

Criteria	Rationale	Limit / criteria implementation	Exp. 1	Exp. 2
Number of responses collected (including returning participants)			120	320
Test restarted	A participant may take the test several times; in this case any run after the first one should be excluded.	Same IP and browser performing the test part of the survey several times.	-1	-1
Number of unique participants who completed the experiment			119	319
Test paused	If the participant paused the test for too long, he/she did not stay focused on the test and the influence of the learning context may have been compromised.	More than 60 seconds for any trial.	-1	-1
Attention lacking	If the participant keeps moving away from the test and failing to answer before the trial's end, he did not stay focused enough on the test.	Threshold on the sum of missing answers and moves away from the browser window: • Exp 1: maximum of 6 (over 48 trials). • Exp 2: maximum of 1 (over 5 trials).	-2	-5
Too low accuracy	When a participant answers wrongly too often, he/she did not understand the task or is not performing it seriously.	Threshold on accuracy (over all conditions): • Exp 1: minimum of 60%. • Exp 2: minimum of 40% (2 correct answers over the 5 trials).	-2	-4
Not enough valid trials	The experiment requires a minimum number of valid trials for a participant to get the intended experience, especially in terms of learning context (percentage of violations). Individual trials may be invalid because of technical issues or participant not abiding by the instructions, which happens more frequently with online experiments. See Table S3 for which trials are flagged invalid.	Thresholds on the number of valid trials: • Exp 1: in each block, valid rare trials ≥ 3 (optimal = 4) and ratio of frequent over rare trials between 2.5 and 4 (optimal ratio = $12/4 = 3$). • Exp 2: allow only 1 invalid trial, among the 4 familiarization trials. The violation trial must be valid.	-7	-35
Participant comments	Participants may report problems or deviant strategies in the debriefing page.	Decision on an individual basis, before looking at experimental results.	-0	-0
Total number of excluded participants			-12	-45
Number of participants included in accuracy and debriefing analysis			(10.1%) 107	(14.1%) 274
Extreme response times	Participants with extreme response times can weight abnormally on group means. Furthermore, very slow participants may be using an extremely cautious response strategy that prevents the studied effect from manifesting.	Median response time farther than 3 median absolute deviations from the group median (computed in one pass only, no recursion after outliers exclusion).	-1	-8
Missing RT data	Response times are computed on correct answers only, but some participants may have no correct answer on a condition, preventing median RT computation.	No correct answer in some experimental condition. (for Exp 3 only the single test trial was analyzed, so incorrect response on that trial means exclusion)	-0	-157
Number of participants included in RT analysis			103	109

Table S1. Exclusion criteria for participants and number of participants excluded, in Experiments 1 and 2. See next table for Experiment 3.

Criteria	Rationale	Limit / criteria implementation	Experiment 3
Number of responses collected (including returning participants)			1401
Test restarted	A participant may take the test several times; in this case any run after the first one should be excluded.	Exp 3: Same MTurk worker ID returning, or no worked ID recorded	-170
Number of unique participants who completed the experiment			1231
Test paused	If the participant paused the test for too long, he/she did not stay focused on the test and the influence of the learning context may have been compromised.	More than 60 seconds for any trial.	-2
Attention lacking	If the participant keeps moving away from the test and failing to answer before the trial's end, he/she did not stay focused enough on the test.	Moves away from browser window were not measured due to a software limitation. Missing answers are counted as invalid below.	-0
Too many invalid trials*	The experiment requires a minimum number of valid trials for a participant to get the intended experience. Individual trials may be invalid because of technical issues or participant not abiding by the instructions, which happens more frequently with online experiments. See Table S3 for which trials are flagged invalid.	Allow only 1 invalid trial, among the 4 familiarization trials.	-319
Too low accuracy	In Experiment 3 we used a more strict accuracy criterion in order to ensure that participants stayed focused and got sufficiently accustomed to the different outcomes.	Maximum of 1 error on familiarization trials (corresponds to minimum 75% correct, as there were four trials).	-59
Invalid violation trial*	Validity of the violation trial is assessed separately, in order to allow more participants to be included in the analysis of non-violation trials. See Table S3 for which trials are flagged invalid.	The single violation trial must be valid.	-56
Participant comments	Participants may report problems or deviant strategies in the debriefing page.	Decision on an individual basis, before looking at experimental results.	-0
Total number of excluded participants			-436 (35.4%)
Number of participants included in accuracy and debriefing analysis			795
Extreme response times	Participants with extreme response times can weight abnormally on group means. Furthermore, very slow participants may be using an extremely cautious response strategy that prevents the studied effect from manifesting.	Median response time farther than 3 median absolute deviations from the group median (computed in one pass only, no recursion after outliers exclusion).	-28
Missing RT data	Response times are computed on correct answers only, but some participants may have no correct answer on a condition, preventing median RT computation.	No correct answer in some experimental condition. (for Exp 3 only the single test trial was analyzed, so incorrect response on that trial means exclusion)	-458
Number of participants included in RT analysis			309

Table S2. Exclusion criteria for participants and number of participants excluded in Experiment 3. These criteria were identical to those of Experiment 2 except that certain browser events were not measured, and the threshold of accuracy was set higher. Exclusions for Experiment 3 is presented in a separate table because criteria were applied in a slightly different order, to allow separate processing of familiarization trials.

*These two criteria were summed together in Experiment 2. The number of excluded subjects is not affected by this difference in reporting.

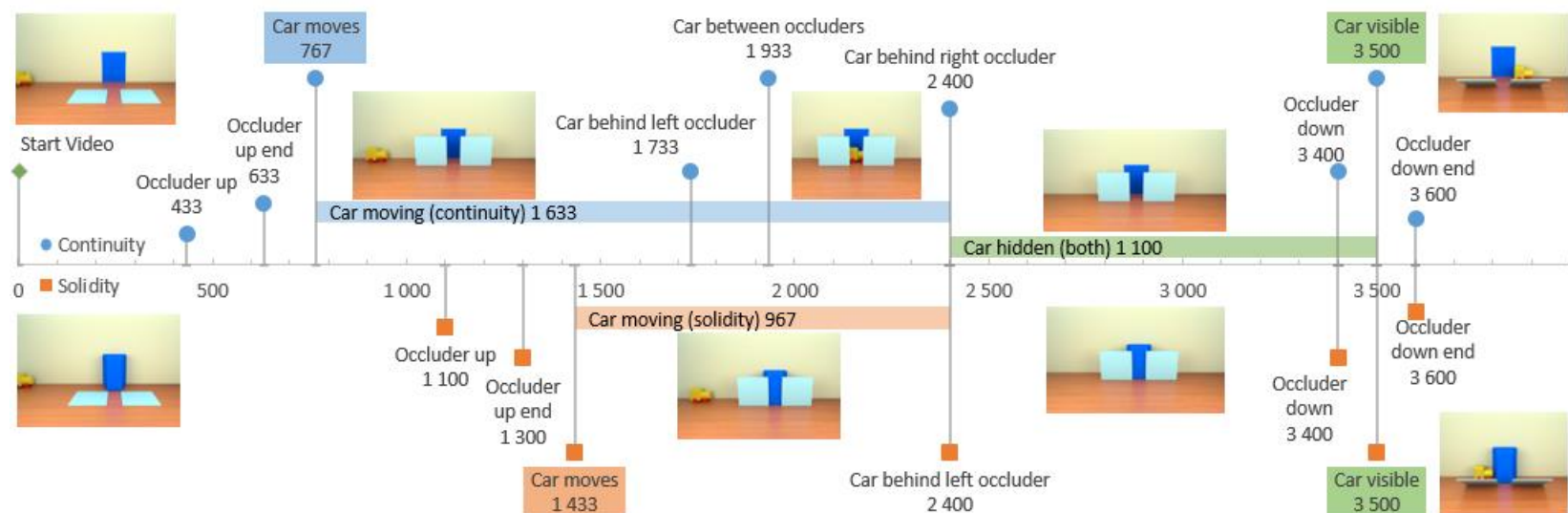
Criteria	Explanation	Limit / Implementation	Experiment 1	Experiment 2	Experiment 3	
					Familiarization	Test
Number of total trials among the kept participants			5136	1370	3640	851*
Missing answer	No answer in the 4 seconds given to answer means that something went wrong or participant was distracted.	No answer within 4s window.	-4	-6	-132	-50
Early answer	If the answer is given before the car reappears and not corrected afterwards, it can't be based on its actual position.	RT < 0 (not recorded for Experiment 3, however these trials are implicitly counted as missing above)	-8	-26	Counted as missing	
Many early answer attempts	A single early answer maybe a slip, but when a participant keeps trying to answer too early, he/she is not abiding by the instructions.	More than one answer attempt before the car is visible.	-9	-3	-51	-6
Participant left browser window	If a participant switches to another window on his computer while the video is playing, he is not attending it correctly	window.onblur or window.onfocus event after the car starts moving and before the participant answers.	-12	-1	Not recorded	
Video flow interrupted	The video may stop for buffering in the middle of a trial, preventing a correct display of the physical event to test.	Buffering event after car starts moving and before it reappears. Experiment 3 uses a different system, where buffering happens before the trials starts.	-34	-4	Not recorded / not applicable	
Total Invalid Trials included in analysis			-67 5069	-40 1330	-183 3457	-56 795

Table S2. Exclusion criteria for trials and number of trials excluded from analysis (among trials from non-excluded participants). For Experiment 3, familiarization and violation phases are counted separately, as more participants were included in the analysis of familiarization trials.

*This number refers to the number of trials/participants included before excluding based on test trial validity.

A

Continuity versus Solidity - Experiment 1 - Video timeline (ms)



B

Continuity versus Solidity - Experiment 2 - Video timeline (ms)

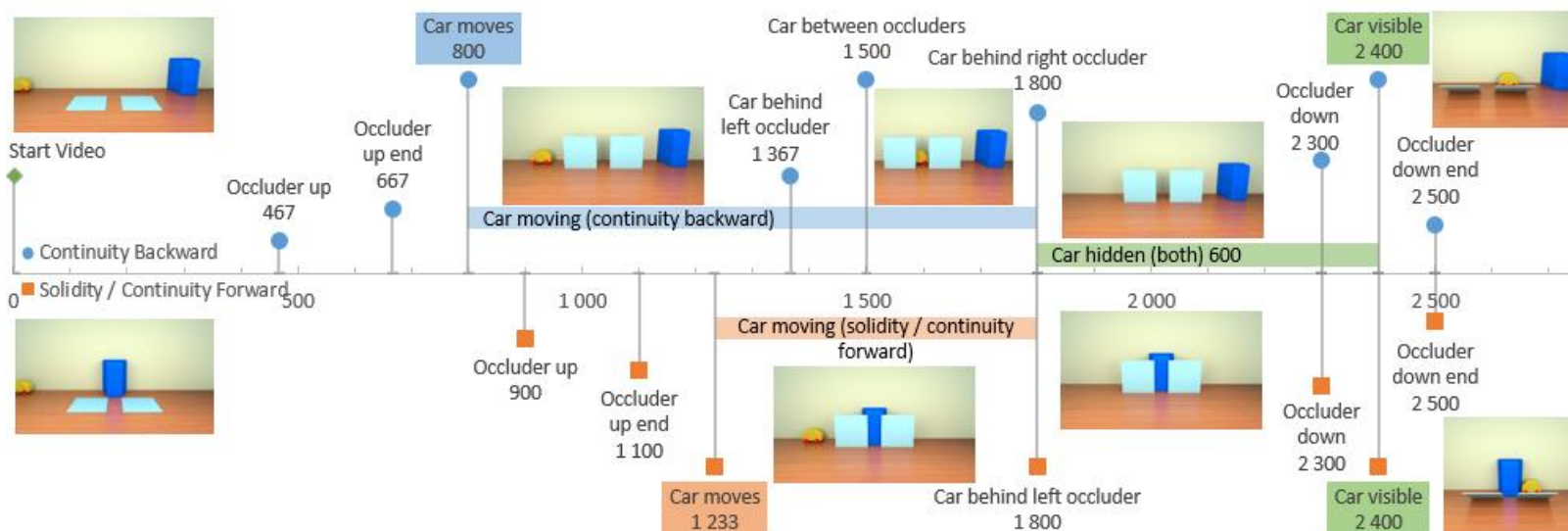


Figure S1. Timelines of events in stimuli videos from Experiment 1 (A) and experiment 2 (B), with example frames. Movement in the videos stops at 3600ms in Experiment 1 and 2500ms in Experiment 2. Actual videos continued for several seconds (repeat of the final frame), to give participants 4 seconds to answer after the car became visible.

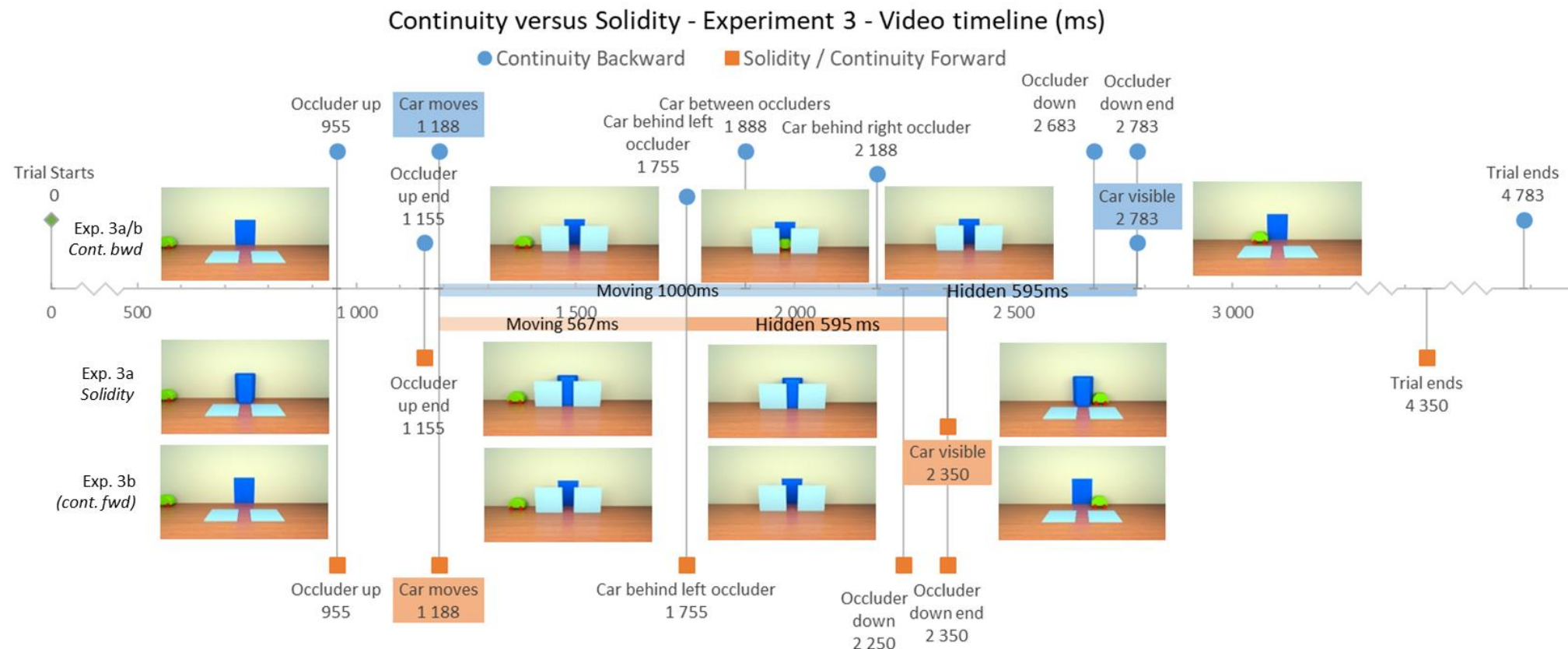


Figure S2. Timelines of events in stimuli videos from Experiment 3, with example frames. Above timeline (blue): Continuity backward events from both Experiments 3a and 3b. Below timeline (red): Solidity events from Experiment 3a (upper); Continuity forward events from Experiment 3b (lower). The final frame was always shown for two full seconds after the movement had stopped, leaving participants time to respond.