Response to the RPIB consultation on the Design of the Future Retail Payments Infrastructure

Questions 11, 17, 17a and 18 — Agentic payments and advanced delegation

Respondent: Matthew T. Kirby, Independent Researcher, River Falls, Wisconsin, USA — matthew.kirby@mandatepatch.org. I consent to publication of this response with attribution.

Scope: Questions 17, 17a and 18 (with a short observation on Question 11), with reference to “Agentic payments and advanced delegation” and Box G.

Summary

The consultation identifies four matters requiring further consideration for agentic payments: how users understand and control what agents are permitted to do, who is accountable when transactions go wrong, how agent-initiated transactions are identified, and how agents are authenticated. This response addresses the first three. It does not address agent authentication, which is well served by existing authentication standards and is not where the evidentiary gap lies.

Box G’s £200 auction instruction is a single amount constraint. Safe agentic payments require a lifecycle for that authority when execution diverges from the constraint — and from constraints Box G does not name: what was bought, from whom, in what quantity, and whether the resulting transaction still matches what the user authorised.

The published frameworks already recognise points at which existing authority is insufficient, a mandate does not match the proposed checkout, or buyer review is required. What they do not share is a transaction-scoped lifecycle artefact for narrowly authorising a changed transaction while preserving both the standing mandate and an auditable lineage from original authority to exception.

I filed US provisional application 64/141,321 on 26 August 2026. The specification is published openly as Mandate Lifecycle Extensions for Agentic Payment Credentials, Technical Disclosure Commons, Defensive Publications Series no. 11517 (August 2026), under CC BY 4.0. A reference implementation of the artefact layer is published at github.com/MandatePatch/mandatepatch under Apache-2.0.

My principal recommendation to the Board is narrower than adoption of any particular protocol: the next-generation messaging architecture should be capable of carrying opaque, integrity-protected references that let authorised parties join an executed agentic payment to the authority and evaluation evidence that permitted it. The semantic evaluation itself need not sit in the clearing kernel.

The gap, in the published protocols’ own text

Across the published frameworks the same architectural boundary appears in different forms.

OpenAI and Stripe’s Agentic Commerce Protocol defines the message resolution value requires_buyer_review — “Buyer must authorize before order placement.” The published protocol therefore recognises a point at which execution should stop for buyer review. It does not itself define a signed buyer-authorisation artefact expressing the changed transaction the buyer approved.

The Universal Commerce Protocol, in its AP2 Mandates extension, cryptographically binds its checkout mandate to a checkout state and defines mandate_scope_mismatch where “the mandate is bound to a different checkout.” The extension defines the mismatch. It does not define a transaction-scoped amendment or supersession object that changes only the failed predicate while preserving lineage to the original mandate.

Google’s Agent Payments Protocol takes a different approach in Human-Not-Present mode. The user signs open mandates carrying constraints; once the Shopping Agent has assembled an appropriate checkout, the agent may sign the closed checkout with its Agent Key and provide both the user-signed open mandate and the agent-signed closed mandate to verifiers. AP2 therefore distinguishes original human authority from autonomous closure. What I have not found in the published protocol is a transaction-scoped delta-authorisation object for the case where the proposed checkout no longer fits the standing authority but the user wishes to approve only that exception rather than replace the standing mandate.

The common gap is not that these protocols are incapable of rejecting or escalating a transaction. They can. The gap is what the successful re-entry of human authority produces.

A law firm’s presentation to the W3C/GS1 Workshop on E-commerce for Humans and AI Agents (Zurich, 8–9 September 2026) uses an example with the same shape as Box G: a mandate to buy “a cabin suitcase under €200” that the checkout turns into a €180 item, €15 delivery and a €49-a-year loyalty plan whose terms the agent accepts. The presentation’s conclusion is that standards “do not decide liability” but must “preserve the facts law needs to determine it.” That is the division this response assumes: the infrastructure should preserve joinable evidence, and the substantive rule can be set elsewhere.

Why outcome success is an unsafe proxy

Two documented agent failures illustrate different points in the lifecycle, and they should not be conflated.

Perception and instruction fidelity. In a June 2025 test of OpenAI’s Operator on a grocery task, the agent failed to transcribe items accurately — reading the “(5)” after bananas as “15” — omitted an item and added several unchecked ones (Timothy B. Lee, Understanding AI, 26 June 2025). This was not a payment-stage failure. It shows that the transaction state an agent constructs can already differ from the user’s instruction before payment begins.

Payment and fulfilment sequencing. In a December 2024 test of Perplexity’s shopping agent, the user checked out for toothpaste and the bank statement showed a payment to the agent rather than to the retailer; three hours later the agent reported it could not complete the purchase because the item was out of stock (Maxwell Zeff, TechCrunch, 2 December 2024). This is not semantic mandate drift. It demonstrates that apparent payment movement can precede successful completion of the underlying commercial task.

A July 2026 controlled-invalidation study makes the measurement problem sharper. It is a security stress test rather than an estimate of real-world prevalence: in its primary matrix, agents reached the visible endpoint in 262 of 270 runs while only 55 of 270 were authorised completions. Its own conclusion distinguishes endpoint success as a utility metric from authorised commit as a security property (arXiv:2607.10487).

These are different failure mechanisms. Their common lesson is narrower than any of them individually: successful-looking execution is not sufficient evidence that an agent remained inside valid authority.

Why the rails cannot reconstruct the authority decision

The difficulty is not principally message capacity. Authority is created and evaluated on one plane while money moves on another, and the evidence connecting the two is not consistently projected across that boundary.

Cards — loss of semantic context before settlement. Rich cart, line-item and agent-authorisation information can exist at the checkout or agent-protocol layer. The point is not that card authorisation messages contain only an amount and a credential; they carry materially more transaction and risk data. The narrower observability point is that the public evidence I have examined does not expose a standard transaction-level join from an agent mandate and its conformance evaluation through to settlement. At final sterling settlement the abstraction is clearer still: card schemes settle net obligations into the Bank's RTGS infrastructure rather than presenting each underlying consumer checkout.

Open Banking — evaluation exists, but the authority evidence does not project. This is the clearest case, because the evaluation is explicit. Under the Read/Write Standard v4.0, where a payment would breach one or more of a variable recurring payment consent's ControlParameters, the ASPSP must return an error — code U014 in that version — identifying the offending control parameter in the Field element. That is a real, machine-readable evaluation of an attempted payment against a standing authority, and it remains at the API boundary.

The public Faster Payments mapping does identify this class of payment. Field 61.1 PAYMENT SUB-TYPE is "set by the FPS Institution with a V prefix for any FPS transaction initiated by a PISP for both sVRP and cVRP only," and the scheme separately requires that the PISP be identified through field 122 REGULATORY REPORTING. A conforming variable recurring payment is therefore not indistinguishable from a manually keyed one.

What the mapping does not project is the authority evidence. EndToEndIdentification maps into FPS field 62. By contrast, ConsentId — the identifier of the consent against which the ASPSP evaluated the payment — appears in none of the mapping tables on that page, whether to ISO 8583, Bacs Standard 18 or MT103. Nor is there any projection for the applicable ControlParameters, or for the result of the evaluation itself.

The distinction matters. Faster Payments can know what kind of payment this is without knowing which authority it satisfied. Where a control is breached, the API request is rejected and no FPS transaction exists on which to record a verdict. Where the payment conforms, an FPS transaction exists but carries no lineage explaining why it was permitted.

Faster Payments and Bacs — semantic authority is not a clearing function. Neither system is designed to interpret the semantic bounds of a delegated authority, and neither should be. Increasing message capacity alone therefore does not solve the problem: in the non-conforming case there may be no interbank payment at all, and in the conforming case there is a payment but no common evaluation attached to it. The prerequisite is not a larger field. It is that a structured evaluation be produced, retained, and stably referenced.

Interoperability between forms of money — Q11

On Question 11, I support the proposed approach at the clearing and messaging layer. My observation concerns three questions that sit above it and are not yet named. The consultation defines interoperability at that layer, where liabilities are exchanged frictionlessly and at par. Those three questions are: who performs the conversion, who bears its cost, and what claim, against whom, the payer stands in legal relation to at the

moment of authorisation. “At par” is a statement about the exchange rate, not about fees. Where the payer’s agent selects the form of money, these become authorisation questions, and the joinable-evidence capability described under Question 17 is what would let them be answered after the fact.

Capabilities the ecosystem and infrastructure should support — Q17

The consultation observes that agentic payments “raise important policy and design challenges, many of which sit beyond the core infrastructure.” The Board’s April 2026 meeting summary records that agentic payments “were unlikely to require fundamental changes to the core infrastructure per se, provided it remained flexible enough to accommodate future developments,” and that “consent, liability, identification and scheme rules would require further consideration.” I agree with both halves. The capability described below is what the flexibility proviso has to mean in practice, and the requirement it implies can sit in scheme rules and the services layer rather than in the clearing message set. Because the Board is now defining handover artefacts for DeliveryCo, this is the moment at which such a requirement can enter the design record cheaply. I would therefore distinguish sharply between capabilities belonging at product, scheme or service layers and the narrower capability the core clearing and messaging infrastructure should support.

Items 1 to 4 need not execute inside the clearing kernel. The core should simply avoid foreclosing them. Item 5 is the core-infrastructure design requirement.

1. Committed intent — product or service layer. The authority should contain a cryptographically bound representation of what the user authorised, not merely numeric ceilings, so execution can be evaluated against the permitted transaction rather than only against an amount cap. The infrastructure need not understand that representation.

2. Suspension rather than binary termination — product or service layer. Where execution no longer fits the governing authority, the workflow should be able to suspend and produce a durable divergence record, rather than being forced to choose between silent continuation and irreversible termination.

3. Transaction-scoped amendment — product or service layer. A user should be able to authorise a narrowly defined exception for one transaction without silently rewriting the standing authority. The amendment should identify the transaction and the changed predicates, be consumed on use, and leave the standing mandate available for future transactions on its original terms.

4. Out-of-band confirmation bound to execution state — product or service layer. Where the user is brought back into the loop, the approval should bind to the exact transaction state being resumed. A notification asking only “approve?” is materially weaker evidence than a signed approval tied to the transaction state, the divergence, and the amendment being authorised. Shopify’s presentation to the W3C/GS1 workshop reports the problem from production — “the agent updated their cart — nothing happened on screen” — and asks for “a user-in-the-loop primitive” and for “regulatory/business disclosure the agent must not drive past.” Interface synchronisation of that kind is complementary to binding the approval to the transaction state; it is not a substitute for it.

5. Joinable evidence references — the core design requirement. I would encourage the Board to ensure the future messaging architecture is sufficiently extensible to carry stable, transaction-bound references to authority evidence where agentic-payment arrangements require them. For an executed agentic transaction those references should be capable of identifying, directly or through a governed evidence service: the authority artefact under which execution occurred; the relevant conformance evaluation; and, where applicable, the divergence and the user-approved amendment that allowed execution to resume.

These should be opaque, integrity-protected identifiers or cryptographic commitments — not mandate payloads.

The card-scheme standards body has now proposed a layer of this shape. On 1 September 2026 EMVCo opened for comment a draft EMV Agentic Payments – Framework for Specifications which "introduces the concept of 'Intent Services'": "a shared, interoperable layer that enables payment participants to register, reference, retrieve and manage consumer-authorised intent before, during and after a transaction," described as "registering intent, maintaining lifecycle and state information, and enabling the authorised retrieval of intent-related data," and intended to "complement cryptographic assurance provided by existing industry solutions, such as Verifiable Intent, with a common coordination point." EMVCo says the framework will help identify potential future enhancements to other EMV technologies, including EMV 3-D Secure, EMV Payment Tokenisation, EMV Secure Remote Commerce and the EMV Digital Payment Credential. That is a registry of authority evidence with governed retrieval, proposed for the card rails. Two observations follow for the Board. First, a UK account-to-account architecture that cannot carry a reference to such a registry would leave one rail with joinable authority evidence and the other without. Second, EMVCo's published verbs are register, reference, retrieve and manage; whether the draft provides for amending a registered intent when the transaction changes is not stated in the public release, and it is the amendment case that generates the divergence and re-authorisation evidence described in items 2 to 4. Personal instructions and detailed semantic content need not be exposed to the clearing layer, and dereferencing the underlying evidence can remain governed outside the core.

The core infrastructure therefore need not decide whether a fifteen-banana cart was inside a user's authority. It should make it possible for an authorised party later to establish which authority and which evaluation governed the payment. That single property preserves future options for schemes, regulators, payment providers, dispute-resolution bodies and courts without requiring the clearing kernel to become an agent-policy engine.

Additional message capacity alone is not sufficient. A useful architecture must ensure an evaluation can be produced, retained and referenced — not merely reserve space for data no upstream participant is required to generate.

Barriers — Q17a

First, there is no common transaction-scoped amendment path. The published protocols expose buyer-review, mismatch, rejection and autonomous-closure states, but share no artefact for authorising a narrow transaction exception while retaining cryptographic lineage to the standing authority. That fragmentation matters, because a future UK ecosystem should not have to infer the meaning of human re-entry differently for every agent protocol.

This is not only a reading of specifications. The Merchant Advisory Group, a US merchant association, told the same workshop that the ecosystem shows "competing approaches to mandate handling, tokenization, evidence design, and agent interaction"; that a transaction "may be authenticated at Authorization and still be disputed because the agent exceeded its authority, acted on stale instructions, selected the wrong product, or completed a purchase outside the consumer's approved limits"; and that merchants need evidence of what the agent was authorised to do "generated at Authorization and preserved through the dispute window."

Second, there is no public cross-rail mandate-conformance measure. I have found no public transaction-level conformance measure with a defined cross-rail denominator. Four things would have to be defined before supervisors, schemes or participants could answer how often agentic transactions remain within authority. Evaluation coverage: what share of agent-initiated transactions carry sufficient authority and evaluation evidence for conformance to be determinable at all. Mandate conformance: among evaluable transactions, the share executing without an unapproved divergence from the governing authority. Divergence incidence: how

often execution encounters a material authority mismatch, and of what type — amount, payee, item or service meaning, quantity, cumulative budget, or freshness. Unauthorised continuation: the share of transaction value moving after an authority evaluation has failed, without a fresh user authorisation resolving the failure.

Coverage must be reported first. Otherwise a high conformance rate can be manufactured by excluding transactions for which no evidence exists. This is why joinability is not merely a forensic convenience — it is a prerequisite for measurement, and measurement precedes allocation.

Third, protocol vocabularies are already diverging. ACP describes buyer review; UCP exposes mandate-scope mismatch; AP2 distinguishes open human-signed mandates from closed agent-signed ones. These are not necessarily incompatible designs, but they use different objects and different lifecycle vocabulary for related authority boundaries — precisely the fragmentation the consultation warns "could diverge, increasing costs." The infrastructure should not choose a winner among them. It should preserve enough common evidence semantics that the payment layer does not become permanently coupled to one. The FIDO Alliance Payments Technical Working Group, which now hosts Google's AP2 and Mastercard's Verifiable Intent, lists among the standards still needed the "lifecycle and validity of delegated authority, including expiry, revocation, audience or destination binding, and replay resistance." That list is the credential lifecycle. It does not yet name amendment of a live authority, which is the lifecycle question this response is about.

Fourth, consumer-protection architecture differs materially by payment method. Section 75 of the Consumer Credit Act 1974 can make a creditor jointly and severally liable with a supplier for qualifying breach or misrepresentation in debtor-creditor-supplier arrangements, and section 75(4) preserves that operation "notwithstanding that the debtor, in entering into the transaction, exceeded the credit limit or otherwise contravened any term of the agreement." The authorised push payment reimbursement regime addresses a different category of harm: PSR policy expressly distinguishes reimbursable APP scams from civil disputes, such as payments to legitimate suppliers where goods are not received or are defective (PS25/5, paragraphs 3.3 and 3.4).

The point is not that account-to-account payments lack protection. It is that routes to redress differ materially across payment architectures. Agentic orchestration makes that increasingly consequential, because software may compare or select payment methods on the user's behalf. Where liability and redress vary by method, preserving evidence of what the user authorised and what the agent actually did becomes more important, not less.

Consumer protection and fraud capabilities — Q18

For the core infrastructure I would prioritise the joinable evidence-reference capability described above, because it is a consumer-protection capability before it is a liability rule.

The Payments Vision Delivery Committee (PVDC) Strategy asks the infrastructure to "enable appropriate consumer protection across payment methods" (Outcome 3(c)). The Committee's update on roles and responsibilities, of 2 July 2026, is more specific: "There is an opportunity for the core infrastructure to support appropriate protections, including dispute resolution and redress mechanisms. The infrastructure will need to enable compliance with protections which are set out in regulation, as well as in scheme rules. This includes considering interactions with emerging technologies and innovations, such as programmable payments and agentic AI payments, and any associated new rules or regulations." It adds that "the core infrastructure scheme should support an appropriate minimum level of consumer protection across the board," while enhanced protection may sit with product level arrangements, and that the core should seek to design out financial crime "including through supporting data sharing and reporting."

A joinable authority reference is exactly a capability that enables compliance with protections set elsewhere. It does not tell the Bank what the substantive rule should be. It does not move semantic evaluation into the clearing kernel. It does not require mandate contents to become payment data. It allows the ecosystem to reconstruct the evidence needed to apply whatever protection, fraud-control, liability or dispute-resolution rules legitimately govern the transaction — including rules that do not exist yet.

Without it, a dispute about an agent-initiated payment must reconstruct authority indirectly, from proprietary platform logs, interface records and inconsistent protocol objects. With it, an executed payment can point to the authority state and evaluation lineage that permitted the money to move.

The point was made from the certification side at the W3C/GS1 workshop in Zurich on 9 September 2026, which I attended. A presenter from the payments-certification laboratory Fime characterised the liability shift that card payments rely on today as absent once an agent is in the loop, and as unresolved — his slide read “Liability shift ??” — under the emerging stack of signed mandates, verifiable intent and agent identity. None of the trust questions on that slide asked what happens when the executed transaction departs from the mandate. A joinable authority reference is the capability that lets that question be answered from evidence rather than from a platform’s own account of events, and so is the capability on which any future liability rule can actually be applied.

Conclusion

My recommendation is deliberately narrow. The core need not interpret user intent, need not execute agent-policy logic, and need not carry natural-language mandates or detailed checkout semantics in payment messages. It needs only to preserve the ability to join an executed agentic payment to the authority and evaluation evidence that permitted it.

The mechanisms I have developed — semantic divergence gating against a committed baseline, transaction-scoped amendment, supersession lineage, state-bound asynchronous step-up and presentation attestation — are one implementation of a layer that can produce such evidence, and are documented openly at Technical Disclosure Commons no. 11517. The infrastructure requirement is more general than that implementation: when software acts under delegated human authority, the payment should not become evidentially orphaned from the authority that permitted it.

I would be glad to provide technical detail to the Board or its Design Authority.

Equality Act 2010

The consultation identifies delegated payments as a route to inclusion, giving the example of authorising “a carer or trusted person to assist with managing payments.” Delegation of that kind can be particularly important for disabled people, older people, and people managing fluctuating capacity. It is also where an inability to reconstruct what was authorised carries the greatest consequence, because the principal may be least able to review a confirmation at the moment it is presented, or to reconstruct events afterwards from memory.

I would note that the capability proposed in this response does not depend on whether the delegate is software or a person. A design that carries authority references only for agentic payments would leave human-delegate journeys — the ones the consultation invokes for inclusion — with the same evidentiary gap. I have not identified a differential impact arising from the capability itself. The risk I would flag is of scoping it too narrowly, so that the journeys most associated with protected characteristics are the ones it does not reach.

Selected references

- Bank of England, RPIB consultation on the Design of the Future Retail Payments Infrastructure, 25 June 2026 — "Agentic payments and advanced delegation," Box G, Questions 17, 17a and 18.
- Payments Vision Delivery Committee, Strategy for Future Retail Payments Infrastructure, 7 November 2025 (Outcome 3(c)); and Payments Vision Delivery Committee update: Roles and Responsibilities in the future Retail Payments Ecosystem, gov.uk, updated 2 July 2026 (section 5).
- Open Banking Limited, Read/Write Standard v4.0 — Domestic VRPs; Domestic Payment Message Formats.
- OpenAI / Stripe, Agentic Commerce Protocol, Agentic Checkout specification (17 April 2026).
- Universal Commerce Protocol, AP2 Mandates extension.
- Google, Agent Payments Protocol specification, Human-Not-Present mode.
- Santos-Grueiro, I., Temporary Authority, Permanent Effects: Commit-Time Authorization for LLM Agents, arXiv:2607.10487, 11 July 2026.
- Lee, T. B., "Computer-use agents seem like a dead end," Understanding AI, 26 June 2025.
- Zeff, M., "The race is on to make AI agents do your online shopping for you," TechCrunch, 2 December 2024.
- Payment Systems Regulator, PS25/5, APP scams reimbursement — consolidated policy statement, May 2025.
- Consumer Credit Act 1974, sections 75 and 75(4).
- Bank of England, The Retail Payments Infrastructure Board (RPIB) — meeting summaries for 22 April 2026 and 15 June 2026, https://www.bankofengland.co.uk/payments/the-national-payments-vision/retail-payments-infrastructure-board (page last updated 1 September 2026; accessed 8 September 2026).
- Merchant Advisory Group (S. Cole), position statement (26 June 2026) and slides, W3C/GS1 Workshop on E-commerce for Humans and AI Agents, Zurich, 8–9 September 2026 — https://github.com/w3c/ecommerce-for-humans-and-AI-Agents-workshop/issues/18 ; https://www.w3.org/2026/09/ecommerce-agents/slides/Steve_Cole.pdf
- FIDO Alliance Payments Technical Working Group, position statement (8 July 2026), same workshop — https://github.com/w3c/ecommerce-for-humans-and-AI-Agents-workshop/issues/22
- Polimeni, A. (Polimeni.Legal), position statement (10 July 2026), same workshop — https://github.com/w3c/ecommerce-for-humans-and-AI-Agents-workshop/issues/32 ; slides presented by Caluori, C., on behalf of Polimeni.Legal, 9 September 2026 — https://www.w3.org/2026/09/ecommerce-agents/slides/Claudia_Caluori.pdf
- Grigorik, I. and Weiss, Y. (Shopify), slides, same workshop, 8 September 2026 — https://www.w3.org/2026/09/ecommerce-agents/slides/Yoav_Weiss.pdf
- Di Manno, J.-L. (Fime), From KYA to Runtime Trust: Closing the Gap in Agentic Commerce, slides, same workshop, 9 September 2026, pp. 5–7 — https://www.w3.org/2026/09/ecommerce-agents/slides/Jean_Luc_Di_Manno_Fime.pdf (accessed 9 September 2026); position statement (13 July 2026) — https://github.com/w3c/ecommerce-for-humans-and-AI-Agents-workshop/issues/34
- EMVCo, EMVCo Requests Feedback on Framework for Secure, Interoperable and Scalable Card-Based Agentic Payments, 1 September 2026 — https://www.emvco.com/news/emvco-requests-feedback-on-framework-for-secure-interoperable-and-scalable-card-based-agentic-payments/ (accessed 9 September 2026); EMV Agentic Payments – Framework for Specifications v1.0 DRAFT, comment period to 30 September 2026 (document not consulted; account-gated).
- Kirby, M. T., Mandate Lifecycle Extensions for Agentic Payment Credentials, Technical Disclosure Commons, Defensive Publications Series no. 11517, CC BY 4.0.