Find a clip described by the sentence: 'a man being filmed from the backseat of a car'. Provide a detailed description of the clip.

<tst> A man is being filmed from the backseat of a car, with the camera capturing the urban landscape and other vehicles in motion. <ent> <tst>

Legend: — <ent> — Left <tst> --- Right <tst> ▭ Pred Segment ▭ GT Segment

$p(f \mid s)$

Time (s)

(a) **Temporal Grounding**

Find all the clips belonging to the action category: 'showing something to the camera'. Provide a concise description for each clip.

Legend: — <ent> — Left <tst> --- Right <tst> ▭ Pred Segment ▭ GT Segment

<tst> A hand reaches for a bottle of water and show it to the camera. <ent>

$p(f \mid s)$

<tst> A hand presents a bottle of water, highlighting its significance in the frame. <ent>

$p(f \mid s)$

<tst> A hand holds a banana, drawing attention to its vibrant yellow color. <ent> <tst>

$p(f \mid s)$

Time (s)

(b) **Temporal Action Localization**