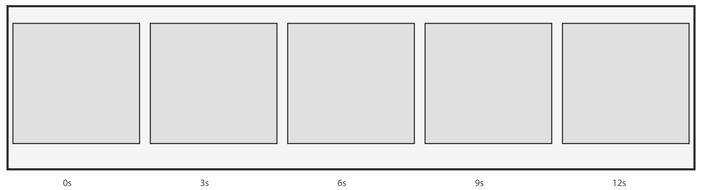
HERBench: High Evidence Requirement Benchmark

(1) Referring & Tracking

[AGAR] Appearance-Grounded Attribute Recognition

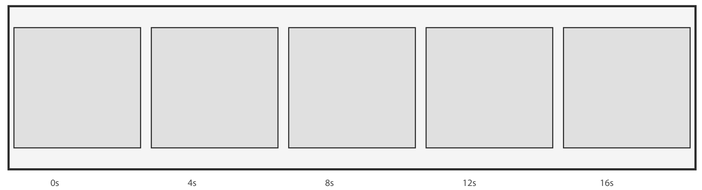
How does the person's pace appear throughout their movement in the video?



Answer: They maintain a consistent pace without stopping

[AGBI] Appearance-Grounded Behavior Interactions

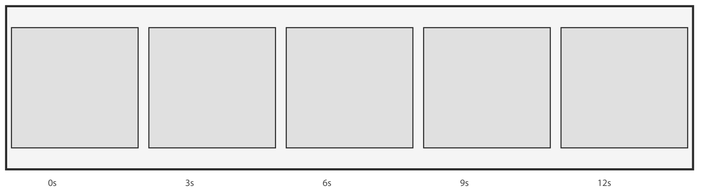
Who is accompanying the person as they move across the middle horizontal section?



Answer: A female wearing a light-colored coat and dark boots

[AGLT] Appearance-Grounded Localization Trajectory

In the video, how does the person move from the left to the right edge?

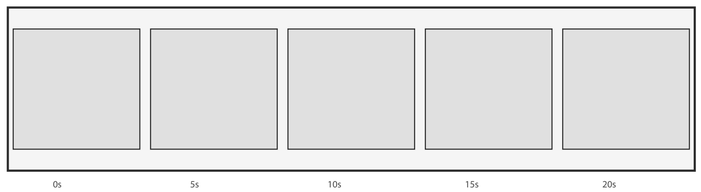


Answer: Mostly straight without noticeable stops

(2) Multi-Entity Aggregation & Numeracy

[MEGL] Multi-Entities Grounding & Localization

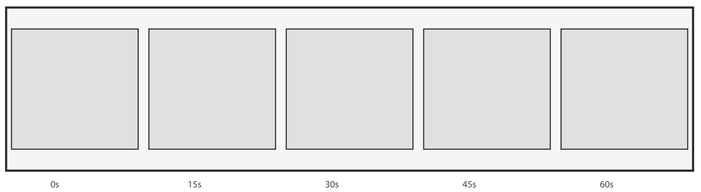
Which people appeared in the video? (descriptions must match exactly)



Answer: All of them

[RLPC] Region-Localized People Counting

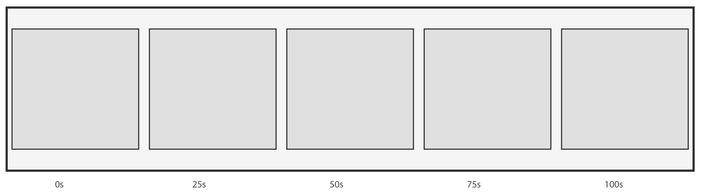
How many people entered the frame through the top edge?



Answer: 219-298

[AC] Action Counting

How many times does the action-object pair 'stir potatoes' occur?

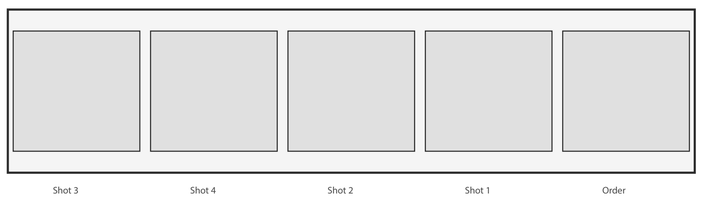


Answer: 31

(3) Temporal Reasoning & Chronology

[TSO] Temporal Shot Ordering

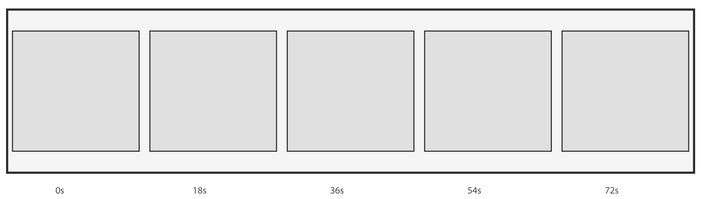
What is the correct chronological order of these 4 shots?



Answer: 3->4->2->1

[ASII] Action Sequence Integrity & Identification

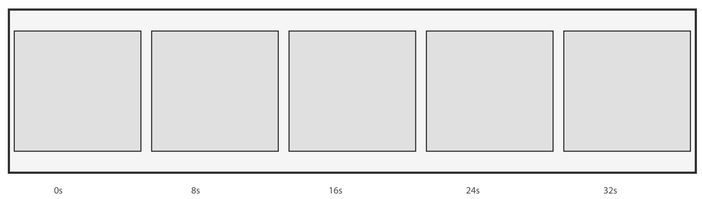
What is the correct temporal order of the 5 narrated events?



Answer: Close door -> Spoon rice -> Pick kettle -> Pick sriracha -> Pick soy sauce

[MPDR] Multi-Person Duration Reasoning

Who entered and exited the frame first?

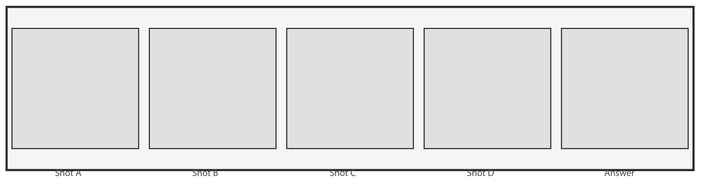


Answer: Person 2 entered first and exited first

(4) Global Consistency & Verification

[SVA] Scene Verification Arrangement

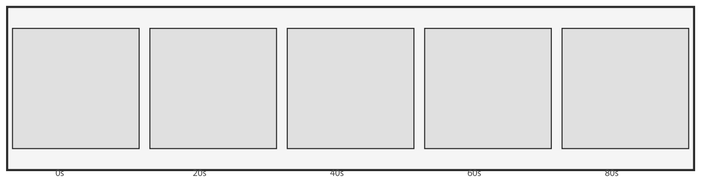
From the correctly described shots, which appears last in the video?



Answer: Person with dark hair at window overlooking stormy sea

[FAM] False Action Memory

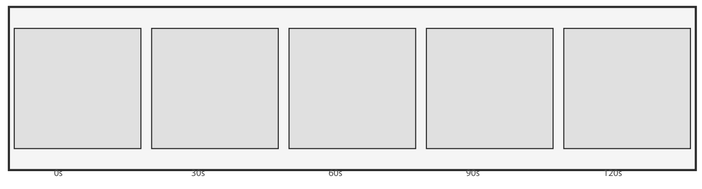
Which of the following actions did NOT occur in the video?



Answer: Stir chopped onion

[FOM] False Object Memory

Which object did the camera wearer NOT interact with?



Answer: Sweet potato