



Reliable affordances: A generative modeling approach for test-retest reliability of the affordances task

Ran Littman¹ · Shachar Hochman^{1,2,3} · Eyal Kalanthroff¹

Accepted: 16 April 2023 / Published online: 1 May 2023
© The Psychonomic Society, Inc. 2023

Abstract

The affordances task serves as an important tool for the assessment of cognition and visuomotor functioning, and yet its test–retest reliability has not been established. In the affordances task, participants attend to a goal-directed task (e.g., classifying manipulable objects such as cups and pots) while suppressing their stimulus-driven, irrelevant reactions afforded by these objects (e.g., grasping their handles). This results in cognitive conflicts manifesting at the task level and the response level. In the current study, we assessed the reliability of the affordances task for the first time. While doing so, we referred to the “reliability paradox,” according to which behavioral tasks that produce highly replicable group-level effects often yield low test–retest reliability due to the inadequacy of traditional correlation methods in capturing individual differences between participants. Alongside the simple test–retest correlations, we employed a Bayesian generative model that was recently demonstrated to result in a more precise estimation of test–retest reliability. Two hundred and ninety-five participants completed an online version of the affordances task twice, with a one-week gap. Performance on the online version replicated results obtained under in-lab administrations of the task. While the simple correlation method resulted in weak test–retest measures of the different effects, the generative model yielded a good reliability assessment. The current results support the utility of the affordances task as a reliable behavioral tool for the assessment of group-level and individual differences in cognitive and visuomotor functioning. The results further support the employment of generative modeling in the study of individual differences.

Keywords Affordances · Test–retest reliability · Bayesian generative model · Individual differences · Stimulus-driven behavior · Task conflict

Introduction

Over four decades ago, James J. Gibson presented the seminal concept of affordances to describe the relationships that exist between organisms and their environments, indicating that “the affordances of the environment are what it offers the animal” (Gibson, 1979, p.127). According to this view, common manipulable objects, such as tools, handles, or kitchenware, automatically trigger responses that have acquired a strong association with them, resulting in automatic and specific motor plans for interacting with them (Makris et al., 2013; Proverbio et al., 2011; Tucker & Ellis, 2001). In the classic version of the affordances task, participants classify images of manipulable objects according to a certain rule (e.g., natural vs. manufactured; upright vs. inverted) by responding with their right or left hand (Tucker & Ellis, 1998, 2004; Wilf et al. 2013). Typically, the objects have a prominent handle and thus trigger an automatic grasping

Ran Littman and Shachar Hochman contributed equally to this work.

✉ Ran Littman
ran.littman@mail.huji.ac.il

✉ Shachar Hochman
s.hochman@surrey.ac.uk

¹ Department of Psychology, The Hebrew University of Jerusalem, Mt. Scopus, 91905 Jerusalem, Israel

² Department of Psychology, Ben-Gurion University of the Negev, Beer-Sheva, Israel

³ School of Psychology, University of Surrey, Guildford GU2 7XH, UK

response in one hand (e.g., a cup with the handle facing right or left will trigger a grasping response in the corresponding hand). Reactions decelerate and are more erroneous when the relevant response (classifying the object) and the irrelevant, stimulus-driven, grasping response activate different hands (*incongruent* condition) than when they activate the same hand (*congruent* condition). Recent studies have elaborated on this finding by adding a neutral condition to the task and demonstrated that two cognitive conflicts exist in the affordances task—a response conflict between responding with the relevant versus the irrelevant hand, and a task conflict between the goal-directed classification task and the stimulus-driven grasping task (Littman & Kalan-throff, 2021, 2022). While response conflict manifests only in incongruent trials, task conflict exists in both incongruent and congruent trials. Thus, typical results indicate a congruency effect (longer reaction time [RT] to incongruent than to congruent trials, indicating a response conflict), a reversed facilitation effect (congruent RT > neutral RT, indicating task conflict), and an interference effect (incongruent RT > neutral RT, which encompasses both task and response conflicts; Littman & Kalan-throff, 2022). Since its presentation in the seminal work by Tucker and Ellis (1998), the affordances task has been employed in a variety of studies and in various iterations to promote our understanding of human cognition, attention, and visuomotor functioning. However, despite its importance in experimental science, an evaluation of the task's psychometric properties, including its test–retest reliability, has not been undertaken. Critically, the lack of reliability measures poses a significant limitation to our ability to infer valid conclusions regarding aspects of individual differences measured by the task. Thus, our primary goal here was to establish the test–retest reliability of the affordances task.

For cognitive tasks, test–retest reliability is often assessed by correlating RT performance on different occasions of assessment (Enkavi et al., 2019). However, such efforts often result in low test–retest measures, falling short of the minimal satisfactory value of 0.7 (Barch et al., 2008), even with the most well-established tasks (Draheim et al., 2021; von Bastian et al., 2020). The “reliability paradox” (Hedge et al., 2018) refers to the phenomenon according to which behavioral tasks that produce highly replicable group-level effects often fail to capture individual differences between participants in the same task, for example, by yielding low test–retest reliability (Enkavi et al., 2019; Haines et al., 2020). For instance, in the case of the Stroop task, which yields a very robust and replicable interference effect (MacLeod, 1991; Stroop, 1935), simple test–retest correlations often yield satisfactory reliabilities for the congruent, incongruent, and neutral conditions individually, but significantly lower and unsatisfactory correlations for the

robust Stroop interference effect, calculated as the difference between the incongruent and the congruent conditions (Bender et al., 2016; Hedge et al., 2018; Strauss et al., 2005). The gap between the satisfactory reliability of the task's conditions and the low reliability of the task's effects, which otherwise produce robust replicable group-level effects, presumably stems from the nature of the well-studied “reliability of difference scores” (see review in Draheim et al., 2019). This term refers to the difference between the scores of two conditions that are highly correlated and often result in a value that is unstable across administrations, which in turn yields low reliability. Nevertheless, the focus on congruency effects and not on congruency conditions is crucial for the evaluation of the operation of “refined” processes of cognitive control that are evident beyond general performance RT. Thus, as many robust group-level effects result in unsatisfactory test–retest estimations, various researchers have raised concerns regarding their employment in the measurement of individual differences, or have deemed them unsuitable to do so (Dang et al., 2020; Elliott et al., 2019; Gawronski et al., 2017; Hedge et al., 2018; Schuch et al., 2022; Wennerhold et al., 2020).

A few recent studies have suggested that hierarchical Bayesian models could serve as key instruments to bypass the limitations of the common summary statistics practice to obtain test–retest reliability (Chen et al., 2021; Haines et al., 2020; Romeu et al., 2020; Rouder & Haaf, 2019). As mentioned above, summary statistics correlate mean RT differences across times of assessment. However, the summary statistics approach has two major limitations: (a) it ignores the specific distribution for each participant from which this mean derives, and (b) it neglects to account for trial-level variance, which constitutes an important source of data variability (Chen et al., 2021; Haines et al., 2020). As opposed to summary statistics, hierarchical Bayesian models are generative. That is, for a likelihood function predetermined by the researcher (e.g., lognormal distribution for RT, Bernoulli distribution for accuracy), hierarchical Bayesian models allow one to simulate data on the trial level based on the specific distribution calculated for each participant in the different experimental conditions (McElreath, 2020). Thus, hierarchical Bayesian models address the limitations mentioned above by (a) choosing a likelihood distribution for the model, which in turn provides each participant with their own specific distribution, and (b) incorporating the data in the model on the trial level, thus accounting for trial-level variance. That is to say, while summary statistics ignore the uncertainty (i.e., measurement error) associated with each participant's summary score, generative models specify a single model that jointly captures individual- and group-level uncertainty. Given that means alone are often imprecise when characterizing entire distributions, models that capture

the entire shape of participants' RT distributions may yield very different inferences. Therefore, because hierarchical Bayesian models are generative, they provide individualized distribution per participant (per experimental condition) and can provide an improved estimation of test–retest reliability. In this respect, hierarchical Bayesian models can provide a more reliable and more accurate estimation of task test–retest reliability than summary statistics.

Recently, researchers have compared the test–retest estimates of several well-established cognitive paradigms by using both mean RT correlations and hierarchical generative models (Chen et al., 2021; Haines et al., 2020; Snijder et al., 2022). The generative models consistently inferred higher test–retest measures relative to the summary statistics approach, and in many cases resulted in substantial differences in test–retest estimations. Moreover, Haines et al. (2020) demonstrated how the generative model estimates are highly consistent across replications of the same task, whereas estimates based on summary statistics may vary considerably. A core process that differentiates the two methods is the hierarchical pooling that takes place in the generative models' method and refers to the regression of individual-level parameters toward the group-level mean. Simply put, hierarchical pooling improves the estimation for each participant such that when there is great inconsistency in the participant's data (e.g., large variability), the estimation benefits from the group-level mean to yield a better estimate of the participant's performance. Importantly, Haines et al. (2020) showed that generative models do not automatically generate higher test–retest reliability than summary statistics, but that the process of hierarchical pooling only occurs to the extent it is warranted by the data.

In the present study, we followed the method employed by Haines et al. (2020) to assess the test–retest reliability of the affordances task for the first time, first by using traditional summary statistics and then by employing a Bayesian generative model. As mentioned above, the task has been shown to serve as a valuable behavioral tool for the assessment of visuomotor functioning, cognitive conflicts, and the activation of cognitive control at the task level and at the level of response (Buccino et al., 2009; Goslin et al., 2012; Grezes & Decety, 2002; Littman & Kalanthroff, 2021, 2022; Rice et al., 2007; Schulz et al., 2018). Furthermore, in recent years there is a growing interest in scientific psychopathological models that focus on the imbalance between stimulus-driven habitual behaviors and goal-directed behaviors (Gillan et al., 2014, 2015; Kalanthroff et al., 2017, 2018b; Robbins et al., 2012). Thus, establishing the affordances task's test–retest reliability is important for its utilization as a behavioral and neurocognitive tool for the assessment of control over stimulus-driven habitual behaviors. Finally, we administered the task online at both time points. Given the rising popularity

of online administration of cognitive tasks (e.g., Feenstra et al., 2018; Gillan & Daw, 2016), establishing reliability measures for an online version of the affordances task may be useful for both cognitive and clinical scientists.

Method

Participants

Three hundred and thirty-one students from the Hebrew University of Jerusalem, Ben Gurion University of the Negev, and Achva Academic College (all in Israel) took part in the experiment for course credit or small monetary compensation (~12 USD). Participants had normal or corrected-to-normal vision, were native speakers of Hebrew, and were naïve as to the purposes of the experiment. The experiment was approved by the Hebrew University institutional ethics committee (HUJI-500119). Informed consent was obtained from all participants prior to their participation in the experiment. The participants were instructed to register for the experiment only if they were able to complete its second part precisely one week after the first part (on the same day and hour in which they completed the first part). The results of 20 participants who did not complete the second part of the study and of an additional 10 participants who failed to complete the second part within six hours of the designated time were removed from the analyses. Following Hedge et al. (2018), the results of six participants were removed due to having more than 30% missed trials (three participants) or due to accuracy rates below 60% in either session (three participants). The analyzed sample thus consisted of 295 participants (226 female, 69 male) between the ages of 18 and 42 ($M = 24.5$, $SD = 3.1$). The proportion of left-handed participants was 10.8%¹.

Materials and methods

The experiment was programmed and administered online using Gorilla Experiment Builder (Anwyl-Irvine et al., 2020). Participants were instructed to complete the experiment on their private PCs, in a quiet environment, devoid of interruptions, and after turning off their mobile phones. The experiment was limited to participation via stationary or laptop computers; tablet devices or mobile phones were not permitted. The program matched the image resolution so that stimuli size was held constant across participants' monitors. The participants completed the affordances task

¹ A supplementary analysis indicated no significant effects for handedness.

twice, with a one-week gap between the two administrations. One day prior to the Time 2 designated administration, participants were reminded via email to complete the second part at the same time of day in which Time 1 was administered. The versions of the task in Time 1 and Time 2 were identical except for the practice block, which consisted of 90 practice trials in Time 1 and 30 practice trials in Time 2. We designed a longer practice block at Time 1 to familiarize participants with the keyboard keys. Icons of the relevant response keys appeared at the bottom of the screen throughout the practice block and disappeared during the experiment. By the end of the practice block in each session, a minimum 80% accuracy rate was required to start the experimental block. If a participant fell short of reaching this requirement, an additional 30 practice trials were added.

Prior to undertaking the affordances task, participants viewed four brief video clips, each lasting two seconds, in which a male/female hand (consistent with the participant's gender) reached and grasped a teapot or a cup by the handle, each turning left in one clip and right in a second clip (for similar procedures see: Garrido-Vásquez & Schubö, 2014; Littman & Kalanthroff, 2021, 2022; Tipper et al., 2006). This was done in line with previous suggestions according to which affordances tendencies may become more prominent under conditions that emphasize the object's graspability or the contextual correspondence of perception and action (Girardi et al., 2010; Lu & Cheng, 2013; Netelenbos & Gonzalez, 2015). Next, participants completed a practice block and then performed an experimental block consisting of 288 trials. Each trial began with a 500 ms fixation (white plus at the center of a black screen) followed by the target stimulus, which appeared for 1500 ms or until key-press, and an additional 500 ms of a black screen. Trials in which there was no response within 1500 ms were coded as missed trials and were not further analyzed. The target stimuli consisted of one of three black-and-white images (a cup, a teapot, or a house), obtained from the Amsterdam Library of Object Images (ALOI; Geusebroek et al., 2005). Stimuli were 767×574 pixels and appeared at the center of a black screen. Each stimulus appeared either in its upright form or in its inverted form on a random selection of 50% of the trials, a common procedure in affordances tasks (e.g., Iani et al., 2019; Saccone et al., 2016; Tucker & Ellis, 1998). Participants were instructed to indicate whether each object appeared in its upright or inverted form as quickly and as accurately as possible by pressing the "A" key with their left index finger or the "L" key with their right index finger. The mapping rule was counterbalanced across participants. To provoke an affordances effect, the cup and the teapot stimuli had a horizontal handle that could appear to the right or the left side of the object. In half of the trials, the handle direction was congruent with the correct response

key (i.e., both on the right or both on the left), while in the other half the direction of the handle and the correct response key were incongruent (i.e., left-facing handle and right correct response key, and vice versa). House images were previously shown to function as a neutral condition in the affordances task (Littman & Kalanthroff, 2022), serving as large objects that do not afford grasping tendencies (Chao & Martin, 2000). Within each presented orientation (upright vs. inverted), the trials were equally divided into neutral, congruent, and incongruent conditions with equal proportions and random order. As tools were previously shown to evoke affordances effects when presented in their functional orientation, but not in other, unfunctional orientations (Bub et al., 2018; Iani et al., 2019; Littman & Kalanthroff, 2022; Masson et al., 2011), we focused our analyses on the upright trials only (and indeed, the inverted trials data did not produce an affordances effect).

Statistical analysis

We began by trimming RTs shorter than 150 ms (0.06% of the data). To evaluate the within-task effects, a two-way analysis of variance (ANOVA) with repeated measures was applied to the RT data of correct responses with congruency conditions (congruent vs. neutral vs. incongruent) and time of assessment (Time 1, Time 2) as within-subject factors. Next, we assessed test–retest correlations of the RT data of correct responses. First, we employed the traditional summary statistics method and calculated Pearson's r correlations between the mean RTs of Times 1 and 2 for the congruency conditions (congruent, incongruent, and neutral) and the congruency effects (congruency, interference, and reversed facilitation). Following this, we assessed test–retest reliability for the congruency conditions and congruency effects by using the Bayesian generative model similar to the one presented by Haines et al. (2020). In this model, group-level normal distributions are considered prior distributions on the individual-level parameters. This allows information to be pooled across participants such that each individual-level estimate influences its corresponding group-level mean and standard deviation estimates, which in turn influence all other individual-level estimates. This interplay between the individual- and group-level parameters is the hierarchical pooling, a core feature of hierarchical models, which increases the precision of individual-level estimates and allows for the group- and individual-level model parameters to be estimated simultaneously (Gelman & Pardoe, 2006). Here, we analyzed our data by using a generative model with a lognormal link function, in which changes in stimulus difficulty produce changes in both means and variances of RT distributions (Rouder et al., 2015). Similar to RT distributions, lognormal is a positive-only right-skewed distribution.

Table 1 Mean (standard deviation for congruency conditions, standard error for congruency effects) reaction time (ms) and accuracy rates on the affordances task at Time 1 and Time 2

	Congruent	Neutral	Incongruent	Congruency	Interference	Reversed facilitation
Time 1	524 (83) 94.8% (5.3%)	516 (82) 93.8% (6.5%)	535 (84) 92.8% (6.9%)	10*** (2) $\eta_p^2 = 0.08$	19*** (2) $\eta_p^2 = 0.23$	-9*** (1.9) $\eta_p^2 = 0.07$
Time 2	507 (84) 95.5% (4.9%)	498 (83) 94.7% (5.4%)	512 (87) 94.3% (5.7%)	5*** (1.9) $\eta_p^2 = 0.03$	14*** (1.9) $\eta_p^2 = 0.15$	-9*** (1.8) $\eta_p^2 = 0.08$

Congruency = RT for incongruent minus RT for congruent condition. Interference = RT for incongruent minus RT for neutral condition. Reversed facilitation = RT for neutral minus RT for congruent condition

*** $p < .001$

For further discussion on analyzing RT data with a lognormal link function, see Lo and Andrews (2015). We used four chains, each using 3000 iterations and 1500 sample warm-ups. A figure of the Rhat distribution and caterpillar plots are presented in section S1 of the Supplementary Material.

We estimated parameters using Stan (version 2.2.1), a probabilistic programming language that uses a variant of Markov chain Monte Carlo to estimate posterior distributions for parameters within Bayesian models (Carpenter et al., 2017). For the generative model analysis, we report the highest maximum a posteriori (MAP) probability estimates of the posteriors, i.e., the value associated with the highest probability density (the "peak" of the posterior distribution) which serves as an estimation of the mode for continuous parameters. To further illustrate the interpretability of the posterior distributions, we report 89% posterior highest density intervals (HDI), which were deemed to be stable in Bayesian analyses (Kruschke, 2014; Makowski et al., 2019). The HDI is a generalization of the concept of the mode, but it is an interval rather than a single value.

Results

We began by inspecting the within-tasks effects for each administration time. Table 1 illustrates RTs and accuracy rates. As can be seen in Table 1, the task yielded significant congruency effects in both administrations.

To assess test–retest reliability, we began by evaluating correlations for the congruency conditions and effects by calculating Pearson's r correlations between Time 1 and Time 2 for each pair of values. As can be seen in Table 2, Pearson's r test–retest measures for the congruent, incongruent, and neutral conditions resulted in acceptable values between .72 and .75. However, in line with the "reliability

paradox," when inspecting test–retest correlations for the congruency, interference, and reversed facilitation effects, test–retest measures dropped significantly, yielding weak correlations—between the values of .22 and .29.

Next, we assessed test–retest correlations by using the Bayesian generative model method. In comparison to the traditional test–retest (Pearson) evaluation, the generative model method resulted in a modest improvement for the test–retest values for the congruent, incongruent, and neutral trials, yielding acceptable to good correlations, between .77 and .79 (see Table 2). More importantly, the generative model method resulted in considerably higher test–retest values for the congruency, interference, and reversed facilitation effects. The affordances task yielded acceptable to good test–retest measures for all three congruency effects, yielding correlations between .70 and .83 (see Table 2).

Table 2 Test–retest correlations between performance at Time 1 and Time 2 for the affordances task estimated by Pearson's r correlations and the Bayesian generative model

Condition / effect	r	95% CI	MAP	89% HDI
Congruent	.72	[.66, .77]	.79	[.75, .83]
Neutral	.75	[.68, .78]	.78	[.74, .82]
Incongruent	.73	[.69, .79]	.77	[.73, .82]
Congruency	.26	[.15, .36]	.77	[.58, .92]
Interference	.29	[.18, .39]	.83	[.65, .94]
Reversed facilitation	.22	[.11, .33]	.70	[.46, .89]

Congruency = RT for incongruent minus RT for congruent condition. Interference = RT for incongruent minus RT for neutral condition. Reversed facilitation = RT for neutral minus RT for congruent condition. An elaborated table of Pearson's r correlations that includes t and p values is presented in section S2 of the Supplementary Material

Discussion

In the present study, we evaluated the reliability of the affordances task for the first time. To that end, we administered the task online twice, with a one-week gap. We assessed the task's test–retest reliability both by using traditional summary statistics and by employing a hierarchical Bayesian generative model that was recently suggested as more suitable for the assessment of individual differences (Chen et al., 2021; Haines et al., 2020; Rouder & Haaf, 2019; Snijder et al., 2022). The task's online administration replicated common group-level results obtained on in-lab experiments, yielding congruency, interference, and reversed facilitation effects (see Table 1). Test–retest correlations for the three congruency conditions were satisfactory in both traditional summary statistics and the Bayesian generative model. However, for the congruency effects (congruency, interference, and reversed facilitation), which are the more important measures in the task, the test–retest correlations obtained by the use of traditional summary statistics were weak whilst the employment of the Bayesian generative model improved those correlations considerably, resulting in satisfactory and reliable test–retest estimations. These results raise several important points for consideration.

First, the current findings provide the first assessment of the affordances task test–retest reliability, indicating the task's stability for the study of individual differences. These findings may prove significant for future investigations of visuomotor and neurocognitive processes. Recent studies have demonstrated that the affordances task consists of cognitive conflicts at both the task level and the level of response (Littman & Kalanthroff, 2021, 2022). Response conflict manifests in the task as a conflict between responding with one's right versus left hand, illustrated by the longer RT to incongruent than to congruent trials (i.e., the congruency effect). Task conflict, which evolves between competing task demands (Kalanthroff et al., 2018a; Littman et al., 2019), manifests as a conflict between the goal-directed object classification task versus the stimulus-driven object grasping task. Hence, task conflict is indicated by longer RT to congruent (conflict-laden) than to neutral (conflict-free) trials (i.e., the reversed facilitation effect, see Littman & Kalanthroff, 2022). The current results provide evidence of good test–retest reliability for the congruency and reversed facilitation effects, thus supporting the task's reliability in the assessment of task and response conflicts. Importantly, while past studies mainly demonstrated the emergence of task conflict under conditions that trigger mental reactions such as word-reading in the Stroop task (Goldfarb & Henik, 2007; Parris, 2014) and object recognition in the object-interference task (La Heij et al., 2010; La Heij & Boelens, 2011; Prevor & Diamond, 2005), the affordances task is the

first to demonstrate the emergence of task conflict under conditions that trigger a behavioral reaction (object-grasping). As such, the affordances task serves as a nonlinguistic, behavioral measure of task conflict that is potentially closer to participants' everyday experiences. The current findings also illustrate the affordances task as a promising tool for the assessment of control over stimulus-driven habitual behaviors in healthy populations as well as in pathological populations characterized by increased reliance on stimulus-driven habitual behaviors, such as obsessive-compulsive disorder patients (Gillan et al., 2014, 2015; Kalanthroff et al., 2017, 2018b; Robbins et al., 2012), patients with substance use or behavioral addictions (Voon et al., 2015), and individuals suffering from a pre-supplementary motor area brain lesion (Haggard, 2008). Importantly, while stimulus-driven habitual behaviors have been demonstrated using various tasks, the current findings support the use of the affordances task as a unique measure of the specific cognitive control impairments that result in increased reliance on stimulus-driven habitual behaviors.

An important point regarding the affordances task needs to be acknowledged. Although many researchers attribute the affordances effect to the automatic activation of grasping responses, an alternative view has been suggested. According to this suggestion, the affordances effect represents a spatial correspondence effect, essentially similar to the Simon effect for stimulus location, and not grasping tendencies (Proctor & Miles, 2014). According to this approach, the effect is not triggered by a conflict between a correct response behavior and an incongruent activation of a stimulus-driven behavior, but rather by a conflict between a correct response behavior and an incongruent spatial cue. In other words, this alternative approach suggests that the conflict would be evident regardless of the graspability characteristics of the presented stimulus, since only its asymmetrical spatial form determines the conflict. Behavioral studies which inspected the two alternatives yielded mixed results: while some concluded that the observed affordances effects may be explained by a mere spatial correspondence effect (Cho & Proctor, 2010; Proctor et al., 2017; Song et al., 2014; Xiong et al., 2019), others have demonstrated that dissociable Simon and affordances effects can co-occur and that the affordances effect may emerge even in the absence of spatial correspondence (Azaad & Laham, 2019; Buccino et al., 2009; Iani et al., 2019; Netelenbos & Gonzalez, 2015; Pappas, 2014; Saccone et al., 2016; Scerrati et al., 2020; Symes et al., 2005). Importantly, a wide body of brain imaging studies has demonstrated the activation of premotor areas when participants view manipulable objects (Chao & Martin, 2000; Creem-Regehr & Lee, 2005; Grafton et al., 1997; Grezes & Decety, 2002; Proverbio et al., 2011), an activation which is absent in classic Simon tasks (e.g., Kerns, 2006),

and unique patterns of brain activity for manipulable objects that go beyond the effects of spatial correspondence (Bucino et al., 2009; Rice et al., 2007). Nonetheless, the findings of recent studies refined the initial concept of complete automaticity of the affordances effect and suggested that the affordances effect becomes more behaviorally evident when objects are presented in their functional orientation (Bub et al., 2018; Masson et al., 2011), and under conditions which emphasize the object's graspability (Girardi et al., 2010; Lu & Cheng, 2013). To ascertain the emergence of an affordances effect, we followed the specific suggestions made by these studies. In doing so, we believe that the current study results reflect a reliable measure for control over motor stimulus-driven behavior.

Second, the current study also allows us to evaluate the task's functioning and reliability under online administration conditions. In recent years, the online administration of cognitive tasks has gained popularity due to its ability to save resources, allow large sample sizes, and reach diverse populations across the globe (Feenstra et al., 2018; Gillan & Daw, 2016; Hansen et al., 2016; Haworth et al., 2007; Ruano et al., 2016). This tendency became even more prominent following the COVID-19 pandemic when the administration of in-lab experiments became limited or impossible for periods of time. Recently, a wide body of studies has reported encouraging findings following online administrations of a variety of cognitive tasks (Anwyl-Irvine et al., 2020; Crump et al., 2013; de Leeuw & Motz, 2016; Hilbig, 2016; Ratcliff & Hendrickson, 2021; Semmelmann & Weigelt, 2017; Simcox & Fiez, 2014). Chiefly, these studies reported results that were comparable to those typically obtained under in-lab administrations. The results of the current study are comparable to those of previous studies which used similar task designs in a laboratory setting (e.g., Littman & Kalanthroff, 2021, 2022; Saccone et al., 2016; Tucker & Ellis, 1998). Specifically, a comparison of the current study results to those reported by Littman and Kalanthroff (2022), Experiment 1, which used an identical design but was administered in the lab, yielded very similar results, albeit with minor differences in general RTs, which were somewhat shorter in the current study than those reported by Littman and Kalanthroff (2022). This full data is presented in section S3 of the Supplementary Material. Most importantly, the effects found in the current study were all in the same direction as the ones reported by Littman and Kalanthroff (2022), and were all significant, yielding medium to large effect sizes. Furthermore, the current results provide essential data regarding the reliability of an online administration of the task, together with the application of generative modeling to behavioral data obtained online. Alongside their advantages, web-based experiments are limited to the extent that administration may be less standardized in comparison to in-lab administration and may contain additional noise sources. Here, the

replication of the task's effects under these (noisier) conditions strengthens their replicability and the utility of the web-based administration of the affordances task.

Lastly, the inspection of test–retest reliability using a traditional method of assessment (Pearson's r) resulted in weak test–retest correlations for the congruency, interference, and reversed facilitation effects. These findings replicate the “reliability paradox” that is often observed when using summary statistics to assess individual differences in cognitive tasks that yield robust group-level effects (Haines et al., 2020; Rouder & Haaf, 2019), typically resulting in low estimates of the congruency effects (Bender et al., 2016; Hedge et al., 2018; Paap & Sawi, 2016; Soveri et al., 2018; Strauss et al., 2005). Following this, the application of the hierarchical Bayesian generative model resulted in a significant improvement in test–retest evaluation of the congruency, interference, and reversed facilitation effects, all yielding acceptable or good test–retest reliability. These results are in line with recent findings that illustrated the utility of generative models in the assessment of individual differences features (Chen et al., 2021; Haines et al., 2020; Rouder & Haaf, 2019). Recently, Haines et al. (2020) have demonstrated how the employment of generative models results in richer and more accurate test–retest estimations for a variety of well-established cognitive paradigms including Stroop, flanker, and Posner tasks. Additionally, Chen et al. (2021) indicated how the use of generative models accounts for trial-level variability and incorporates it into the model, allowing for a more precise evaluation of reliability in comparison to the summary statistics approach, in which trial-level variability is considered as measurement error. Importantly, the employment of generative models does not automatically result in an inflation of test–retest measures but does so only when such changes are warranted by the data (see Haines et al., 2020). The results of the current study are in line with the recent findings presented by Chen et al. (2021) and Haines et al. (2020), demonstrating the importance of employing finer, more able tools (such as Bayesian generative models) for the psychometric assessment of cognitive tasks. Such methods may deepen our understanding of the tasks themselves, their psychometric properties, and the cognitive structures they are designed to measure.

The findings from our study demonstrate that the affordances task can yield reliable individual differences. However, this is only the first step in a broader psychometric investigation. It is crucial to further examine the variability of these individual differences for clinical use and to determine their relationships to other constructs in the larger context. Additionally, our study suggests that Bayesian hierarchical models are an effective method for understanding these individual differences (Draheim et al., 2019, 2021), and it is recommended to continue using this approach to account for uncertainty in the affordances task. Further

research is needed to fully understand the psychometric potential of the affordances task.

Conclusion

The affordances task can serve as an important tool to study aspects of cognitive control and visuomotor functioning. In the current study, we assessed the task's test–retest reliability for the first time by using a hierarchical Bayesian generative model in an online administration. The affordances task yielded good test–retest properties, supporting its applicability in the study of individual differences. The employment of the generative model replicated recent findings that demonstrated its higher precision in the assessment of test–retest reliability in comparison to traditional methods of assessment which are based on summary statistics. The employment of Bayesian generative models may be used for future evaluations of individual differences and the reliability of cognitive tasks.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.3758/s13428-023-02131-3>.

Acknowledgments We thank Maya Schonbach, Nathaniel Haines, Eran Eldar, Hadar Naftalovich, and Eldad Keha for their helpful input and support during the work process of this article.

Data and code availability The experiment was not pre-registered. The data and code for the experiment are available at https://osf.io/rs8n9/?view_only=bfb5e272c6e34498b530795d9641753c

Funding This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors. RL and SH are supported by the Israel Science Foundation (grant no. 1341/18). This source was not involved in this study at any stage or in any aspect.

Declarations

Ethical approval This study was performed in line with the principles of the Declaration of Helsinki. Approval was granted by the Hebrew University institutional ethics committee (HUJI-500119).

Consent Informed consent was obtained from all participants prior to their participation in the experiment.

Competing interests The authors have no competing interests to declare.

References

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods*, 52(1), 388–407.

Azaad, S., & Laham, S. M. (2019). Sidestepping spatial confounds in object-based correspondence effects: The Bimanual Affordance Task (BMAT). *Quarterly Journal of Experimental Psychology*, 72(11), 2605–2613.

Barch, D. M., Carter, C. S., Executive, C. N. T. R. I. C. S., & Committee. (2008). Measurement issues in the use of cognitive neuroscience tasks in drug development for impaired cognition in schizophrenia: a report of the second consensus building conference of the CNTRICS initiative. *Schizophrenia Bulletin*, 34(4), 613–618.

Bender, A. D., Filmer, H. L., Garner, K. G., Naughtin, C. K., & Dux, P. E. (2016). On the relationship between response selection and response inhibition: An individual differences approach. *Attention, Perception, & Psychophysics*, 78(8), 2420–2432.

Bub, D. N., Masson, M. E. J., & Kumar, R. (2018). Time course of motor affordances evoked by pictured objects and words. *Journal of Experimental Psychology: Human Perception and Performance*, 44(1), 53–68.

Buccino, G., Sato, M., Cattaneo, L., Rodà, F., & Riggio, L. (2009). Broken affordances, broken objects: A TMS study. *Neuropsychologia*, 47(14), 3074–3078.

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., ..., Riddell, A. (2017). Stan: A probabilistic programming language. *Grantee Submission*, 76(1), 1–32.

Chao, L. L., & Martin, A. (2000). Representation of manipulable man-made objects in the dorsal stream. *NeuroImage*, 12(4), 478–484.

Chen, G., Pine, D. S., Brotman, M. A., Smith, A. R., Cox, R. W., & Haller, S. P. (2021). Trial and error: A hierarchical modeling approach to test-retest reliability. *NeuroImage*, 245, 118647.

Cho, D. T., & Proctor, R. W. (2010). The object-based Simon effect: Grasping affordance or relative location of the graspable part? *Journal of Experimental Psychology: Human Perception and Performance*, 36(4), 853–861.

Creem-Regehr, S. H., & Lee, J. N. (2005). Neural representations of graspable objects: are tools special? *Cognitive Brain Research*, 22(3), 457–469.

Crump, M. J. C., McDonnell, J. V., & Gureckis, T. M. (2013). Evaluating Amazon's Mechanical Turk as a tool for experimental behavioral research. *PloS One*, 8(3), e57410.

Dang, J., King, K. M., & Inzlicht, M. (2020). Why are self-report and behavioral measures weakly correlated? *Trends in Cognitive Sciences*, 24(4), 267–269.

de Leeuw, J. R., & Motz, B. A. (2016). Psychophysics in a Web browser? Comparing response times collected with JavaScript and Psychophysics Toolbox in a visual search task. *Behavior Research Methods*, 48(1), 1–12.

Draheim, C., Mashburn, C. A., Martin, J. D., & Engle, R. W. (2019). Reaction time in differential and developmental research: A review and commentary on the problems and alternatives. *Psychological Bulletin*, 145(5), 508–535.

Draheim, C., Tsukahara, J. S., Martin, J. D., Mashburn, C. A., & Engle, R. W. (2021). A toolbox approach to improving the measurement of attention control. *Journal of Experimental Psychology: General*, 150(2), 242–275.

Elliott, M. L., Knodt, A. R., Ireland, D., Morris, M. L., Poulton, R., Ramrakha, S., ..., Hariri, A. R. (2019). Poor test-retest reliability of task-fMRI: New empirical evidence and a meta-analysis. *BioRxiv*, 681700.

Enkavi, A. Z., Eisenberg, I. W., Bissett, P. G., Mazza, G. L., MacKinnon, D. P., Marsch, L. A., & Poldrack, R. A. (2019). Large-scale analysis of test–retest reliabilities of self-regulation measures. *Proceedings of the National Academy of Sciences*, 116(12), 5472–5477.

Feenstra, H. E. M., Vermeulen, I. E., Murre, J. M. J., & Schagen, S. B. (2018). Online self-administered cognitive testing using the Amsterdam cognition scan: Establishing psychometric properties and normative data. *Journal of Medical Internet Research*, 20(5), e192.

Garrido-Vásquez, P., & Schubö, A. (2014). Modulation of visual attention by object affordance. *Frontiers in Psychology*, 5, 59.

- Gawronski, B., Morrison, M., Phills, C. E., & Galdi, S. (2017). Temporal stability of implicit and explicit measures: A longitudinal analysis. *Personality and Social Psychology Bulletin*, *43*(3), 300–312.
- Gelman, A., & Pardoe, I. (2006). Bayesian measures of explained variance and pooling in multilevel (hierarchical) models. *Technometrics*, *48*(2), 241–251.
- Geusebroek, J.-M., Burghouts, G. J., & Smeulders, A. W. M. (2005). The Amsterdam library of object images. *International Journal of Computer Vision*, *61*(1), 103–112.
- Gibson, J. J. (1979). *The Ecology Approach to Visual Perception* (Classic ed.). Psychology Press.
- Gillan, C. M., & Daw, N. D. (2016). Taking psychiatry research online. *Neuron*, *91*(1), 19–23.
- Gillan, C. M., Morein-Zamir, S., Urcelay, G. P., Sule, A., Voon, V., Apergis-Schoute, A. M., Fineberg, N. A., Sahakian, B. J., & Robbins, T. W. (2014). Enhanced avoidance habits in obsessive-compulsive disorder. *Biological Psychiatry*, *75*(8), 631–638.
- Gillan, C. M., Otto, A. R., Phelps, E. A., & Daw, N. D. (2015). Model-based learning protects against forming habits. *Cognitive, Affective, & Behavioral Neuroscience*, *15*(3), 523–536.
- Girardi, G., Lindemann, O., & Bekkering, H. (2010). Context effects on the processing of action-relevant object features. *Journal of Experimental Psychology: Human Perception and Performance*, *36*(2), 330–340.
- Goldfarb, L., & Henik, A. (2007). Evidence for task conflict in the Stroop effect. *Journal of Experimental Psychology: Human Perception and Performance*, *33*(5), 1170–1176.
- Goslin, J., Dixon, T., Fischer, M. H., Cangelosi, A., & Ellis, R. (2012). Electrophysiological examination of embodiment in vision and action. *Psychological Science*, *23*(2), 152–157.
- Grafton, S. T., Fadiga, L., Arbib, M. A., & Rizzolatti, G. (1997). Premotor cortex activation during observation and naming of familiar tools. *Neuroimage*, *6*(4), 231–236.
- Grezes, J., & Decety, J. (2002). Does visual perception of object afford action? *Evidence from a neuroimaging study*. *Neuropsychologia*, *40*(2), 212–222.
- Haines, N., Kvam, P. D., Irving, L. H., Smith, C., Beauchaine, T. P., Pitt, M. A., ..., Turner, B. (2020). *Learning from the Reliability Paradox: How Theoretically Informed Generative Models Can Advance the Social, Behavioral, and Brain Sciences*. PsyArXiv.
- Haggard, P. (2008). Human volition: towards a neuroscience of will. *Nature Reviews Neuroscience*, *9*(12), 934–946.
- Hansen, T. I., Lehn, H., Evensmoen, H. R., & Håberg, A. K. (2016). Initial assessment of reliability of a self-administered web-based neuropsychological test battery. *Computers in Human Behavior*, *63*, 91–97.
- Haworth, C. M. A., Harlaar, N., Kovas, Y., Davis, O. S. P., Oliver, B. R., Hayiou-Thomas, M. E., Frances, J., Busfield, P., McMillan, A., & Dale, P. S. (2007). Internet cognitive testing of large samples needed in genetic research. *Twin Research and Human Genetics*, *10*(4), 554–563.
- Hedge, C., Powell, G., & Sumner, P. (2018). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behavior Research Methods*, *50*(3), 1166–1186.
- Hilbig, B. E. (2016). Reaction time effects in lab-versus Web-based research: Experimental evidence. *Behavior Research Methods*, *48*(4), 1718–1724.
- Iani, C., Ferraro, L., Maiorana, N. V., Gallese, V., & Rubichi, S. (2019). Do already grasped objects activate motor affordances? *Psychological Research*, *83*(7), 1363–1374.
- Kalanthroff, E., Davelaar, E. J., Henik, A., Goldfarb, L., & Usher, M. (2018a). Task conflict and proactive control: A computational theory of the Stroop task. *Psychological Review*, *125*(1), 59–82.
- Kalanthroff, E., Henik, A., Simpson, H. B., Todder, D., & Anholt, G. E. (2017). To do or not to do? Task control deficit in obsessive-compulsive disorder. *Behavior Therapy*, *48*(5), 603–613.
- Kalanthroff, E., Steinman, S. A., Schmidt, A. B., Campeas, R., & Simpson, H. B. (2018b). Piloting a Personalized Computerized Inhibitory Training Program for Individuals with Obsessive-Compulsive Disorder. *Psychotherapy and Psychosomatics*, *87*(1), 52–54.
- Kerns, J. G. (2006). Anterior cingulate and prefrontal cortex activity in an fMRI study of trial-to-trial adjustments on the Simon task. *Neuroimage*, *33*(1), 399–405.
- Kruschke, J. (2014). *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan*. Academic Press.
- La Heij, W., & Boelens, H. (2011). Color-object interference: Further tests of an executive control account. *Journal of Experimental Child Psychology*, *108*(1), 156–169.
- La Heij, W., Boelens, H., & Kuipers, J.-R. (2010). Object interference in children's colour and position naming: Lexical interference or task-set competition? *Language and Cognitive Processes*, *25*(4), 568–588.
- Littman, R., Keha, E., & Kalanthroff, E. (2019). Task Conflict and Task Control: A Mini-review. *Frontiers in Psychology*, *10*, 1598.
- Littman, R., & Kalanthroff, E. (2021). Control over task conflict in the stroop and affordances tasks: An individual differences study. *Psychological Research*, *85*(6), 2420–2427.
- Littman, R., & Kalanthroff, E. (2022). Neutral affordances: Task conflict in the affordances task. *Consciousness and Cognition*, *97*, 103262.
- Lo, S., & Andrews, S. (2015). To transform or not to transform: Using generalized linear mixed models to analyse reaction time data. *Frontiers in psychology*, *6*, 1171.
- Lu, J., & Cheng, L. (2013). Perceiving and interacting affordances: A new model of human-affordance interactions. *Integrative Psychological and Behavioral Science*, *47*(1), 142–155.
- MacLeod, C. M. (1991). Half a century of research on the Stroop effect: an integrative review. *Psychological Bulletin*, *109*(2), 163–203.
- Makowski, D., Ben-Shachar, M. S., & Lüdtke, D. (2019). bayestestR: Describing effects and their uncertainty, existence and significance within the Bayesian framework. *Journal of Open Source Software*, *4*(40), 1541.
- Makris, S., Grant, S., Hadar, A. A., & Yarrow, K. (2013). Binocular vision enhances a rapidly evolving affordance priming effect: Behavioural and TMS evidence. *Brain and Cognition*, *83*(3), 279–287.
- Masson, M. E. J., Bub, D. N., & Breuer, A. T. (2011). Priming of reach and grasp actions by handled objects. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(5), 1470–1484.
- McElreath, R. (2020). *Statistical rethinking: A Bayesian course with examples in R and Stan*. Chapman & Hall/CRC.
- Netelenbos, N., & Gonzalez, C. L. R. (2015). Is that graspable? Let your right hand be the judge. *Brain and Cognition*, *93*, 18–25.
- Paap, K. R., & Sawi, O. (2016). The role of test-retest reliability in measuring individual and group differences in executive functioning. *Journal of Neuroscience Methods*, *274*, 81–93.
- Pappas, Z. (2014). Dissociating Simon and affordance compatibility effects: Silhouettes and photographs. *Cognition*, *133*(3), 716–728.
- Parris, B. A. (2014). Task conflict in the Stroop task: When Stroop interference decreases as Stroop facilitation increases in a low task conflict context. *Frontiers in Psychology*, *5*, 1182.
- Prevor, M. B., & Diamond, A. (2005). Color-object interference in young children: A Stroop effect in children 3½–6½ years old. *Cognitive Development*, *20*(2), 256–278.
- Proctor, R. W., Lien, M.-C., & Thompson, L. (2017). Do silhouettes and photographs produce fundamentally different object-based correspondence effects? *Cognition*, *169*, 91–101.

- Proctor, R. W., & Miles, J. D. (2014). Does the concept of affordance add anything to explanations of stimulus–response compatibility effects? In B. H. Ross (Ed.), *The psychology of learning and motivation* (Vol. 60, pp. 227–266). Academic Press.
- Proverbio, A. M., Adorni, R., & D’aniello, G. E. (2011). 250 ms to code for action affordance during observation of manipulable objects. *Neuropsychologia*, *49*(9), 2711–2717.
- Ratcliff, R., & Hendrickson, A. T. (2021). Do data from mechanical Turk subjects replicate accuracy, response time, and diffusion modeling results? *Behavior Research Methods*, *53*(6), 2302–2325.
- Rice, N. J., Valyear, K. F., Goodale, M. A., Milner, A. D., & Culham, J. C. (2007). Orientation sensitivity to graspable objects: an fMRI adaptation study. *Neuroimage*, *36*, T87–T93.
- Robbins, T. W., Gillan, C. M., Smith, D. G., de Wit, S., & Ersche, K. D. (2012). Neurocognitive endophenotypes of impulsivity and compulsivity: Towards dimensional psychiatry. *Trends in Cognitive Sciences*, *16*(1), 81–91.
- Romeu, R. J., Haines, N., Ahn, W.-Y., Busemeyer, J. R., & Vassileva, J. (2020). A computational model of the Cambridge gambling task with applications to substance use disorders. *Drug and Alcohol Dependence*, *206*, 107711.
- Rouder, J. N., & Haaf, J. M. (2019). A psychometrics of individual differences in experimental tasks. *Psychonomic Bulletin & Review*, *26*(2), 452–467.
- Rouder, J. N., Province, J. M., Morey, R. D., Gomez, P., & Heathcote, A. (2015). The lognormal race: A cognitive-process model of choice and latency with desirable psychometric properties. *Psychometrika*, *80*(2), 491–513.
- Ruano, L., Sousa, A., Severo, M., Alves, I., Colunas, M., Barreto, R., ..., Bento, V. (2016). Development of a self-administered web-based test for longitudinal cognitive assessment. *Scientific Reports*, *6*(1), 1–10.
- Saccone, E. J., Churches, O., & Nicholls, M. E. R. (2016). Explicit spatial compatibility is not critical to the object handle effect. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(10), 1643–1653.
- Schuch, S., Philipp, A. M., Maulitz, L., & Koch, I. (2022). On the reliability of behavioral measures of cognitive control: retest reliability of task-inhibition effect, task-preparation effect, Stroop-like interference, and conflict adaptation effect. *Psychological research*, *86*(7), 2158–2184.
- Schulz, L., Ischebeck, A., Wriessnegger, S. C., Steyrl, D., & Müller-Putz, G. R. (2018). Action affordances and visuo-spatial complexity in motor imagery: An fMRI study. *Brain and Cognition*, *124*, 37–46.
- Scerrati, E., Iani, C., Lugli, L., Nicoletti, R., & Rubichi, S. (2020). Do my hands prime your hands? The hand-to-response correspondence effect. *Acta Psychologica*, *203*, 103012.
- Semmelmann, K., & Weigelt, S. (2017). Online psychophysics: Reaction time effects in cognitive experiments. *Behavior Research Methods*, *49*(4), 1241–1260.
- Simcox, T., & Fiez, J. A. (2014). Collecting response times using amazon mechanical turk and adobe flash. *Behavior Research Methods*, *46*(1), 95–111.
- Snijder, J. P., Tang, R., Bugg, J., Conway, A. R., & Braver, T. (2022). *On the Psychometric Evaluation of Cognitive Control Tasks: An Investigation with the Dual Mechanisms of Cognitive Control (DMCC) Battery*. PsyArXiv.
- Song, X., Chen, J., & Proctor, R. W. (2014). Correspondence effects with torches: Grasping affordance or visual feature asymmetry? *Quarterly Journal of Experimental Psychology*, *67*(4), 665–675.
- Soveri, A., Lehtonen, M., Karlsson, L. C., Lukasik, K., Antfolk, J., & Laine, M. (2018). Test–retest reliability of five frequently used executive tasks in healthy adults. *Applied Neuropsychology: Adult*, *25*(2), 155–165.
- Strauss, G. P., Allen, D. N., Jorgensen, M. L., & Cramer, S. L. (2005). Test–retest reliability of standard and emotional stroop tasks: An investigation of color-word and picture-word versions. *Assessment*, *12*(3), 330–337.
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*(6), 643–662.
- Symes, E., Ellis, R., & Tucker, M. (2005). Dissociating object-based and space-based affordances. *Visual Cognition*, *12*(7), 1337–1361.
- Tipper, S. P., Paul, M. A., & Hayes, A. E. (2006). Vision-for-action: The effects of object property discrimination and action state on affordance compatibility effects. *Psychonomic Bulletin & Review*, *13*(3), 493–498.
- Tucker, M., & Ellis, R. (1998). On the relations between seen objects and components of potential actions. *Journal of Experimental Psychology: Human Perception and Performance*, *24*(3), 830–846.
- Tucker, M., & Ellis, R. (2001). The potentiation of grasp types during visual object categorization. *Visual Cognition*, *8*(6), 769–800.
- Tucker, M., & Ellis, R. (2004). Action priming by briefly presented objects. *Acta Psychologica*, *116*(2), 185–203.
- von Bastian, C. C., Blais, C., Brewer, G., Gyurkovics, M., Hedge, C., Kalamala, P., ..., Rouder, J. N. (2020). *Advancing the understanding of individual differences in attentional control: Theoretical, methodological, and analytical considerations*. PsyArXiv.
- Voon, V., Derbyshire, K., Rück, C., Irvine, M. A., Worbe, Y., Enander, J., ..., Sahakian, B. J. (2015). Disorders of compulsivity: a common bias towards learning habits. *Molecular Psychiatry*, *20*(3), 345–352.
- Wennerhold, L., Friese, M., & Vazire, S. (2020). Why self-report measures of self-control and inhibition tasks do not substantially correlate. *Collabra: Psychology*, *6*(1), 9.
- Wilf, M., Holmes, N. P., Schwartz, I., & Makin, T. R. (2013). Dissociating between object affordances and spatial compatibility effects using early response components. *Frontiers in Psychology*, *4*, 591.
- Xiong, A., Proctor, R. W., & Zelaznik, H. N. (2019). Visual salience, not the graspable part of a pictured eating utensil, grabs attention. *Attention, Perception, & Psychophysics*, 1–10.

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.