of the decision, also rose, but barely: $F(2, 357) = 3.22$, $p = .041$, $\eta^2 = .018$, a total shift of $d = 0.31$ against $d = 1.77$ for extraneous load. Germane load did not move at all, $F(2, 357) = 0.36$, $p = .695$. Because the cell variances differ, Welch tests were also computed and return the same conclusions throughout.

The ratio rather than either figure alone is the informative quantity. Raising intensity did make the decision somewhat harder, and the paper does not claim otherwise; the claim is that the increase in task difficulty is negligible beside the increase in the cost of filtering the stream. The intrinsic effect is also unstable across specifications, returning $F = 1.93$ ($p = .146$) in the two-factor model and $F = 3.01$ ($p = .050$) with demographic covariates, whereas the extraneous effect exceeds $F = 84$ in every specification. The increments accumulate faster in the upper half of the range: workload rose by 8.5 points then 14.4, extraneous load by 0.62 then 1.33.

Table 1. Study 1: cognitive-load indicators by personalization intensity

Indicator	Low, mean (SD)	Moderate, mean (SD)	High, mean (SD)	F / r	p	ηp^2
Raw workload index (0–100)	39.13 (12.85)	47.59 (14.56)	61.98 (15.16)	79.21	< .001	.307
Intrinsic load (1–7)	3.36 (0.93)	3.61 (1.05)	3.68 (1.13)	3.22	.041	.018
Extraneous load (1–7)	3.05 (1.09)	3.67 (1.05)	5.00 (1.17)	98.18	< .001	.355
Germane load (1–7)	4.19 (1.03)	4.31 (1.32)	4.27 (1.14)	0.36	.695	.002
Response latency (seconds)	4.68 (1.73)	5.39 (2.07)	7.26 (2.44)	48.09	< .001	.212
Workload with extraneous load	—	—	—	r = .62	< .001	—
Workload with latency	—	—	—	r = .37	< .001	—
Extraneous load with latency	—	—	—	r = .41	< .001	—

Notes. *F* statistics are one-way analyses of variance on personalization intensity, $df = (2, 357)$, computed from the cell means, standard deviations and cell sizes reported in this table; ηp^2 is the corresponding partial eta-squared, and the estimated marginal means of the one-way model are the cell means shown. Welch tests that relax the equal-variance assumption are reported in Section 4.1. Five indicators from three instruments; the differentiated scale yields three subscales. The raw workload index follows Hart (2006) and Hart and Staveland (1988); the differentiated subscales follow Leppink et al. (2013) and Klepsch et al. (2017). Response latency is mean seconds to decision per item. Scales are reported in the original units. The final panel reports pairwise correlations among the global, presentation-specific and behavioural indicators.

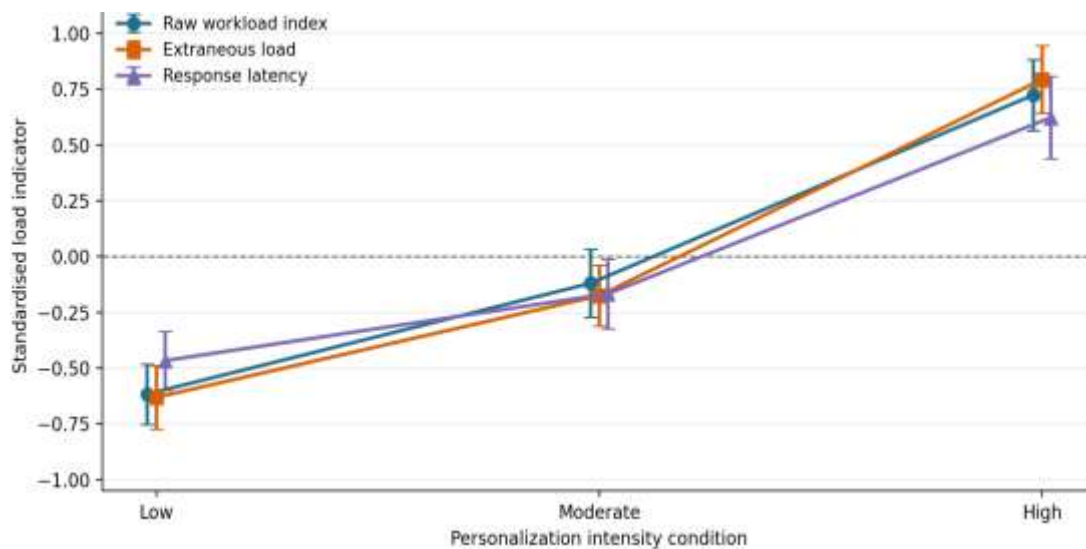


Figure 1. Study 1: three representative cognitive-load indicators across intensity conditions

Notes. Standardised means with 95 per cent confidence intervals. Three of the five indicators in Table 1 are plotted, representing the global, presentation-specific and behavioural indicator; standardisation permits a common vertical scale, and the unstandardised values appear in Table 1.

Engagement followed the ordering H3 predicts, and did so with unusual clarity (Table 2). The mean rose from 3.06 at low intensity to 6.06 at moderate and fell to 3.58 at high, $F(2, 357) = 154.97$, $p < .001$. The low-to-moderate contrast was +3.00 (95% CI [2.68, 3.33]) and the moderate-to-high contrast -2.48 ([-2.84 , -2.13]) on a seven-point scale, with a quadratic term of -2.74 ([-3.05 , -2.43]). The parallel mediation gave +0.49 ([0.25, 0.74]) through perceived usefulness and -0.22 ([-0.35 , -0.10]) through cognitive load, so the two paths carried opposite signs within a single specification, as H1 and H2 jointly require. Trust did not differ across the three conditions, $F(2, 357) = 0.20$, $p = .818$, which is what a measured moderator must show if the manipulation is not to have disturbed it.

One feature of the decomposition requires statement. Perceived usefulness continued to rise at the top of the range, from 4.44 at moderate intensity to 5.90 at high, and the indirect path through it remained positive across that segment (+0.48, [0.27, 0.73]). The downturn therefore occurred while the benefit path was still strengthening, which rules out the most obvious alternative account, that personalization simply stops being useful. The cost path ran in the predicted direction over the same segment (-0.20 , [-0.35 ,

–0.09]) but accounts for approximately eight per cent of a total fall of 2.48 scale points, with a direct effect of –2.76 that neither measured mediator explains; Section 5.3 states what this leaves unresolved.

Table 2. Study 1: engagement and the two competing indirect paths

Effect	Estimate	SE	95% interval	p
Contrast: low to moderate (H3)	+3.00	(0.17)	[2.68, 3.33]	< .001
Contrast: moderate to high (H3)	–2.48	(0.18)	[–2.84, –2.13]	< .001
Quadratic term (H3)	–2.74	(0.16)	[–3.05, –2.43]	< .001
Indirect via perceived usefulness (H1)	+0.49	(0.12)	[0.25, 0.74]	—
Indirect via cognitive load (H2)	–0.22	(0.06)	[–0.35, –0.10]	—
Total indirect	+0.27	(0.14)	[–0.01, 0.55]	—
Direct effect	–0.03	(0.17)	[–0.37, 0.32]	.877
Transparency main effect (control)	–0.16	(0.08)	[–0.31, –0.01]	.040
Transparency × intensity (control)	F(2, 354) = 1.16	—	—	.315

Notes. Engagement is on a 1 to 7 scale. The first three rows test H3, which predicts a positive low-to-moderate and a negative moderate-to-high contrast. Indirect effects are estimated in parallel with bias-corrected bootstrap intervals. The total indirect effect is the sum of two paths of opposite sign and is close to zero for that reason; its standard error is smaller than the two component standard errors would imply under independence, because the two paths share a predictor and an outcome and covary negatively.

We estimate the interaction between cognitive load and trust on engagement at +0.112 ([0.008, 0.217], $p = .035$) (Table 3). The load-engagement association was –0.317 ([–0.467, –0.168]) one standard deviation below the trust mean, –0.205 at the mean and –0.093 ([–0.238, 0.052]) one standard deviation above, where it is no longer distinguishable from zero. The penalty attached to a given level of load is therefore roughly three times larger among low-trust users than among high-trust users: trust attenuates the cost path rather than reversing it, which is what Section 2.4 predicts.

H5 is not supported. The implied turning points were 2.02 (bootstrap 95% CI [1.98, 2.06]) for low-trust respondents and 2.08 ([2.02, 2.14]) for high-trust respondents, a difference of +0.06 ([–0.01, 0.13]) that includes zero (Figure 2). The quadratic-by-trust coefficient that generates the displacement was +0.285 (SE 0.154, $p = .065$). Both vertices fell inside the experimental range and neither interval required truncation, so the null is informative rather than merely imprecise: the location of the optimum is set by the main effects, which are estimated with very large t values, and trust does not move it. The finding is that trust conditions the price of load without changing the point at which the stream becomes too dense. Section 5.1 develops the distinction, which is sharper than the one H5 proposed.

Table 3. Study 1: trust as a condition on the cost path, and implied turning points

Quantity	Estimate	SE	95% interval
Cognitive load → engagement	–0.205	(0.053)	[–0.310, –0.101]
Trust → engagement	+0.111	(0.051)	[0.011, 0.210]
Cognitive load × trust (H4)	+0.112	(0.053)	[0.008, 0.217]
Conditional association at low trust (–1 SD)	–0.317	(0.076)	[–0.467, –0.168]
Conditional association at mean trust	–0.205	(0.053)	[–0.309, –0.101]
Conditional association at high trust (+1 SD)	–0.093	(0.074)	[–0.238, 0.052]
Turning point, low trust (H5)	2.02	—	[1.98, 2.06]
Turning point, high trust (H5)	2.08	—	[2.02, 2.14]
Difference in turning points (H5)	+0.06	—	[–0.01, 0.13]

Notes. Moderation coefficients are standardised; turning points are on the experimental intensity scale, coded 1 for low, 2 for moderate and 3 for high. Turning-point intervals are bootstrap percentile intervals; the quadratic is exactly identified over three levels, so the vertex is treated as an interpolated summary rather than as an estimated parameter. Trust is a continuous measured variable, so the two turning points are model-implied quantities evaluated at plus and minus one standard deviation of trust, not vertices of two subsamples, and all quantities describe conditional associations rather than causal effects.

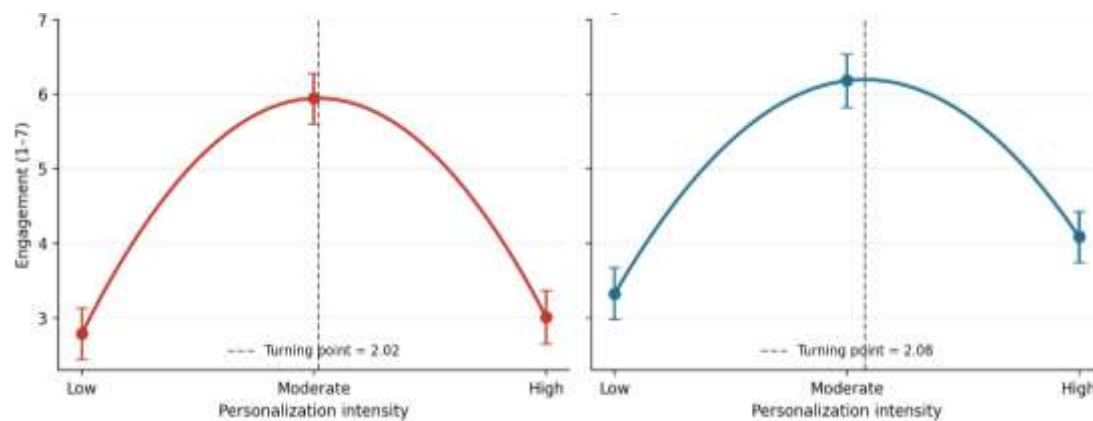


Figure 2. Study 1: engagement across personalization intensity, by trust in the recommendation algorithm

Notes. Model-predicted marginal means with 95 per cent confidence intervals. Both

panels share a single vertical scale. The left panel plots predictions one standard deviation below the trust mean and the right panel one standard deviation above, matching the conditional associations in Table 4. Trust is continuous, so the panels are model-implied rather than subgroup means. Quadratic curves are descriptive summaries of three experimental levels. Dashed lines mark the bootstrap turning points reported in Table 3; the two lie at 2.02 and 2.08 and are therefore close to coincident, which is the visual form of the null result for H5. The difference between the panels is the steepness of the descent, not the position of the peak.

The transparency control provides discriminant evidence. Although the manipulation was strong, it did not move a single cognitive-load indicator: $F(1, 354) = 0.12$ for raw workload, 1.35 for intrinsic load, 0.21 for extraneous load, 0.03 for germane load and 1.47 for response latency ($p = .247$ to $.855$), with no interaction with intensity on any indicator (all $p \geq .692$). It did lower engagement slightly, by -0.16 standardised units ($[-0.31, -0.01]$, $p = .040$), without interacting with intensity there either ($F(2, 354) = 1.16$, $p = .315$). Two sources of algorithmic salience therefore behave differently: making the algorithm visible reaches engagement by a route that does not pass through processing cost, while increasing the density of matched content reaches it through a cost that is large, presentation-specific and moderated by trust. That route is not identified here, and the obvious candidates, privacy concern and reactance, were not measured.

4. Study 2: Two-wave observational panel study

4.1. Purpose

Study 1 establishes whether the mechanism operates when personalization intensity is set by the experimenter. Study 2 asked whether it operates in ordinary use, where intensity is determined by the platform and by the user's own history. The two studies answer different questions and neither substitutes for the other.

4.2. Sample and measures

Respondents were recruited from the same online panel provider as Study 1, with quotas against population margins for age, sex and region, and eligibility required at least weekly use of a domestic commerce or short-video platform, a criterion every respondent met. The wave-one sample was 50.2 per cent female and 48.7 per cent male, with 1.2 per cent declining to state; 30.0 per cent were aged 25 to 34, 24.8 per cent 35 to 44, 20.0 per cent 18 to 24, 15.8 per cent 45 to 54 and 9.4 per cent 55 or over; and 40.4 per cent were resident in the south, 39.0 per cent in the north and 20.6 per cent in the centre. Platform use was 37.9 per cent domestic commerce only, 27.1 per cent short-video only and 35.0 per cent both. Wave one measured perceived personalization, cognitive load, perceived usefulness, trust and a baseline measure of engagement; wave two measured engagement only. All constructs used the scales in Section 3.3, with cognitive load measured by the differentiated instrument alone, since response latency cannot be recorded outside a controlled interface.

4.3. Design and the limits of two waves

Predictors, mediators and a baseline measure of engagement were collected at wave one and the outcome at wave two, three weeks later. The separation addresses common method variance by removing the mediator and the outcome from a single response occasion (Podsakoff et al., 2003, 2024), and the wave-one engagement control means the coefficients describe change rather than level.

The design does less than a longitudinal mediation design would. Perceived personalization, cognitive load, perceived usefulness and trust were all measured at wave one, so two waves separate the mediators from the outcome but not from the predictor: the personalization-to-mediator links rest on contemporaneous measurement and only the mediator-to-engagement links are temporally ordered. Establishing the full ordering would require three waves.

Two responses are available within a two-wave instrument, and both are used. Engagement was measured at wave one as well as wave two and enters as a control, so the coefficients describe change rather than level. And the mediators are measured in the

same wave as the predictor, so the first stage of each indirect path is reported as a contemporaneous association carrying no temporal ordering.

Attrition between waves was examined by comparing retained and lost respondents on five wave-one characteristics: perceived personalization, cognitive load, perceived usefulness, trust and engagement. Listwise deletion was applied, since the wave-two outcome is missing by design for non-completers and no imputation model is available for it.

4.4. Common method bias

Three procedures were applied beyond temporal separation: an unrelated marker variable with the correlation-adjustment procedure (Lindell & Whitney, 2001), full collinearity variance inflation factors (Kock, 2015), and the design controls Podsakoff et al. (2024) identify as effective, applied at instrument construction rather than assessed after the fact.

4.5. Latent interaction estimation

The latent interaction was estimated by latent moderated structural equations (Klein & Moosbrugger, 2000), with a matched-pair product-indicator specification reported alongside as a check on the estimator rather than as an independent test.

4.6. Results

Of 520 wave-one respondents, 421 completed wave two, a retention rate of 81.0 per cent, and retained and lost respondents did not differ on any of five wave-one characteristics ($p = .220$ to $.877$).

The measurement model met every conventional criterion: composite reliabilities from $.90$ to $.92$ for the focal constructs, average variance extracted from $.69$ to $.75$, full collinearity variance inflation factors from 1.22 to 1.53 against a threshold of 3.3 (Kock, 2015), and a largest heterotrait-monotrait ratio of $.489$ against the $.85$ criterion. The marker variable correlated between $-.020$ and $+.059$ with the focal constructs. Item-level detail appears in Supplementary Table S2.

The sign pattern of Study 1 was reproduced in ordinary platform use, and the moderation was stronger there than in the experiment (Table 4). Perceived personalization was positively associated with perceived usefulness at $+.426$ and with cognitive load at $+.280$; usefulness was positively associated with wave-two engagement at $+.398$ ($[.318, .478]$) and load negatively at $-.248$ ($[-.328, -.169]$), with wave-one engagement controlled. Because personalization and both mediators were measured in the same wave, the first two coefficients are contemporaneous associations and carry no temporal ordering. The indirect effects were $+.151$ ($[.108, .197]$) through usefulness and $-.068$ ($[-.105, -.038]$) through load. The latent interaction was $+.188$ ($[.111, .264]$, $p < .001$) under the matched-pair product-indicator specification and $+.194$ under composites, with $R^2 = .377$ for wave-two engagement.

The two indirect paths differed more in magnitude in Study 2 than in Study 1, and a direct effect remained ($+.096$, $[.008, .183]$). The experiment sets intensity at three fixed levels including one chosen to be high, whereas the field measures whatever intensity respondents happen to experience, so a field sample whose modal respondent sits below the vertex would produce this difference without any change in the mechanism.

Table 4. Study 2: structural model and latent interaction

Path / model quantity	Product-indicator	Composite	95% interval
Personalization → perceived usefulness	$+.426$	$+.426$	$ [.34, .51]$
Personalization → cognitive load	$+.280$	$+.280$	$ [.19, .37]$
Perceived usefulness → engagement (W2)	$+.398$	$+.402$	$ [.318, .478]$
Cognitive load → engagement (W2)	$-.248$	$-.250$	$ [-.328, -.169]$
Trust → engagement (W2)	$+.109$	$+.111$	$ [.025, .193]$
Cognitive load × trust → engagement (H4)	$+.188$	$+.194$	$ [.111, .264]$
Indirect via perceived usefulness	$+.151$	—	$ [.108, .197]$
Indirect via cognitive load	$-.068$	—	$ [-.105, -.038]$
Direct effect	$+.096$	—	$ [.008, .183]$
R^2 for engagement (wave 2)	$.377$	—	—

Notes. Structural coefficients are standardised; the log-likelihood and information criterion are not. The latent interaction is estimated by latent moderated structural equations (Klein & Moosbrugger, 2000) with the product-indicator estimate alongside. Fit quantities are reported in the form each estimator supports and are not combined into a single figure. All paths control for wave-one engagement.

The interaction is stable. Across ten robustness exercises reported in Supplementary Table S3, covering estimator choice, sample restrictions, marker-variable adjustment and four alternative operationalisations of load, it ranges from $+.069$ to $+.230$ and retains its sign throughout. The weakest cell is Study 1 estimated on response latency ($+.069$, $p = .166$): the moderation is present under every self-report measure of load but not under the behavioural one, which is the material qualification on H4.

5. General discussion

5.1. Theoretical implications

Personalization intensity raises cognitive load, and load carries a negative indirect effect on engagement alongside the positive path through perceived usefulness. Both studies reproduce that sign pattern, and in Study 1 the two contrasts change sign across the manipulated range as H3 requires. The net effect therefore depends on where a platform sits on the curve. Two features locate the downturn on the cost side. It occurred while perceived usefulness was still rising, from 4.44 at moderate intensity to 5.90 at high, so it cannot be attributed to a failure of relevance. And the cost is asymmetric across the components of load, which is the first of the two findings this paper rests on.

Extraneous load accounts for approximately twenty times the variance that intrinsic load accounts for ($\eta^2 = .355$ against $.018$). Intrinsic load did rise, and we do not claim otherwise: a denser feed makes the decision marginally harder as well as costlier to process. But had the burden fallen there, the finding would have been that personalization complicates choice, which the overload literature already supports (Jacoby et al., 1974). That it falls almost entirely on the presentation-specific component identifies the cost as a processing cost imposed by the stream rather than by the decision, which is the construct the privacy-framed literature has not measured. The transparency control corroborates the boundary from the other side: a manipulation that made the algorithm maximally visible moved no load indicator at all, so algorithmic salience in general does not produce this cost; the density of matched content does.

The second finding concerns trust, and it is sharper than the hypothesis that generated it. Trust conditions the price of cognitive load without changing the quantity of personalization a user can absorb. The load-engagement association attenuated from -0.317 at low trust to -0.093 at high trust in the experiment and from -0.436 to -0.061 in the panel, where the interaction reached $+1.188$; yet the position of the engagement optimum was identical across trust levels, at 2.02 against 2.08, a difference of $+0.06$ ($[-0.01, 0.13]$). Trust is a shield, not an extension. It does not let a platform deliver more matched content before the return turns, but it substantially reduces what the user withdraws once the cost has been incurred. The theoretical consequence is that trust and cognitive load may stand in a substitution relation rather than being independent constructs entered side by side, with trust the thing that allows a user to stop checking. That reading remains an interpretation, since re-verification behaviour was not measured.

The practical consequence is not the one the mediating specification implies. Under mediation, raising trust raises engagement generally. Under the conditional specification estimated here, raising trust is valuable chiefly where load is already high, and reducing load is an alternative to it rather than a complement. A platform choosing between investing in explanation and investing in feed simplification faces a real trade-off under the second reading and none under the first.

5.2. Practical implications

Three implications follow for platform design. On intensity, an interior optimum implies that further matching returns nothing beyond a point, and the point does not move with how much users trust the system. A recommendation team should run a staged reduction in match density for any segment whose extraneous-load index exceeds the sample median, triggered by the load measurement rather than by a fall in engagement, since engagement declines only once the optimum has been passed. On trust, the finding defines what a trust-building programme can buy. It cannot buy headroom; what it buys is a shallower descent once the feed is already dense, with the load penalty falling by roughly two thirds between low- and high-trust users. Trust investment is a mitigation instrument for segments already past the optimum, not a licence to push intensity further. On explanation, making the algorithm more visible reduced no component of load here, so explanation features should not be budgeted as an overload remedy.

5.3. Limitations

Three limitations bear most heavily on these conclusions. First, the measured mediators account for a small share of the downturn. Over the moderate-to-high segment engagement fell by 2.48 scale points while the indirect path through cognitive load carried -0.20 of that fall, approximately eight per cent, and the direct effect of -2.76 is not explained by either measured mediator. The study therefore establishes that a processing cost exists, that it is specifically located and that trust moderates it; it does not establish that this cost is what produces the downturn. Candidates for the remainder, including perceived intrusiveness, reactance and privacy concern, were not measured, and identifying them is the first priority for replication. Second, trust is measured rather than manipulated, so H4 is supported as a conditional association: respondents reporting higher trust may differ systematically in ways that produce the same interaction, and the re-verification mechanism proposed in Section 2.4 was not observed in either study. Third, the two-wave design orders the outcome after the predictors but measures personalization and both mediators in the same wave, so the first stage of each indirect path carries no temporal ordering; a three-wave design would be required to establish precedence. Two further points are recorded rather than developed: H5 was not supported, so the claim that trust displaces the optimum is withdrawn rather than qualified, and both samples are drawn from one national market.

5.4. Conclusion

Taken together, the two studies establish that AI personalization imposes a processing cost alongside the relevance it delivers, that the cost is specific to the presentation component of load rather than to the difficulty of the decision, and that trust in the algorithm conditions its price without moving the point at which the return turns. The contribution is the position assigned to trust: the literature has treated it almost uniformly as a benefit-side antecedent or mediator, and where it has recently been specified as a moderator the moderated link is an earlier one (Chen & Cheng, 2026). Specifying it as a condition on the cost path yields a different prescription for a platform deciding where to invest, and the transparency control supports that reading from the other side. Three

extensions follow. Measuring re-verification directly would convert the mechanism from an interpretation into a test. Measuring perceived intrusiveness and reactance alongside load would identify what carries the remainder of the downturn. And manipulating trust rather than measuring it would convert the conditional association into a causal claim.

References

- Aguinis, H., Beaty, J. C., Boik, R. J., & Pierce, C. A. (2005). Effect size and power in assessing moderating effects of categorical variables using multiple regression: A 30-year review. *Journal of Applied Psychology*, 90(1), 94–107. <https://doi.org/10.1037/0021-9010.90.1.94>
- Aguirre, E., Mahr, D., Grewal, D., de Ruyter, K., & Wetzels, M. (2015). Unraveling the personalization paradox: The effect of information collection and trust-building strategies on online advertisement effectiveness. *Journal of Retailing*, 91(1), 34–49. <https://doi.org/10.1016/j.jretai.2014.09.005>
- Bleier, A., & Eisenbeiss, M. (2015). The importance of trust for personalized online advertising. *Journal of Retailing*, 91(3), 390–409. <https://doi.org/10.1016/j.jretai.2015.04.001>
- Brünken, R., Plass, J. L., & Leutner, D. (2003). Direct measurement of cognitive load in multimedia learning. *Educational Psychologist*, 38(1), 53–61. https://doi.org/10.1207/s15326985ep3801_7
- Chen, C.-J., & Cheng, J. M.-S. (2026). When help feels creepy: A MOA model of AI smart-retail adoption via perceived intrusiveness and the moderating role of trust in AI. *Journal of Retailing and Consumer Services*, 90, Article 104699. <https://doi.org/10.1016/j.jretconser.2025.104699>
- Dai, B., Zhu, J., & Zhu, X. (2025). The impact of AI-recommended content affordances on post-purchase intention in stockout substitution scenarios. *Behavioral Sciences*, 15(11), Article 1507. <https://doi.org/10.3390/bs15111507>
- Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13(3), 319–340. <https://doi.org/10.2307/249008>
- Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General*, 144(1), 114–126. <https://doi.org/10.1037/xge0000033>
- Dietvorst, B. J., Simmons, J. P., & Massey, C. (2018). Overcoming algorithm aversion: People will use imperfect algorithms if they can (even slightly) modify them. *Management Science*, 64(3), 1155–1170. <https://doi.org/10.1287/mnsc.2016.2643>
- Hart, S. G. (2006). NASA-Task Load Index (NASA-TLX); 20 years later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 50(9), 904–908. <https://doi.org/10.1177/154193120605000909>
- Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), *Human mental workload* (Advances in Psychology, Vol. 52, pp. 139–183). North-Holland. [https://doi.org/10.1016/S0166-4115\(08\)62386-9](https://doi.org/10.1016/S0166-4115(08)62386-9)
- Hassan, N., Abdelraouf, M., & El-Shihy, D. (2025). The moderating role of personalized recommendations in the trust–satisfaction–loyalty relationship: An empirical study of AI-driven e-commerce. *Future Business Journal*, 11(1), Article 66. <https://doi.org/10.1186/s43093-025-00476-z>
- Hollebeek, L. D., Glynn, M. S., & Brodie, R. J. (2014). Consumer brand engagement in social media: Conceptualization, scale development and validation. *Journal of Interactive Marketing*, 28(2), 149–165. <https://doi.org/10.1016/j.intmar.2013.12.002>
- Iyengar, S. S., & Lepper, M. R. (2000). When choice is demotivating: Can one desire too much of a good thing? *Journal of Personality and Social Psychology*, 79(6), 995–1006. <https://doi.org/10.1037/0022-3514.79.6.995>
- Jacoby, J., Speller, D. E., & Kohn, C. A. (1974). Brand choice behavior as a function of information load. *Journal of Marketing Research*, 11(1), 63–69. <https://doi.org/10.1177/002224377401100106>
- Klein, A., & Moosbrugger, H. (2000). Maximum likelihood estimation of latent interaction effects with the LMS method. *Psychometrika*, 65(4), 457–474. <https://doi.org/10.1007/BF02296338>
- Klepsch, M., Schmitz, F., & Seufert, T. (2017). Development and validation of two instruments measuring intrinsic, extraneous, and germane cognitive load. *Frontiers in Psychology*, 8, Article 1997. <https://doi.org/10.3389/fpsyg.2017.01997>
- Kock, N. (2015). Common method bias in PLS-SEM: A full collinearity assessment approach. *International Journal of e-Collaboration*, 11(4), 1–10. <https://doi.org/10.4018/ijec.2015100101>
- Leppink, J., Paas, F., Van der Vleuten, C. P. M., Van Gog, T., & Van Merriënboer, J. J. G. (2013). Development of an instrument for measuring different types of cognitive load. *Behavior Research Methods*, 45(4), 1058–1072. <https://doi.org/10.3758/s13428-013-0334-1>
- Li, Y., Deng, X., Hu, X., & Liu, J. (2024). The effects of e-commerce recommendation system transparency on consumer trust: Exploring parallel multiple mediators and a moderator. *Journal of Theoretical and Applied Electronic Commerce Research*, 19(4), 2630–2649. <https://doi.org/10.3390/jtaer19040126>
- Lindell, M. K., & Whitney, D. J. (2001). Accounting for common method variance in cross-sectional research designs. *Journal of Applied Psychology*, 86(1), 114–121. <https://doi.org/10.1037/0021-9010.86.1.114>
- Logg, J. M., Minson, J. A., & Moore, D. A. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes*, 151, 90–103. <https://doi.org/10.1016/j.obhdp.2018.12.005>
- Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of Management Review*, 20(3), 709–734. <https://doi.org/10.2307/258792>

- McClelland, G. H., & Judd, C. M. (1993). Statistical difficulties of detecting interactions and moderator effects. *Psychological Bulletin*, 114(2), 376–390. <https://doi.org/10.1037/0033-2909.114.2.376>
- McKnight, D. H., Carter, M., Thatcher, J. B., & Clay, P. F. (2011). Trust in a specific technology: An investigation of its components and measures. *ACM Transactions on Management Information Systems*, 2(2), Article 12. <https://doi.org/10.1145/1985347.1985353>
- Mehrabian, A., & Russell, J. A. (1974). *An approach to environmental psychology*. MIT Press.
- Ngo, V. M. (2025). Balancing AI transparency: Trust, certainty, and adoption. *Information Development*. Advance online publication. <https://doi.org/10.1177/02666669251346124>
- Paas, F. G. W. C. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. *Journal of Educational Psychology*, 84(4), 429–434. <https://doi.org/10.1037/0022-0663.84.4.429>
- Peltier, J. W., Dahl, A. J., & Schibrowsky, J. A. (2024). Artificial intelligence in interactive marketing: A conceptual framework and research agenda. *Journal of Research in Interactive Marketing*, 18(1), 54–90. <https://doi.org/10.1108/JRIM-01-2023-0030>
- Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, 88(5), 879–903. <https://doi.org/10.1037/0021-9010.88.5.879>
- Podsakoff, P. M., Podsakoff, N. P., Williams, L. J., Huang, C., & Yang, J. (2024). Common method bias: It's bad, it's complex, it's widespread, and it's not easy to fix. *Annual Review of Organizational Psychology and Organizational Behavior*, 11, 17–61. <https://doi.org/10.1146/annurev-orgpsych-110721-040030>
- Schmutz, P., Heinz, S., Métrailler, Y., & Opwis, K. (2009). Cognitive load in eCommerce applications—Measurement and effects on user satisfaction. *Advances in Human-Computer Interaction*, 2009, Article 121494. <https://doi.org/10.1155/2009/121494>
- Skulmowski, A., & Rey, G. D. (2020). Subjective cognitive load surveys lead to divergent results for interactive learning media. *Human Behavior and Emerging Technologies*, 2(2), 149–157. <https://doi.org/10.1002/hbe2.184>
- Sun, H., Xie, P., & Sun, Y. (2025). The inverted U-shaped effect of personalization on consumer attitudes in AI-generated ads: Striking the right balance between utility and threat. *Journal of Advertising Research*, 65(2), 237–258. <https://doi.org/10.1080/00218499.2025.2462405>
- Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4
- Sweller, J., van Merriënboer, J. J. G., & Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. *Educational Psychology Review*, 31(2), 261–292. <https://doi.org/10.1007/s10648-019-09465-5>
- Tankelevitch, L., Kewenig, V., Simkute, A., Scott, A. E., Sarkar, A., Sellen, A., & Rintel, S. (2024). The metacognitive demands and opportunities of generative AI. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Article 680, pp. 1–24). Association for Computing Machinery. <https://doi.org/10.1145/3613904.3642902>
- Teepapal, T. (2025). AI-driven personalization: Unraveling consumer perceptions in social media engagement. *Computers in Human Behavior*, 165, Article 108549. <https://doi.org/10.1016/j.chb.2024.108549>
- Teepapal, T. (2026). Corrigendum to “AI-driven personalization: Unraveling consumer perceptions in social media engagement” [Computers in Human Behavior 165 (2025) 108549]. *Computers in Human Behavior*, 174, Article 108797. <https://doi.org/10.1016/j.chb.2025.108797>
- Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly*, 27(3), 425–478. <https://doi.org/10.2307/30036540>
- Vieira, V. A. (2013). Stimuli–organism–response framework: A meta-analytic review in the store environment. *Journal of Business Research*, 66(9), 1420–1426. <https://doi.org/10.1016/j.jbusres.2012.05.009>