Online Resource 2. Coding agreement and sensitivity checks

Breaches in the Flow: A Mismatch Model of Micro-Decisions in Everyday Activity

Atrina Oraee and Saadi Lahlou
Department of Psychological and Behavioural Science, London School of Economics and Political Science
Correspondence: a.oraee@lse.ac.uk

1. Scope and data

These post-hoc checks use the retained final codes in the master workbook. They examine agreement between AO and two additional coders, FH and NL, and whether the association between initial behavioural markers and reported decisions is present under alternative coding and participant summaries. They do not constitute a new coding exercise.

The main worksheet contains 3,212 timestamped breaches from 60 participants. Decision is recorded as 1 for 2,436 breaches and 0 for 776. Initial-marker columns are distinguished from following-code columns. The analyses below use the initial-marker columns; a later Give up code is not counted as an initial Give up marker.

Definitions used throughout

PHS-exclusive: at least one initial Pause, Hesitation or Scan label, with neither initial Retry nor initial Give up. **RG-exclusive:** at least one initial Retry or Give up label, with no initial Pause, Hesitation or Scan. Breaches with labels from both groups, or from neither group, are retained as separate categories. Individual-marker summaries can overlap because a breach can carry more than one label.

The seven stored initial flags are Pause, Hesitation, Retry, Give up, Scan, IDK and None. IDK and None are retained for the coding-agreement checks; they are not additional substantive behavioural markers. An all-zero seven-flag vector is distinguished from an explicitly selected None label.

Analytical participant IDs S001–S060 follow first appearance in the timestamped master rows; event IDs E0001–E3212 follow those rows. These identifiers differ from the participant numbers used in the thesis. Public records omit names, source clip names and demographic details.

All five substantive initial-marker fields and Decision are complete in the main sample. One IDK value, in E1393, is missing and remains blank. No values were imputed. The accompanying reproducibility files retain this missing value and the verified interview-attendance flags.

2. Matched coding records

AO and observer records were matched on participant, clip and exact elapsed time. Excel clock-time and elapsed-duration representations were converted to integer microseconds without rounding or a tolerance window. Each retained pair had seven complete binary flags, at least one selected flag for each coder, and a binary Decision value for each coder. Exclusions were applied sequentially in the order below.

Table S1. Sequential inclusion and exclusion

Stage	AO–FH	AO–NL
Nonempty source records	328	341
Excluded: no exact master match	6	15
Exact matches before code checks	322	326
Excluded: incomplete binary initial flags	24	23
Complete initial flags in both records	298	303
Excluded: either seven-flag vector all zero	7	3
Excluded: incomplete binary Decision	0	0
Eligible matched breaches	291	300
Participants in eligible sample	9	8

The FH worksheet also contains 206 entirely empty formatting rows, excluded before counting the 328 source records. They are not missing coding observations. The two eligible samples share 106 breaches and include 485 distinct breaches from 14 distinct participants. They are analysed separately. An unmatched key does not establish that a coder missed an event.

Table S2. Agreement on the full initial-label vector

Pair	Exact seven-flag agreement	Vector κ (95% CI)	Flags and Decision both agree
FH	246/291 (84.5%)	0.803 (0.711 to 0.895)	228/291 (78.4%)
NL	231/300 (77.0%)	0.706 (0.591 to 0.798)	215/300 (71.7%)

Vector κ treats each observed seven-flag combination as a category. Joint agreement requires the complete vector and Decision to agree on the same breach. All vector-κ bootstrap replicates were defined.

3. Label-level contingency counts

Tables S3a–b show the complete 2 × 2 counts for each retained initial flag and for Decision. A zero is a recorded binary value in an eligible pair. Counts are not restricted to breaches where a marker was present.

Table S3a. AO–FH binary counts (n = 291)

Code	Both 0	AO 0 / observer 1	AO 1 / observer 0	Both 1
Pause	232	18	4	37
Hesitation	208	12	5	66
Retry	223	3	6	59
Give up	278	3	1	9
Scan	179	5	11	96
IDK	286	4	0	1
None	283	0	5	3
Decision	100	7	20	164

Table S3b. AO–NL binary counts (n = 300)

Code	Both 0	AO 0 / observer 1	AO 1 / observer 0	Both 1
Pause	243	22	1	34
Hesitation	226	14	9	51
Retry	210	13	7	70
Give up	287	5	0	8
Scan	168	9	22	101
IDK	294	6	0	0
None	291	1	7	1
Decision	83	13	20	184

Each row sums to its sample size. Positive and negative agreement are reported separately on the next page because high agreement on rare labels can largely reflect both coders recording zero. IDK and None remain visible even where their positive agreement is low.

4. Agreement estimates and uncertainty

Raw agreement is (n00 + n11)/n; positive agreement is 2n11/(2n11 + n01 + n10); negative agreement is 2n00/(2n00 + n01 + n10). Cohen's κ uses the two coders' marginal proportions for expected agreement. No coder is treated as ground truth.

Table S4a. AO–FH agreement

Code	Raw	Positive	Negative	κ (95% CI)	Undefined κ
Pause	92.4%	77.1%	95.5%	0.727 (0.573 to 0.881)	0
Hesitation	94.2%	88.6%	96.1%	0.847 (0.714 to 0.942)	0
Retry	96.9%	92.9%	98.0%	0.909 (0.846 to 0.977)	0
Give up	98.6%	81.8%	99.3%	0.811 (0.576 to 0.946)	0
Scan	94.5%	92.3%	95.7%	0.880 (0.816 to 0.943)	0
IDK	98.6%	33.3%	99.3%	0.329 (0.000 to 0.856)	49
None	98.3%	54.5%	99.1%	0.539 (0.000 to 1.000)	6
Decision	90.7%	92.4%	88.1%	0.805 (0.626 to 0.977)	0

Table S4b. AO–NL agreement

Code	Raw	Positive	Negative	κ (95% CI)	Undefined κ
Pause	92.3%	74.7%	95.5%	0.705 (0.542 to 0.778)	0
Hesitation	92.3%	81.6%	95.2%	0.768 (0.650 to 0.878)	0
Retry	93.3%	87.5%	95.5%	0.830 (0.615 to 0.930)	0
Give up	98.3%	76.2%	99.1%	0.754 (0.357 to 0.906)	4
Scan	89.7%	86.7%	91.6%	0.783 (0.668 to 0.864)	0
IDK	98.0%	0.0%	99.0%	0.000 (0.000 to 0.000)	34
None	97.3%	20.0%	98.6%	0.191 (-0.012 to 0.420)	46
Decision	89.0%	91.8%	83.4%	0.752 (0.660 to 0.814)	0

Intervals are pointwise 95% percentile intervals from 10,000 participant-cluster bootstrap samples (seed 20261116). Participants are sampled with replacement and all their retained breaches are kept. Undefined replicates are omitted from the relevant quantiles and counted above. Positive agreement has the same undefined counts as κ; raw and negative agreement have none. Detailed intervals for all agreement measures are in the CSV files.

These intervals are exploratory: FH and NL contribute only nine and eight participant clusters. The 0.000–0.000 interval for NL's IDK κ follows from the observed coding margins; it is not evidence of certainty about population agreement.

5. Decision patterns under each coder's codes

For each matched sample, AO and the observer are compared on identical breaches. Each coder's own initial flags define the groups, and that coder's own Decision values define the rates. Group membership can therefore differ between coders. The contrast is the PHS-exclusive rate minus the RG-exclusive rate.

Table S5. Same-event contrasts; decisions / breaches (%)

Sample	Codes	PHS-exclusive	RG-exclusive	Difference, points (95% CI)	Mixed; neither
FH	AO	164/212 (77.4%)	15/69 (21.7%)	55.6 (46.1 to 67.3)	1/1 (100.0%); 4/9 (44.4%)
FH	FH	148/213 (69.5%)	19/69 (27.5%)	41.9 (26.1 to 58.3)	1/1 (100.0%); 3/8 (37.5%)
NL	AO	183/212 (86.3%)	16/80 (20.0%)	66.3 (52.6 to 77.3)	0/0 (undefined); 5/8 (62.5%)
NL	NL	170/201 (84.6%)	23/90 (25.6%)	59.0 (52.5 to 64.8)	1/1 (100.0%); 3/8 (37.5%)

Table S6. Individual initial markers; decisions / breaches (%)

Marker	AO in FH sample	FH	AO in NL sample	NL
Pause	38/41 (92.7%)	41/55 (74.5%)	34/35 (97.1%)	45/56 (80.4%)
Hesitation	57/71 (80.3%)	58/78 (74.4%)	54/60 (90.0%)	59/65 (90.8%)
Retry	14/65 (21.5%)	15/62 (24.2%)	14/77 (18.2%)	19/83 (22.9%)
Give up	6/10 (60.0%)	8/12 (66.7%)	6/8 (75.0%)	9/13 (69.2%)
Scan	75/107 (70.1%)	67/101 (66.3%)	100/123 (81.3%)	90/110 (81.8%)

The grouped contrast remains positive under both observers' codes, although the rates differ from AO's. This does not establish replication of every marker. Initial Give up was associated with a reported decision in 8/12 FH-coded and 9/13 NL-coded breaches. AO's rates on those same samples were 6/10 and 6/8, compared with 18/58 for initial Give up in the full sample.

Paired bootstrap changes in group rates (observer minus AO), using the same participant resamples for each coder: FH, PHS: -7.9 points (95% CI -22.2 to -0.1); FH, RG: +5.8 points (95% CI -0.5 to 9.6); NL, PHS: -1.7 points (95% CI -8.1 to 5.9); NL, RG: +5.6 points (95% CI 0.9 to 9.0).

All four group-contrast intervals and paired changes had 10,000 defined replicates. Individual-marker intervals, contributing participant counts and undefined replicates are retained in the CSV files. These samples overlap and are not combined as independent replications.

6. Full sample and RRIW-only sensitivity

The main-sample analysis uses the final AO codes and the same exclusive groups defined on page 1. RIW refers to the replay interview and RRIW to the reinforced replay interview. All 60 participants completed RIW; 51 also completed RRIW. Attendance was joined through the verified participant crosswalk, rather than inferred from participant numbers.

Table S7. Pooled rates and participant-cluster intervals

Measure	All participants	RRIW completed
Participants / breaches	60 / 3,212	51 / 2,872
All breaches: decisions / total	2,436/3,212 (75.8%)	2,162/2,872 (75.3%)
PHS-exclusive: decisions / total	2,211/2,407 (91.9%)	1,951/2,135 (91.4%)
PHS contributing participants	60	51
PHS rate: 95% CI	89.7 to 93.8%	88.9 to 93.5%
RG-exclusive: decisions / total	44/263 (16.7%)	40/242 (16.5%)
RG contributing participants	49	43
RG rate: 95% CI	11.5 to 22.1%	11.1 to 22.1%
PHS − RG: percentage points	75.1	74.9
Difference: 95% CI, points	69.7 to 80.4	69.1 to 80.5
Other breaches: decisions / total	181/542 (33.4%)	171/495 (34.5%)

The nine RIW-only participants account for 340 breaches and 274 reported decisions (80.6%). Restricting the analysis to RRIW completers retains 2,872 breaches and 2,162 decisions. This restriction checks whether the descriptive association remains in that subset. It is not a randomized comparison of the two interview protocols.

For each sample, 10,000 bootstrap replicates resampled participants with replacement and retained all their breaches (seed 20261116). Intervals use the 2.5th and 97.5th percentiles. All group-rate and difference replicates were defined. Pooled rates weight participants according to the number of contributing breaches.

The full sample contains 541 breaches with none of the five substantive initial markers and one with markers from both groups; these 542 breaches form the “Other” row. They are not used in the exclusive-group contrast. Their presence does not by itself indicate missing coding.

7. Variation between participants

Equal-participant group means average each person's rate within a group. A person without that group is omitted from its mean, rather than assigned a rate of zero. Because the two means have different contributors, the paired comparison is restricted to participants with at least one breach in each exclusive group.

Table S8. Participant-weighted and omission checks

Measure	All participants	RRIW completed
Mean participant PHS rate (contributors)	91.6% (n=60)	90.9% (n=51)
Mean participant RG rate (contributors)	15.9% (n=49)	16.0% (n=43)
Participants with both groups	49	43
Paired mean difference, points (95% CI)	75.0 (67.2 to 82.1)	74.2 (65.6 to 81.7)
Paired median difference, points	82.9	81.8
Positive / zero / negative differences	47 / 0 / 2	41 / 0 / 2
Leave-one-participant-out range, points	74.0 to 76.3	73.6 to 76.2

For the paired interval, participants with both groups are resampled equally, using their individual PHS-minus-RG differences. All 10,000 replicates were defined. The two reverse cases occur in both analyses: S014 had 23/27 PHS decisions and 1/1 RG decision (−14.8 points); S049 had 25/26 and 1/1 (−3.8 points). Each contributes only one RG breach. All 49 full-sample paired results, including these cases, are retained in the participant table.

The leave-one-participant-out check recalculates the pooled contrast after removing each participant in turn. The difference remains positive in all 60 full-sample and all 51 RRIW-only omissions.

Table S9. Individual initial-marker rates; decisions / breaches (%)

Marker	All participants	RRIW completed
Pause	420/431 (97.4%)	383/393 (97.5%)
Hesitation	724/766 (94.5%)	626/665 (94.1%)
Retry	32/213 (15.0%)	30/197 (15.2%)
Give up	18/58 (31.0%)	16/53 (30.2%)
Scan	1,079/1,223 (88.2%)	954/1,090 (87.5%)

The five rows can overlap. These are initial-marker counts, not all occurrences of the corresponding behaviour later within a breach.

8. Worked example

Retry followed by giving up without a reported decision. The participant repeatedly tried to vacuum the gap between a bed and bedside table, then moved on to another area. The master row separates the initial Retry marker from the following Abandon and Give up codes.

Table S10. Linked record

Field	Recorded value
Participant / event	S001 / E0028
Recorded activity time	03:44
Visually matched footage segment	03:39–03:47
Initial marker	Retry
Following codes	Abandon; Give up
Decision code	0

Interviewer: “Did you decide to stop? I mean, did you consider moving the bedside table to try and reach it?”
Participant: “No, I didn’t consider. I don’t think it was a decision. I was tired and wanted it [hoovering the room] to finish to go to sleep.”

The exchange is reproduced from Oraee (2024, pp. 150–151; thesis participant P19). The original footage at 03:39–03:47 was visually matched to the storyboard in Figure 15 (p. 92). The interview audio was not newly transcribed or verified for this supplement. The question itself proposes moving the bedside table; the answer should be read in that interview context. S001 and thesis P19 belong to different identifier systems.

The example illustrates the recorded distinction between behaviour and reported decision. It does not establish the absence of all decision processes, and it does not supply a separate avoidance classification. Give up is a following code in this row, not an initial Give up label.

Interpretation and reproducibility

The agreement analyses concern retained final codes after replay interviews. They do not measure independent detection of all breaches, accuracy against ground truth, or the validity of retrospective experience reports. The coding checks use selected, overlapping samples with few participant clusters. Participant resampling addresses dependence within people; it does not remove common-source measurement or establish the proposed cognitive mechanism. These results therefore support a descriptive association and specify its limits.

Online Resource 3 contains the de-identified analysis records, Python scripts and complete result tables. It permits the matching exclusions and statistics to be reproduced without the private name-to-ID crosswalk. Source-workbook calculations and the portable public-data calculations were checked for agreement. The master workbook was read without alteration. These checks were post hoc and were not preregistered.

Source: Oraee, A. (2024). *Micro decision-making in everyday life: An investigation with Subjective Evidence-Based Ethnography* [Doctoral thesis, London School of Economics and Political Science]. Figure 15, p. 92, and pp. 150–151.