The Supplementary materials include the following sub-sections:

1. Experiment 4: A pre-registered replication of Experiment 1.
2. Analyses of response time data using lenient data screening criteria.
3. Error Rate analysis
4. RT data by effector and experimental condition
5. Qualifications that may be of interest to a more specialized audience.
6. An extended summary of computational modeling findings.
7. Instructions for accessing the study’s data and scripts.
8. Supplementary material references.

1. Experiment 4: A Pre-registered replication of Experiment 1

Experiment 4 is an exact replication of Experiment 1, pre-registered on OSF (https://osf.io/ysugp). Note that on the pre-registration form we committed to stringent data exclusion criteria which are more restrictive than justified compared with norms used on previous studies using this paradigm; e.g., filtering out responses slower than 850ms (see main text, pre-processing subsection under Experiment 1) and required to obtain good estimates of aggregated mean response time (Simmons et al., 2011). Additionally, we planned to sample 30 participants (to achieve an 85% power for a Cohen's d of 0.5). The analysis of the predicted *incompatibility effect on the incompatible condition* proved to be inconclusive at this point, although we did find conclusive support for the critical prediction of no difference between the irrelevant and Baseline condition (see sequential analyses below, Figure S1C). Therefore, we continued to sample participants, using a stopping rule of reaching a conclusive Bayes factor on the key contrasts of baseline condition vs. compatible, incompatible, and irrelevant distractor conditions (Dienes, 2008).

Fifty-nine participants were recruited, Seventy percent were female, ages 18-41 (M = 23.12, SD = 4.29). The final sample provided a statistical power of ~0.91 to find a small effect size of Cohen's d = 0.4. The demographics data for three subjects were not collected due to experimenter error.

The experiment was identical to Experiment 1 save for the following changes:

(a) Each trial began with a variable SOA of 600, 1200, or 1800MS (see below). During which the pronated hand stimuli were shown. The hand movement (for 170ms, as on Experiment 1) began only after the designated SOA has elapsed and the response-cue was displayed. The last frame of the movement and the symbolic cue remained on the screen for 800ms, and were removed after 970ms (170ms movement + 800ms cue and last frame of movement). Then, they were replaced with the blank blue frame (see Figure 1 on the main text) for the remainder of the trial, which always lasted 4500ms.

(b) Participants performed a single 24-trials-long practice block (each trial parameter combination was repeated once; order random, see below) and then continued to perform 4 experimental blocks, each 48-trials long. Blocks were interleaved by self-paced breaks.

(c) Each experimental block included 48 trials. A block was composed of trials representing 2 repetitions of the following 24 combinations: 2 Response Cues (‘1’ for the index finger, ‘2’ for the ring finger) X 4 Conditions (compatible, incompatible, irrelevant, baseline) X 3 ISI lengths (600, 1200 or 1800MS).

Results

Pre-Processing

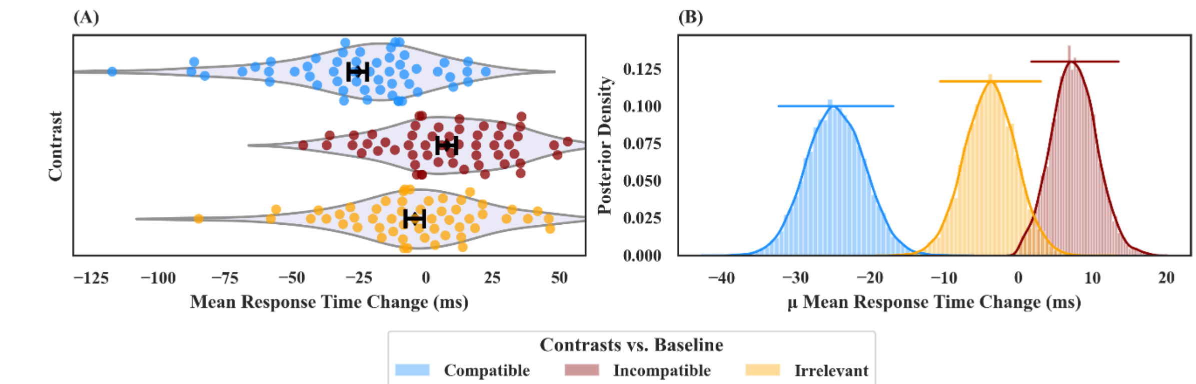
The 24 practice trials were discarded. We screened and analyzed the 192 trials from the experimental blocks (see specification above). Invalid trials included incorrect or omitted responses (6.82%), slow responses (>950ms, 2.77%) fast responses (<150ms, 0.43%), and trials in which participants did not press the keys down correctly at the beginning of the trial (none). Finally, all data from 4 participants (<60% valid trials, 6.78% of 59) were removed. Total filtration amounted to 13.08% of the data.

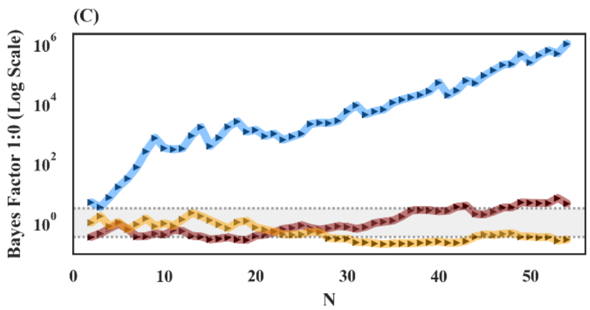
Analysis

For descriptive statistics and inferential plots see Figure S1.

A Repeated-Measures ANOVA shows that the effect of experimental condition on Mean Response Time is significant [$F(2, 104) = 43.11, p < 0.001$, Partial-$\eta^2 = .44$] . We continue by comparing the compatible, incompatible, and irrelevant conditions to the baseline condition.

The contrasts fully matched our predictions and fully replicate the results of Experiment 1 - RT was significantly lower for the compatible condition [Lower-tail $t(54) = -6.59$, $p < 0.001$, Mean Response Time Change -25.38ms, Cohen's $d = -0.90$ 95%-CI (-1.21, -0.58), BF1:0 = 1347273.110], significantly higher for the incompatible condition [Upper-tail $t(54) = 2.55$, $p = 0.007$, Mean Response Time Change 7.86ms, Cohen's $d = 0.35$ 95%-CI (0.07, 0.62), BF1:0 = 5.500] and not different for the irrelevant condition [Two-tail $t(54) = -1.15$, $p = 0.256$, Mean Response Time Change -4.02ms, Cohen's $d = -0.16$ 95%-CI (-0.42, 0.11), BF1:0 = 0.270]. Bayesian estimation of the differences in RT between conditions (shown in Figure S1B) supports the conclusions from the frequentist analysis (see Figure S1A). Finally, a sequential analysis of the Bayes Factor for each contrast confirmed the robustness of the pattern (Figure S1C).





(D)

Condition	Mean Response Time (MS)	Error Rate (%)
Baseline	558.96 (66.57)	3.45 (5.77)
Compatible	533.59 (51.55)	2.31 (3.8)
Incompatible	566.83 (64.89)	5.3 (7.1)
Irrelevant	554.94 (65.37)	3.67 (4.74)

Figure S1: Results of Experiment 4 – a compatible distractors accelerated response speed (blue; all plots), and incompatible ones slowed it down (red) while irrelevant distractor did not affect response speed (orange). (A) Estimated marginal means (X-Axis), 95%-CI and violin-plots of RT differences comparing each condition with the Baseline condition (Y-Axis), (B) Bayesian estimation of posterior distributions of RT differences between conditions (X-Axis) (condition, depicted by color); Stacked Horizontal lines indicate 95%-HDI of the posterior distribution. (C) Sequential analysis of Bayes factor (Y-Axis, Log-scaled) for contrasting each condition and the Baseline condition (Color); Shaded area indicates an inconclusiveness region (1/3 < $BF_{1:0}$ < 3). (D) Means (and SD's) of Response Speed and Error Rate on the different experimental conditions on Experiment 4.

2. Analyses of response time data using lenient data screening criteria

In this section, we provide an analysis that matches that of response speed performed for each experiment in the study but using a dataset that is filtered using more permissive inclusion criteria. Specifically, only incorrect responses, premature responses (i.e., responses emitted before the presentation of the cue and distractor), and subjects that were at the chance level or below in terms of accuracy were removed. Crucially, *none* of the results qualitatively changed compared to the results presented in the main text.

Experiment 1

Invalid trials included incorrect or omitted responses (8.39%) and trials with premature responses (<0ms). Finally, all data from one participant with below-chance accuracy (<50%, 1.79% of 56) on at least one of the conditions was removed. Total filtration amounted to 11.06%. Descriptive statistics are brought in Table S1.

Table S1: Means (SDs) of Response Speed and Error Rate on the different experimental conditions on Experiment 1, when minimal filters were used to preprocess the data.

CONDITION	Mean Response Time	Error Rate
BASELINE	485.88 (67.99)	7.21 (7.93)
COMPATIBLE	467.49 (60.61)	5.94 (7.19)
INCOMPATIBLE	509.57 (86.89)	10.24 (9.75)
IRRELEVANT	483.53 (80.91)	7.76 (7.75)

Repeated-Measures ANOVA showed a significant effect of experimental condition on Mean Response Time [$F(2, 89) = 32.75$, $p < 0.001$, Partial-η^2= 0.378]. Average response time were significantly lower for the compatible condition - Lower-tail $t(54) = -3.73$, $p < 0.001$, Mean Response Time Change -18.39ms, Cohen's d = -0.51 95%-CI [-0.79, -0.23], BF1:0 = 112.120; significantly higher response time on the incompatible condition Upper-tail $t(54) = 4.54$, $p <$

0.001, Mean Response Time Change 23.69ms, Cohen's d = 0.62 95%-CI [0.33, 0.90], BF1:0 = 1260.180 and not different (than No movement condition) for the irrelevant condition, Two-tailed t(54) = -0.47, p = 0.637, Mean Response Time Change -2.35ms, Cohen's d = -0.07 95%-CI [-0.33, 0.20], BF1:0 = 0.160.

Experiment 2

Trials with incorrect or omitted responses (9.86%) and with premature responses (<0ms) we not further analyzed. Finally, All data from 2 participants whose accuracy was below-chance level (<50%, 7.69% of 26) on at least one of the conditions were removed. Total filtration amounted to 12.54%.

Descriptive statistics are brought in Table S2.

Table S2: Means (SDs) of Response Speed and Error Rate on the different experimental conditions on Experiment 1, when minimal filters were used to preprocess the data.

CONDITION	**Mean Response Time**	**Error Rate**
BASELINE	527.5 (75.03)	4.5 (8.36)
COMPATIBLE	508.68 (72.79)	4.58 (8.4)
INCOMPATIBLE	538.3 (86.91)	6.08 (5.75)
IRRELEVANT	529.72 (83.54)	4.5 (5.65)

Repeated-Measures ANOVA supports a significant effect of experimental condition on Mean Response Time [F(2, 40) = 19.55, p < 0.001, Partial-η^2 = 0.460]. As predicted mean response time was lower for the compatible condition t(23) = -4.63, p < 0.001, Mean Response Time Change -18.82ms, Cohen's d = -0.97 95%-CI [-1.45, -0.47], BF1:0 = 472.570, higher for the incompatible condition Upper-tail t(23) = 2.34, p = 0.014, Mean Response Time Change 10.80ms, Cohen's d = 0.49 95%-CI [0.06, 0.91], BF1:0 = 4.000, but no different for the irrelevant condition Two-tail t(23) = 0.62, p = 0.543, Mean Response Time Change 2.22ms, Cohen's d = 0.13 95%-CI [-0.27, 0.53], BF1:0 = 0.260.

Experiment 3

Trials with incorrect or omitted responses (12.51%) and with premature responses (<0ms) were not further analyzed. Finally, all data from 3 participants (~4%) who performed below-chance level (<50%, on at least one of the conditions were removed. Total filtration amounted to 15.03%. Descriptive statistics are brought in Table S3.

Table S3: Means (SDs) of Response Speed and Error Rate on the different experimental conditions on Experiment 3, when minimal filters were used to preprocess the data.

CONDITION	Mean Response Time	Error Rate
BASELINE	519.49 (82.22)	7.49 (6.88)
COMPATIBLE	500.26 (82.2)	6.85 (7.0)
INCOMPATIBLE	532.63 (85.43	10.52 (9.37)
IRRELEVANT	533.56 (89.07)	8.64 (7.29)

Repeated-Measures ANOVA showed a significant effect of experimental condition on Mean Response Time [$F(2, 133) = 28.92$, $p < 0.001$, Partial-$\eta^2 = 0.302$]. As predicted mean response time was lower for the compatible condition $t(67) = -4.17$, $p < 0.001$, Mean Response Time Change -19.23ms, Cohen's d = -0.51 95%-CI [-0.76, -0.26], BF1:0 = 470.410 and higher for the incompatible condition $t(67) = 2.82$, $p = 0.003$, Mean Response Time Change 13.14ms, Cohen's d = 0.34 95%-CI [0.10, 0.59], BF1:0 = 9.900. However, in contrast with our predictions we found significantly slower RTs on the irrelevant condition Two-tail $t(67) = 2.91$, $p = 0.005$, Mean Response Time Change 14.07ms, Cohen's d = 0.35 95%-CI [0.11, 0.60], BF1:0 = 6.190.

Experiment 4

Trials with incorrect or omitted responses (6.82%) and with premature responses (<0ms) were not further analyzed. Finally, all data from 4 (6.78%) participants who performed below-chance level (<50%) on at least one of the conditions were removed. Total filtration amounted to 10.58%. Descriptive statistics are brought in Table S4.

Table S4: Means (SDs) of Response Speed and Error Rate on the different experimental conditions on Experiment 3, when minimal filters were used to preprocess the data.

CONDITION	**Mean Response Time**	**Error Rate**
BASELINE	577.82 (85.82)	3.45 (5.77)
COMPATIBLE	551.99 (65.51)	2.31 (3.8)
INCOMPATIBLE	587.31 (81.51)	5.3 (7.1)
IRRELEVANT	577.2 (77.92)	3.67 (4.74)

Repeated-Measures ANOVA showed a significant effect of experimental condition on Mean Response Time [$F(2, 96) = 17.46$, $p < 0.001$, Partial-$\eta^2 = 0.244$]. In accordance with our predictions we found significantly lower response time for the compatible condition - Lower-tail $t(54) = -4.13$, $p < 0.001$, Mean Response Time Change -25.83ms, Cohen's d = -0.56 95%-CI [-0.84, -0.28], BF1:0 = 361.330, significantly higher error rate for the incompatible condition $t(54) = 1.83$, $p = 0.036$, Mean Response Time Change 9.49ms, Cohen's d = 0.25 95%-CI [-0.02, 0.52], BF1:0 = 1.340 and no difference in error rate for the irrelevant condition Two-tail $t(54) = -0.12$, $p = 0.908$, Mean Response Time Change -0.62ms, Cohen's d = -0.02 95%-CI [-0.28, 0.25], BF1:0 = 0.150.

3. Error rate analysis

In the literature of automatic imitation, the error rate is considered to be a secondary measure while the reaction time is the primary measure (Heyes, 2011). We present below the analysis of the error rate performed over all experiments. Our general prediction would be that the error rate would follow the pattern of RT's (a) lower on the compatible condition, (b) higher on the incompatible condition, and (c) similar on the irrelevant condition, all in comparison to the baseline condition. Generally, we found a pattern that is in congruence with our predictions although with smaller effect sizes and Bayes factors compared with RT. Crucially, the error rate analysis supported our key prediction that irrelevant distractors cause *no* automatic imitation. The error rate on the irrelevant condition was always similar to the one found on the Baseline condition as shown by the Bayes factors reported below. For descriptive statistics of error rate please see main text, Tables 1 through 3, and supplementary materials Table S1.

Experiment 1

Repeated-Measures ANOVA showed a significant effect of experimental condition on Error Rate [$F(2, 82) = 10.33$, $p < 0.001$, Partial-$\eta^2 = 0.168$]. In accordance with our predictions we found significantly lower error rate for the compatible condition - Lower-tail $t(51) = -1.85$, $p = 0.035$, Error Rate Change -1.28%, Cohen's $d = -0.26$ 95%-CI [-0.54, 0.02], BF1:0 = 1.420, a significantly higher error rate for the incompatible condition Upper-tail $t(51) = 2.95$, $p = 0.002$, Error Rate Change 3.20%, Cohen's $d = 0.41$ 95%-CI [0.13, 0.69], BF1:0 = 13.930 and no difference in error rate for the irrelevant condition Two-tail $t(51) = 0.65$, $p = 0.516$, Error Rate Change 0.51%, Cohen's $d = 0.09$ 95%-CI [-0.18, 0.36], BF1:0 = 0.190.

Experiment 2

Repeated-Measures ANOVA did not find a significant effect of experimental condition on Error Rate [$F(1, 25) = 0.81$, $p = 0.388$, Partial-$\eta^2 = 0.034$]. In contrast with the expected pattern we did not find a significantly lower error rate for the compatible condition Lower-tail $t(23) = 0.16$, $p = 0.564$, Error Rate Change 0.08%, Cohen's $d = 0.03$ 95%-CI [-0.37, 0.43], BF1:0 = 0.190 or for the incompatible condition Upper-tail $t(23) = 0.86$, $p = 0.199$, Error Rate Change 1.58%, Cohen's $d = 0.18$ 95%-CI [-0.23, 0.58], BF1:0 = 0.470. However, as predicted we found similar error rates for the Baseline and irrelevant conditions Two-tail $t(23) = 0.00$, $p = 1.000$, Error Rate Change 0.00%, Cohen's $d = 0.00$ 95%-CI [-0.40, 0.40], BF1:0 = 0.210.

Experiment 3

Repeated-Measures ANOVA showed that experimental condition significantly affected Error Rate [$F(2, 102) = 9.52$, $p < 0.001$, Partial-$\eta^2 = 0.133$]. In contrast with the expected pattern we did not find a significantly lower error rate for the compatible condition - Lower-tail $t(62) = -0.73$, $p = 0.236$, Error Rate Change -0.50%, Cohen's $d = -0.09$ 95%-CI [-0.34, 0.16], BF1:0 = 0.270. However, as expected the error rate for the incompatible condition was significantly higher, Upper-tail $t(62) = 3.17$, $p = 0.001$, Error Rate Change 2.93%, Cohen's $d = 0.40$ 95%-CI [0.14, 0.66], BF1:0 = 24.610 and no difference in error rate for the irrelevant condition Two-tailed $t(62) = 1.17$, $p = 0.247$, Error Rate Change 0.89%, Cohen's $d = 0.15$ 95%-CI [-0.10, 0.40], BF1:0 = 0.260.

Experiment 4

Repeated-Measures ANOVA shows that experimental condition significantly affected Error Rate [$F(2, 87) = 8.20$, $p = 0.001$, Partial-$\eta^2 = 0.132$]. As expected we found a significantly lower error rate for the compatible condition - Lower-tail $t(54) = -2.61$, $p = 0.006$, Error Rate Change -

1.14%, Cohen's d = -0.35 95%-CI [-0.63, -0.08], BF1:0 = 6.300, a significantly higher error rate on the incompatible condition Upper-tail $t(54) = 2.06$, $p = 0.022$, Error Rate Change 1.86%, Cohen's d = 0.28 95%-CI [0.01, 0.55], BF1:0 = 2.010 and no difference in the error rate of the irrelevant condition Two-tail $t(54) = 0.40$, $p = 0.694$, Error Rate Change 0.23%, Cohen's d = 0.05 95%-CI [-0.21, 0.32], BF1:0 = 0.160.

4. RT data by effector and experimental condition

As mentioned in the General Discussion in the main text, one of the reviewers of this paper found that the RT is also influenced by the effector (finger) that is used to respond to the symbolic cue. Below we discuss this issue and its possible implications to the conclusions of our study. These findings are not clear cut, but they do pose several interesting implications for our claim and for the field in general, all of which would require future verification.

Due to the design of Experiments 1 and 3, we were not able to perform inferential tests on the data if it is aggregated within-participant, condition, and response finger. On both experiments participants performed a total of 120 trials, assuming perfect performance across all participants (which is not the case), each data point would be the aggregated mean of 15 observations which is a small number of observations (Simmons et al., 2011) for RT. Therefore, we merely counted the number of times a specific pattern was found in one of the contrasts analyzed for each experiment on the study (we treated a difference of at least 10ms as being meaningful, similar to the raw effect sizes found on experiments 2 through 4). Descriptive statistics, along with plots of individual and group means as well as bootstrapped 95%-CI can be found in Figures S2-S5 below.

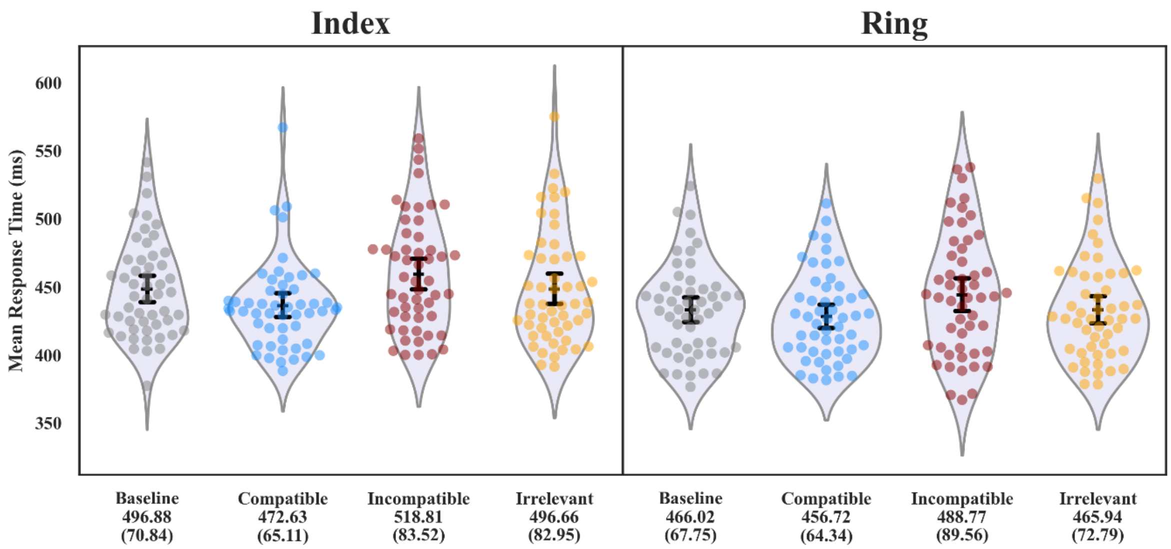


Figure S2:Individual and group mean (with bootstrapped 95%-CI) of responding to the symbolic cue with the two fingers used in Experiment 1.

First, we found that compatible distractors led to faster RTs on all experiments and for all fingers, compared with Baseline (8/8 of the cases; 4 Experiments X 2 effectors). The compatibility effect was about 20ms regardless of response finger. It seems that the compatibility effect was strongest for the index finger (>24ms, see figures S2-S3 and S5), compared to the middle or ring fingers. The most notable difference was in Experiment 2 where the compatibility effect was 25ms and 11ms for the index and middle fingers, respectively. An increased compatibility effect for the index finger is consistent with the conclusion that it is more easily activated (i.e. by external stimuli) due to its chronic relevance. Arguably, its chronic relevance makes the index a potent distractor, *in addition* to the general compatibility effect of observing a compatible distractor.

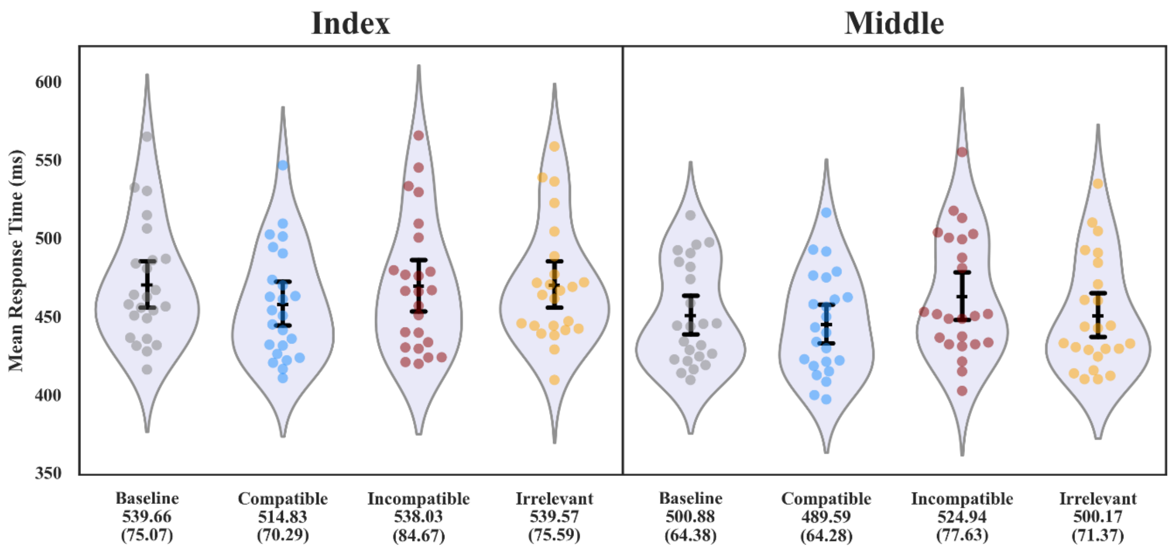


Figure S3: Individual and group means (with bootstrapped 95%-CI) of responding to the symbolic cue with the two fingers used in Experiment 2.

Second, irrelevant distractors led to RTs similar to those found for the Baseline condition, on 7/8 of the cases. The single exception to this pattern is is the irrelevant distractor slowing down responding on Experiment 3 when responding with the ring finger (but not for the middle finger, see Figure S4). First, the consistent lack of incompatibility effect for irrelevant distractors is rather invariant and solidifies the finding that irrelevant distractors are not automatically imitated. Second, the distractor, in this case, was the index finger which compliments the notion of its ‘chronicity’ (i.e. having a lower threshold for activation Higgins, 1996) which limits the efficiency of relevance-based modulation.

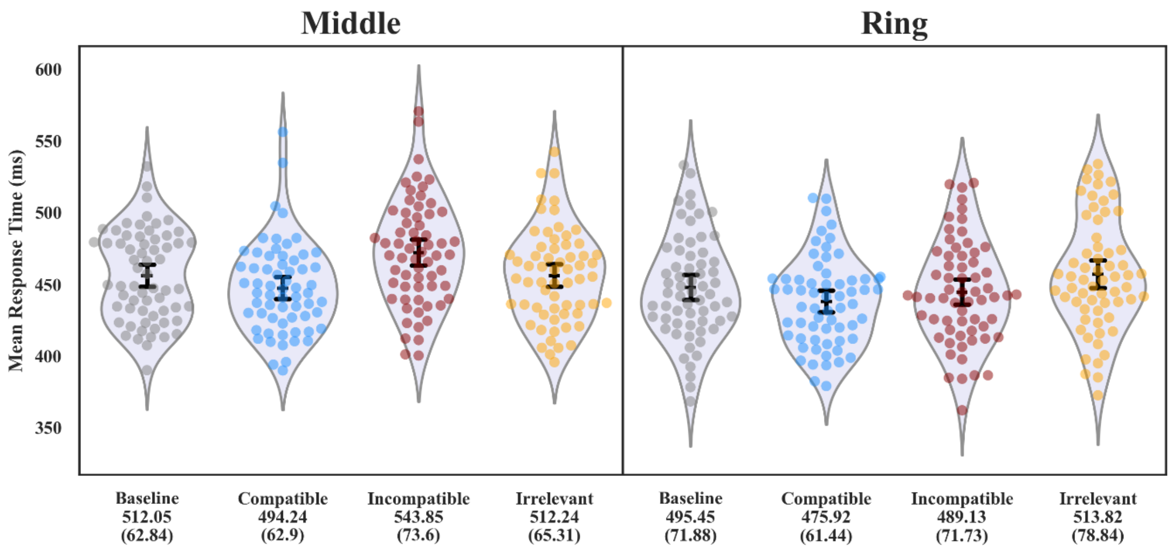


Figure S4: Individual and group means (with bootstrapped 95%-CI) of responding to the symbolic cue with the two fingers used in Experiment 3.

Third, incompatible distractors led to slower RTs compared with Baseline on 5/8 of the cases. When we did find the compatibility effect, it was about 20ms regardless of the finger used. However, in Experiments 2 and 4, when responding with the index finger no incompatibility effect was induced by an incompatible distractor (cf. Experiment 1, see figures S2-S3 and S5). In Experiment 3 the incompatibility effect induced by an incompatible distractor was slightly reversed for responses made with the ring finger. (i.e., ~6ms faster compared with Baseline, see figure S4). It seems then that the incompatibility effect induced by incompatible distractors is more variable compared with those mentioned above. Again, when responding with the index finger, more often than not we did not find an incompatibility effect. If the representation of the index finger is accessible due to its chronicity, it is also easily induced by an 'incompatible' stimulus such as the symbolic cue and hence is more robust to interference from competing response representations, consistent with ROAR.

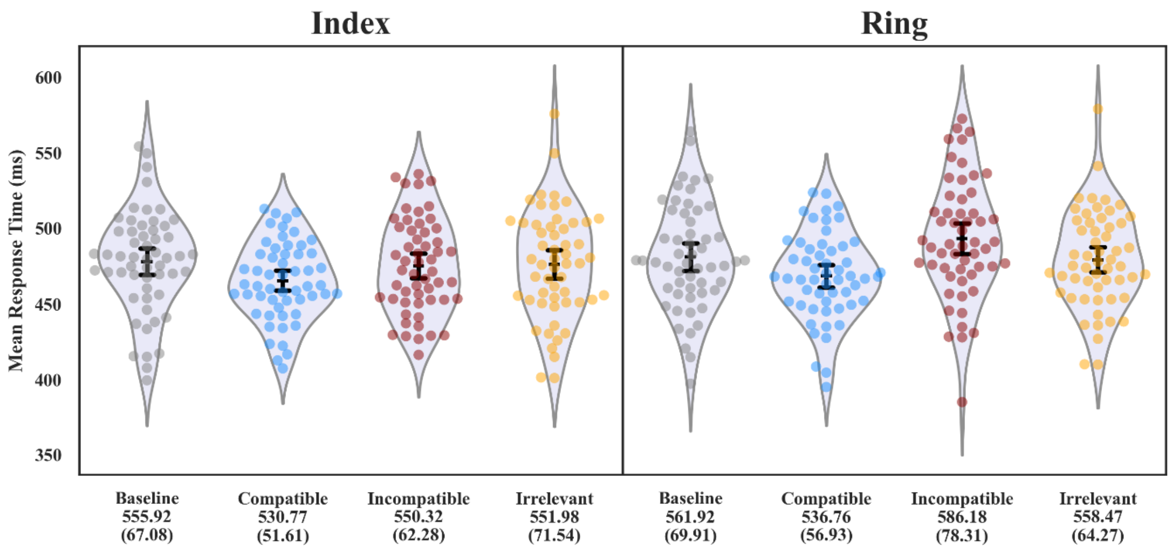


Figure S5: Individual and group means (with bootstrapped 95%-CI) on Experiment 4.

Fourth, as a whole, we find differences in response speed between the fingers used to respond on the task which can amount to ~30ms, although this pattern is not clear cut. Still, it is interesting that in some cases responses made with the index finger were *slower* compared to those made with the ring finger (Experiment 1, but not Experiment 4) or middle Finger (Experiment 2). Conversely, higher accessibility for the index would lead us to predict responses using it would be faster due to its ease of activation. Specifically for the current task, there is evidence of faster RTs using the middle compared with the index finger (Ainley et al., 2014). It is possible though that average response time stems from a different factor such as biomechanical differences between the fingers (as they unfold for keypresses).

It is worthwhile to analyze the data from imitation-inhibition studies at a single finger level. While is not currently a common practice (possibly also due to the limited number of trials collected for each combination), the exploratory analyses above suggest it may lead to novel insights about how mental and physical aspects of responding interact.

In a broader sense, it should be noted that there is a variant of the imitation-inhibition task that is not subject to the effector's identity threat. In that version, participants perform all movements using the same effector (usually the hand) in response to symbolic cues. The responses are usually complex gestures that require multiple fingers or wrist positions (e.g., Heyes, Bird, Johnson, & Haggard, 2005), rather than a keypress. This version is of course free from all force enslavement threats described above, as they relate to the independence of specific finger movements. Ideally, this version could be used in future studies to test whether the effect of automatic imitation is sensitive to the distractors' movement (and crucially, our novel finding of no imitation of irrelevant distractors). Additionally, this version allows more precise scrutiny of differences between effectors (e.g., responding with the left or right fingers) compared with movements (e.g., thumbs up vs. OK). This could help generalize our findings which were obtained using a set of distractors that includes three effectors, performing the same key pressing motion.

5. Additional Qualifications

The following section includes qualifications for the study which may be more specific to the imitation-inhibition paradigm than to the general conclusion of our findings.

Automatic Imitation of the irrelevant finger on Experiment 3

Recall from the main text that we found that irrelevant distractors on Experiment 3 led to an incompatibility effect, similar to incompatible distractors. Curiously, we did not find the same interference effect in terms of error rate (evidence brought in Table 3 in the Main text and available on the code notebook of Experiment 3, see details below). While in comparison with the Baseline condition, the incompatible condition induced a higher error rate (significant difference with strong support from the Bayesian analysis), on the irrelevant condition we found an error rate similar to that of the Baseline condition (with substantial support for the null as shown by the Bayesian analysis).

However, differing from response speed, the error rate shows a slightly less pronounced pattern compared with response speed and was not the main variable of interest. Others have argued (Heyes, 2011) that it is not clear whether it captures the phenomena of imitation as well as response speed (here, given the fact that we did not find a clear compatibility effect in terms of accuracy on the compatible condition). Whether this relative insensitivity of the accuracy measure to the manipulations maps onto ‘imitation’ or is due to the instructions of the task requires further research. It could be the case that in this task participants prioritize accuracy rather than speed in their responding, which leads to less variance on this measure and smaller effect sizes.

Spatial compatibility

One threat to the interpretation of tasks such as the imitation-inhibition is that performance differences do not result from activation of the observed motor plans per se, but due to spatial compatibility effects (i.e. between the observed hand movement and the performed response). To account for this threat, the presented distractors are sometimes horizontally or vertically inverted, to prevent subjects from relying on the location of the motion as a cue to imitate or avoid imitation (e.g., Brass, Bekkering, Wohlschläger, & Prinz, 2000; Bunlon, Marshall, Quandt, & Bouquet, 2015). However, the results are usually similar in terms of compatibility and incompatibility effects, regardless of the orientation of the stimuli and generally, this has not been considered a valid threat following maturation of the field of study (Heyes, 2011).

Notwithstanding, as we used no measures to control for spatial compatibility in the current study (such as using inverted distractors) we should address whether lack of spatial compatibility is a possible explanation for the lack of influence of the 'irrelevant' movements. Note that the lack of spatial compatibility effects may explain only the irrelevant condition in Experiment 2, where the middle finger was the distractor. This alternative explanation does not hold for the lack of incompatibility effect observed in Experiments 1 and 4, in which the ring finger was the irrelevant movement (as it is more spatially compatible with the location of the middle finger, compared to the index finger). Note further that spatial compatibility cannot explain why an incompatibility effect was always observed on the incompatible condition, but not on the irrelevant one (except for Experiment 3). As mentioned above, a future study testing our predictions could use a task that is less prone to spatial compatibility threats, similar to the task used by Heyes et al. (2005).

6. An extended summary of computational modeling findings

We recommend the reader to consult Figure 6 on the main text. Drift rate, the parameter of theoretical interest, when analyzed for each experiment separately. we found that the drift rate for the compatible and incompatible conditions was faster compared with Baseline or slower than Baseline, respectively. Crucially, the drift rate of the *irrelevant* condition was similar to that of the *Baseline* condition. The attentive reader might notice that the drift rate matched the pattern found for response speed, on all but one condition on a single experiment (the irrelevant condition in Experiment 3). Recall that in this experiment we found imitation of irrelevant movements (i.e., slower response speed), and yet it also provides evidence that irrelevant distractors do not affect evidence accumulation (i.e., decreased drift rate). This result is very odd, as it offers that the EZ model doesn't fully capture changes in response speed on the current task, at least in the edge case where the index finger is the irrelevant distractor. Possibly a more advanced model, requiring larger amounts of data would be able to test whether this divergence stems from another fitted parameter (e.g., response bias or variability in drift rate).

Regarding boundary separation, all conditions show lower boundary separation compared to the Baseline condition (albeit insignificantly). We can only speculate that the (un-natural) still hand image presented on the Baseline condition led participants to adopt more conservative criteria for deciding how to respond to the cue (although movement per se was not informative in terms of what would be the correct response). This pattern was less stable on the irrelevant condition than on others, with either lower or equal boundary separation compared with Baseline. It might be the case that an irrelevant movement being 'a movement' is not considered a non-event as the Baseline image, and hence affects the threshold of evidence accumulation, although task-irrelevant.

Regarding the third parameter of the model – Non-Decision Time, we saw no clear pattern and we were not sure what would be the implications of such a pattern if it existed as this parameter is related to non-psychological processes (e.g., neural conduction).

7. Data and Scripts

All data, code modules, and Jupyter notebooks related to the analysis of the data shown jrtr will be made available in a public GitHub repository of the study. Upon submission of the manuscript, they were made available to the editorial staff and reviewers.

The code notebook of each experiment includes:

1. Data wrangling.
2. Data pre-processing.
3. Descriptive statistics for observed (response speed, error rate) and latent variables (drift rate, boundary separation, non-decision time).
4. Diffusion model plot.
5. Bayesian and frequentist analysis of each observed and latent variable.

For additional details regarding dependencies etc., see the README file.

8. References

Ainley, V., Brass, M., & Tsakiris, M. (2014). Heartfelt imitation: high interoceptive awareness is linked to greater automatic imitation. *Neuropsychologia*, *60*, 21–28.

Brass, M., Bekkering, H., Wohlschläger, A., & Prinz, W. (2000). Compatibility between observed and executed finger movements: comparing symbolic, spatial, and imitative cues. *Brain and Cognition*, *44*(2), 124–143.

Bunlon, F., Marshall, P. J., Quandt, L. C., & Bouquet, C. A. (2015). Influence of action-effect associations acquired by ideomotor learning on imitation. *PloS One*, *10*(3).

Dienes, Z. (2008). *Understanding psychology as a science: An introduction to scientific and statistical inference*. Macmillan International Higher Education.

Heyes, C. (2011). Automatic Imitation. *Psychological Bulletin*. https://doi.org/10.1037/a0022288

Heyes, C., Bird, G., Johnson, H., & Haggard, P. (2005). Experience modulates automatic imitation. *Cognitive Brain Research*, *22*(2), 233–240.

Higgins, E. T. (1996). Activation: Accessibility, and salience. *Social Psychology: Handbook of Basic Principles*, 133–168.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*(11), 1359–1366.