

The role of subjective linear orders in probabilistic inferences

Rüdiger F. Pohl · Benjamin E. Hilbig

Published online: 28 September 2012
© Psychonomic Society, Inc. 2012

Abstract In probabilistic inferences concerning which of two objects has the larger criterion value (e.g., which of two cities has more inhabitants), participants may recognize both objects, only one, or neither. According to the mental-toolbox approach, different decision strategies exist for each of these cases, utilizing different probabilistic cues. Possibly, however, participants use these cues to build a subjective rank order that involves all objects, irrespective of their recognition status. The decision process then simply utilizes the distance between two objects in one's subjective order. We tested the role of such linear orders in reanalyses of existing data and in a new experiment. Participants' choices and decision times were determined both by subjective rank-order distances and by the recognition status of the compared objects. To integrate these theoretically inconsistent findings, we discuss the role of the evidential difference (or the degree of conflict) between two objects.

Keywords Inferences · Heuristics · Paired comparisons · Linear order · Reaction times · Recognition · Decision making · Distance effect

Probabilistic inferences are a common task for decision makers, given that to-be judged criteria are rarely known conclusively. Often enough, judgments must be made on the basis of some probabilistic cues (e.g., Gigerenzer, Hoffrage, & Kleinbölting, 1991), and it is thus a central issue to understand how people perform such tasks. In typical

studies on probabilistic inferences, participants are shown pairs of objects (e.g., cities) and are asked to infer for each pair which of the two objects has the higher criterion value (e.g., population). Given that the answer is unknown, whereas both objects are familiar, participants may retrieve probabilistic cues about them (e.g., whether either has an international airport) to reach a decision (cf. Gigerenzer & Goldstein, 1996). If, however, only one or neither of the objects in a pair are known, other strategies are required. This is one of the main assumptions of the “adaptive toolbox” approach (Gigerenzer, Todd, & the ABC Research Group, 1999), which assumes that decision makers are equipped with a mental repertoire of specialized tools (mostly heuristics), selected ad hoc to solve a specific task.

As an alternative, Brown and Tan (2011) recently proposed that participants may engage in magnitude comparisons that are based on subjectively built orders of the objects involved. They tested a set of 16 cars (asking for the more expensive one in each pair) and found that participants' decision times were better explained by assuming magnitude comparisons based on subjective linear orders than by assuming the use of specific heuristics, as predicted by the toolbox approach. However, Brown and Tan's findings were limited to situations in which all of the objects are familiar. Thus, we here extend their approach to testing the role of linear orders with known and unknown objects (and all possible pairings of these). Also, we employ additional dependent measures, namely the deviation rate (measured in relation to one's subjective order) and the rate of inconsistencies across repeated trials. Finally, we analyze the data from two experiments, thus assessing the robustness of the observed effects.

In the following sections, we will more fully describe the idea of linear orders as the basis for magnitude comparisons and what the toolbox approach predicts, assuming multiple

R. F. Pohl (✉) · B. E. Hilbig
Department of Psychology, School of Social Sciences,
University of Mannheim,
68131 Mannheim, Germany
e-mail: pohl@psychologie.uni-mannheim.de

specific strategies. We will then describe tests of these predictions in a reanalysis of a published experiment (Hilbig & Pohl, 2009, Exp. 3) and in a new experiment.

Linear orders

Brown and Tan (2011) suggested that decision makers rely on magnitude comparisons based on the subjective linear order of objects—thus avoiding the need to switch tools from trial to trial. Prior research had shown that in comparative tasks in which well-learned objects are ordered along a criterion, people possess or build subjective linear orders and subsequently base their decisions on this mental representation (see, e.g., Banks, 1977; Moyer & Bayer, 1976; Moyer & Dumais, 1978; Parkman, 1971). The central and most robust finding from these studies is the so-called “symbolic distance effect”: The larger the distance between two objects in their underlying order, the shorter the reaction time (and the fewer errors). This corresponds to what Brown and Tan found for pairs of recognized cars differing in price. In addition to this empirical argument, using such a linear order in repeated paired comparisons would arguably necessitate less effort than having to work through several potential decision strategies over and over (see Pohl, 2011).

Thus, it seems plausible that participants, if confronted with a new set of objects, quickly develop a subjective linear order of these objects and subsequently base comparisons on this order. Importantly, not only recognized objects could be so ordered, but unrecognized ones, too (cf. Marewski, Gaissmaier, Schooler, Goldstein, & Gigerenzer, 2010). For example, cues connected to the unrecognized object’s name might be used: Some city names simply sound big, and others small (see McCloy, Beaman, Frosch, & Goddard, 2010, for examples, and Marewski, Pohl, & Vitouch, 2011, for empirical findings on this topic). Or, participants might simply know more about an object than they admit, so that not all objects called “unrecognized” by participants would necessarily be truly novel (cf. Erdfelder, Küpper-Tetzl, & Mattern, 2011). Indeed, this could also explain why performance for pairs of unrecognized objects is often better than chance (see, e.g., Hilbig, Erdfelder, & Pohl, 2010).¹ In some cases, recognized objects might even score lower than unrecognized ones (cf. Smithson, 2010). For example, a recognized city could have negative cue values (e.g., no university, no airport, no industry) and therefore receive a relatively low position in the linear order. All of

this information would be useful in ordering objects, both recognized and unrecognized.

If, as this reasoning implies, all objects are represented on the *same* continuum, a distance effect should be observed for all types of pairs—that is, with both objects known (labeled “knowledge case”), with only one object known (“recognition case”), or with no objects known (“guessing case”): The farther apart two objects in one’s subjective order, the faster the decision. By the same token, the probability of decisions deviating from one’s order (or showing inconsistencies) should decrease with the distance between objects, again for all types of pairs.

The toolbox approach

The toolbox approach assumes that decision makers possess a repertoire of specific strategies that are tailored to certain situations (Gigerenzer et al., 1999). When inferring which of two objects has the larger criterion value, several different cases and corresponding strategies would need to be differentiated (see Pohl, 2011, for a summary of these strategies). What would these strategies predict for the decision times for pairs of objects differing in their distances to each other on the underlying dimension?

In the case of two recognized objects (knowledge case), one potential strategy would be the fluency heuristic (FH; Hertwig, Herzog, Schooler, & Reimer, 2008; Schooler & Hertwig, 2005). This strategy exploits the difference in recognition times of the two objects and infers that if the difference is large enough, the object recognized more speedily would have the larger criterion value—ignoring any further cues or information (but see Hilbig, Erdfelder, & Pohl, 2011, for a critical evaluation of this aspect). Assuming that recognition times (fluency) decrease with objects’ positions on the continuum (Hertwig et al., 2008; Herzog & Hertwig, *in press*; Marewski & Schooler, 2011), one could deduce that recognition time differences should be larger for two distant than for two close objects. Thus, the FH could predict a distance effect in decision times. Another strategy for knowledge cases would be the take-the-best heuristic (TTB; Gigerenzer & Goldstein, 1996), which assumes that probabilistic cues are searched in the order of their validity until one is found that discriminates between the two objects. The inference is then based on this cue alone (but see Ayala & Hochman, 2009; Bröder, 2000). If none is found, participants resort to guessing. Arguably, two close objects would have more identical cue values than two distant objects, implying that the time needed to find a discriminating cue should be greater for two close than for two distant objects (see Bröder & Gaissmaier, 2007). Thus, TTB could also predict a distance effect for pairs of recognized objects. In contrast, Brown and Tan (2011) argued that

¹ Hilbig et al. (2010) reported the results of ten data sets and found performance in cases of guessing to be slightly greater than 50 % in eight of the ten cases, which is more often than would be expected by chance, $p = .055$ (binomial test).

in their study TTB did not predict the observed distance effect, because the pairs that they used could have been answered by considering the same highly valid cue, thus implying no decision-time differences.

When only one object is recognized, the toolbox proposes to apply the recognition heuristic (RH; Goldstein & Gigerenzer, 1999, 2002): The decision maker simply infers that the recognized object is the larger one—again ignoring all further information (see Hilbig, 2010, for an overview of critical findings). Marewski et al. (2010) recently proposed that the RH will be more often used when recognition times for the recognized object are short rather than long. In other words, recognition times (fluency) may act as a cue not only in the FH, but also in the RH.² Thus, again assuming that recognition times decrease with the objects' relative positions, and further assuming that larger objects (with shorter recognition times) are, on average, more often included in pairs of two distant than in pairs of two close objects, a distance effect could be predicted: The fast RH would be more often applied in pairs of two distant than of two close objects.

Finally, if neither object is recognized, participants are typically assumed to know nothing about the two objects, implying that decision times should also not differ systematically between different distances.

A central prediction, independent of the distance between objects, concerns the decision times for different types of pairs (knowledge, recognition, and guessing). In the toolbox approach, the RH is assumed to be the first and foremost of all heuristics, and should thus lead to shorter decision times than would a strategy that must be selected if the RH cannot be applied (e.g., FH or TTB). So, on average, recognition cases should require less time than knowledge cases. Previous data have supported this prediction (e.g., Pachur & Hertwig, 2006), although only partially (Hilbig & Pohl, 2009). In any case, if all objects within a linear order were treated equally, one would not expect an effect of this nature. Main effects of the type of pair are thus difficult to reconcile with the linear-order perspective. We will return to this problem in the General Discussion.

Reanalyzed data

A previously published study (Hilbig & Pohl, 2009, Exp. 3) contained sufficient data to allow for a reanalysis that addresses the following questions: Is there evidence for the postulated distance effect in decision times? And, if so, does it hold for all types of pairs (i.e., recognition, knowledge, and

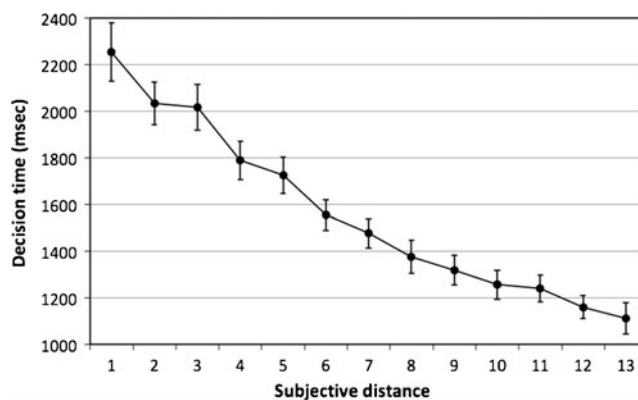


Fig 1 Mean decision times (in milliseconds) as a function of the distance in each participant's subjective rank order (data from Hilbig & Pohl, 2009, Exp. 3). Error bars represent one standard error of the mean

guessing cases)? In that experiment, 68 participants made inferences for 91 pairs of cities (consisting of the 14 largest Swiss cities, excluding the largest, Zurich), judging for each pair which of the two cities is more populous.

To obtain individual rank orders, we assigned ordinal ranks to each city according to the relative choice frequency of that city, separately for each participant. The mean rank correlation (Kendall's tau) between these subjective orders and the true order was $\tau = .47$ ($SD = .13$), $t(67) = 29.083$, $p < .001$. When computed separately for recognized and unrecognized cities, the respective mean correlations were $\tau = .56$ ($SD = .34$), $t(60) = 12.881$, and $\tau = .38$ ($SD = .26$), $t(60) = 11.517$, both $p < .001$. Thus, participants' subjective rank orders were typically valid; that is, they corresponded well with the true order of objects.

We initially analyzed the distance effect for the whole set (summing across types of pairs, leaving in the sample $N = 45$ who provided data for all distances).³ The results show a clear distance effect (see Fig. 1); that is, decision times decreased substantially from Distance 1 ($M = 2,329$ ms, $SD = 888$ ms) to Distance 13 ($M = 1,040$ ms, $SD = 449$ ms), $F(12, 528) = 38.937$, $p < .001$, $\eta_p^2 = .48$. Next, we split the data conditional on the type of pair. Because not all participants provided data for each type of pair and each distance (especially not for the larger ones, which are inherently more rare), we included only those data cells to which at least ten participants contributed data. As is shown in Fig. 2, all three types of pairs showed a distance effect. Linear regressions yielded similar slopes of -70 , -86 , and -90 ms for the recognition, knowledge, and guessing cases, respectively. Finally, we also found a significant main effect of type of pair on decision times (see Fig. 2), $F(2, 120) = 24.378$, $p < .001$, $\eta_p^2 = .09$. Specifically, we

² But note that the FH assumes that fluency differences between two objects are the cue, while here the relative fluency of one object (compared to all others) is taken as a cue.

³ About 1 % of the individual decision times were considered extreme (i.e., larger than the overall mean plus five times the standard deviation) and were excluded from the analysis.

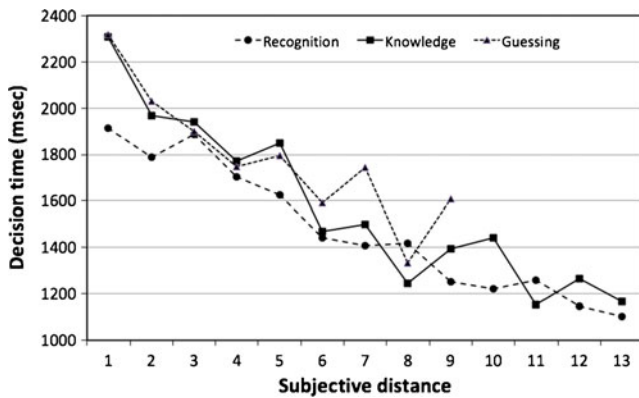


Fig 2 Mean decision times (in milliseconds) for recognition, knowledge, and guessing cases as a function of the distance in each participant's subjective rank order (data from Hilbig & Pohl, 2009, Exp. 3)

observed shorter decision times for recognition cases ($M = 1,518$ ms, $SD = 492$ ms) than for knowledge ($M = 1,852$ ms, $SD = 710$ ms) or guessing ($M = 2,000$ ms, $SD = 842$ ms) cases, whereas the latter two did not differ significantly.

In sum, there was clear evidence for a distance effect in decision times that held for each type of pair, thus supporting the linear-order view. By contrast, the main effect of type of pair is inconsistent with a linear-order perspective. However, one potential caveat of these results is that individual rank orders were not explicitly stated by the participants, but only indirectly reconstructed from their choice data. Thus, we ran a new experiment in which we assessed subjective orders directly and, in addition, sought further dependent measures beyond decision times.

Experiment

This experiment again used the typical city-size task. In addition, to test the robustness of the predicted distance effect across tasks, we manipulated (between subjects) whether participants were asked to indicate the more or the less populous city in each pair. Logically, both questions are equivalent and should not lead to different decision behavior. However, two previous studies that featured the same manipulation (Hilbig, Scholl, & Pohl, 2010; McCloy et al., 2010) both found that more choices were in line with the RH in the “larger” condition than in the “smaller” condition. Also, the first of these studies reported a main effect for decision times, which were shorter for the “larger” question than for the “smaller” question. The latter result corresponds to linear-order findings showing that the “larger” question yields faster answers for comparing large objects and the “smaller” question for comparing small objects—termed a “semantic congruity effect” (see, e.g., Banks, 1977; Banks, Fujii, & Kayra-Stuart, 1976; Moyer & Dumais, 1978). In most RH studies (including the present one), the materials

have consisted exclusively of “large” objects from the upper end of the criterion scale. Consequently, we expected to find a main effect of the question format on decision times. More interestingly, we aimed to test whether the distance effect would be the same for both questions (as Banks et al., 1976, had found).

Finally, we repeated the whole set of paired comparisons for all participants, thus allowing us to measure the consistency of inferences as an alternative measure, which has rarely been considered in the probabilistic-inference literature but which—according to the reasoning above—should corroborate the predicted distance effect.

Method

Participants and materials

A total of 80 students (64 female, 16 male; age: $M = 22.3$ years, $SD = 3.5$ years) participated in return for a flat fee payment of €8 or course credit. They were randomly assigned to one of two groups, differing only in the instructions that they received in the paired-comparison task (see the Procedure section). Group 1 consisted of 42 participants, and Group 2 of 38. As materials, we exhaustively paired the 11 largest Belgian cities (excluding Brussels, the largest one, in order to avoid influences of criterion knowledge—that is, knowledge about a city's rank position—that would make inferences obsolete; cf. Hilbig, Pohl, & Bröder, 2009; Pachur, Bröder, & Marewski, 2008), resulting in 55 pairs that were presented twice to each participant.

Procedure

First, participants were asked, for each single city (presented in random order), whether or not they recognized its name (indicated by pressing one of two keys on the keyboard). The reaction times of these answers (i.e., the recognition times) were recorded. Next followed the paired-comparison task. The 55 pairs of cities were presented one at a time and in random order. Participants were asked to indicate speedily and as accurately as possible which of the cities in a pair was the *larger* one (in Group 1) or the *smaller* one (in Group 2). Choices and reaction times were recorded. After completing the 55 pairs, the same set was presented again, but in a new random order and with altered positions (left/right) on the screen. This repetition was not announced or otherwise signified to the participants; thus, subjectively, the task consisted of 110 trials without any interruption. Finally, participants were asked to subjectively rank order the 11 cities by adjusting their relative positions on the screen (using a mouse) in an initially alphabetically ordered list. The experiment lasted about 30 min in total.

Results and discussion

Overall decision times

For all analyses, the median decision time of each person (per condition) served as the dependent measure. To test whether decision times differed between conditions, we ran a three-way ANOVA with Group (larger? vs. smaller?) as a between-subjects factor and Phase (1 vs. 2) and Type of Pair (recognition, knowledge, and guessing) as within-subjects factors.⁴ The ANOVA revealed several effects: (1) Decision times were shorter in recognition cases ($M = 1,107$ ms, $SD = 301$ ms) than in knowledge cases ($M = 1,468$ ms, $SD = 731$ ms) or guessing cases ($M = 1,397$ ms, $SD = 607$ ms), $F(2, 128) = 18.286$, $p < .001$, $\eta_p^2 = .18$, while the latter two types did not differ significantly.⁵ (2) The decision times of Group 1 ($M = 1,235$ ms, $SD = 624$ ms) trended toward being shorter than those of Group 2 ($M = 1,418$ ms, $SD = 550$ ms), $F(1, 64) = 3.35$, $p = .07$, $\eta_p^2 = .05$. In other words, the “larger” inferences were drawn faster than the logically equivalent “smaller” inferences. (3) Decision times were significantly shorter in Phase 2 ($M = 1,186$ ms, $SD = 468$ ms) than in Phase 1 ($M = 1,461$ ms, $SD = 673$ ms), $F(1, 64) = 53.723$, $p < .001$, $\eta_p^2 = .46$, which most likely resulted from learning or practice effects.

Subjective rank orders

The mean rank correlation (Kendall’s tau) between subjective orders and the true order was $\tau = .38$ ($SD = .19$), $t(79) = 18.534$, $p < .001$. When computed separately for recognized versus unrecognized cities, the respective mean correlations were $\tau = .34$ ($SD = .41$), $t(53) = 5.940$, and $\tau = .16$ ($SD = .33$), $t(79) = 4.269$, both $ps < .001$. Thus, participants’ subjective orders generally had some objective validity. Moreover, recognized cities received predominantly higher subjective ranks than did unrecognized ones. Consequently, in only 6.4 % of all recognition cases was an unrecognized city subjectively considered larger than a recognized one.

Distance effect

To test for distance effects, we analyzed decision times, deviation rates, and inconsistencies conditional upon the distance in subjective rank-order positions of the cities in a pair. Due to the nature of linear orders, the number of cases

⁴ Due to individually extreme portions of recognized or unrecognized cities, not all participants had pairs of all three types. Thus, only 34 participants from Group 1 and 32 from Group 2 were entered in the analysis.

⁵ Type of pair also interacted with phase and with group, mainly due to inconsistent differences between the decision times in knowledge and guessing cases. Recognition cases, however, were always answered fastest.

declines with growing distance in the order. Analyses considering all experimental conditions thus led to many empty cells (or low sample sizes) for larger distances. Therefore, we initially collapsed the data across types of pairs and phases, and could thus keep all participants in the sample. We then ran two-way ANOVAs with Group (larger? vs. smaller?) as the between-subjects factor and Distance (1–10) as the within-subjects factor.

Decision times decreased significantly from Distance 1 ($M = 1,451$ ms, $SD = 587$ ms) to Distance 10 ($M = 1,002$ ms, $SD = 318$ ms), $F(9, 702) = 31.604$, $p < .001$, $\eta_p^2 = .29$, thus showing a distance effect (Fig. 3). As reported above, Group 1 had shorter decision times ($M = 1,115$ ms, $SD = 420$ ms) than did Group 2 ($M = 1,260$ ms, $SD = 441$ ms), $F(1, 78) = 3.961$, $p = .05$, $\eta_p^2 = .05$. There was no interaction between group and distance, $F < 1$; that is, the distance effect was similarly large for both question formats.

Similarly, deviation rates (defined as the proportions of choices deviating from the choice implied by one’s subjective rank order) declined from Distance 1 ($M = 36.4$ %, $SD = 11.6$ %) to Distance 10 ($M = 3.1$ %, $SD = 14.5$ %), $F(9, 702) = 136.726$, $p < .0001$, $\eta_p^2 = .64$ (Table 1). Groups 1 and 2 (with $Ms = 12.6$ % and 14.7 %, respectively) did not differ significantly, and there was no interaction, both $Fs < 1$.

In addition, we analyzed inconsistencies in choices between Phases 1 and 2, computing the probability of providing different judgments for the same pair. These inconsistency rates, too, declined significantly from Distance 1 ($M = 8.8$ %, $SD = 0.4$ %) to Distance 10 ($M = 0.9$ %, $SD = 0.5$ %), $F(9, 702) = 55.719$, $p < .001$, $\eta_p^2 = .46$ (Table 1). The two groups (with $Ms = 3.9$ % and 4.4 %, respectively) did not differ in their proportions of inconsistencies, $F(1, 78) = 1.016$, $p = .32$, and there was no interaction, $F < 1$.

Next, we split the decision time data (shown in Fig. 3) according to the type of pair, excluding cells to which fewer than ten participants contributed data. The resulting mean

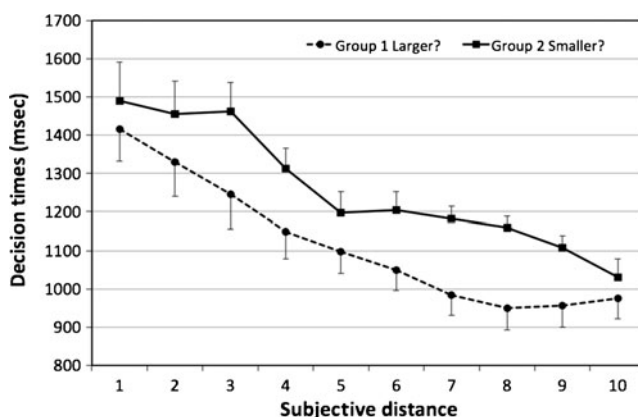


Fig 3 Mean decision times (in milliseconds) for the two experimental groups as a function of the distance in subjective rank orders. Error bars represent one standard error

Table 1 Deviation rates (according to participants' subjective rank orders) and inconsistency rates (between Phases 1 and 2) as a function of rank-order distance (1–10) and type of pair (recognition, knowledge, or guessing)

	Rank-Order Distance									
	1	2	3	4	5	6	7	8	9	10
Deviation	.364	.278	.198	.141	.105	.096	.073	.050	.025	.031
Recognition	.285	.195	.126	.075	.043	.039	.024	.036	.020	.027
Knowledge	.310	.183	.143	.106	.054	.094				
Guessing	.408	.341	.254	.200	.177	.179	.176	.094	.048	
Inconsistency	.088	.085	.066	.048	.029	.034	.023	.021	.009	.009
Recognition	.069	.059	.044	.033	.017	.018	.010	.015	.006	.007
Knowledge	.081	.060	.065	.043	.009	.047				
Guessing	.097	.103	.080	.061	.043	.058	.049	.039	.024	

Results are shown only for cells to which at least ten participants contributed data

decision times showed distance effects for all three types of pairs (see Fig. 4). However, statistical analyses were feasible only for recognition cases, where a sufficient number of participants (viz. 72) entered the analysis. In a two-way ANOVA with Group and Distance as factors, the distance effect for recognition cases was significant, $F(9, 630) = 14.729$, $p < .001$, $\eta_p^2 = .17$. As before, the two groups also differed significantly, $F(1, 70) = 10.058$, $p = .002$, $\eta_p^2 = .13$. As can be seen from Fig. 4, the other two distance effects for knowledge and guessing cases were highly similar. Simple linear regression lines for the mean decision times found slopes of -44 , -69 , and -46 ms for recognition, knowledge, and guessing cases, respectively.

To test whether and how recognition times (fluency) relate to the distance effect in decision times, we ran additional analyses including only cells with data from at least ten participants. Unfortunately, tests of significance were not feasible, because the number of participants who contributed to all included cells was too low. Thus, we report only descriptive results and 95 % confidence intervals (95 % CIs).

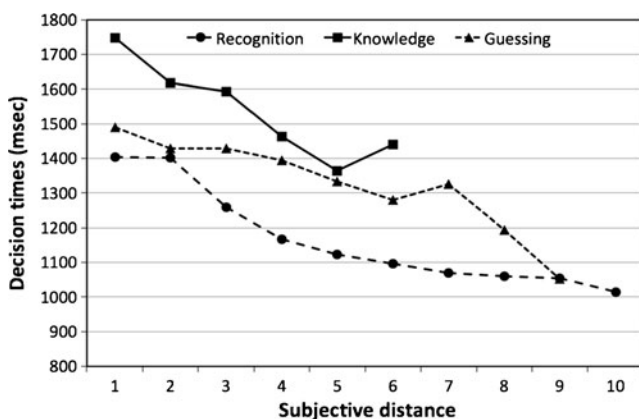


Fig 4 Mean decision times (in milliseconds) for recognition, knowledge, and guessing cases as a function of the distance in participants' subjective rank orders

For knowledge cases, the mean differences in retrieval times of the two recognized cities in a pair increased with increasing subjective distance of these cities, from Distance 1 ($M = 321$ ms, $SD = 304$ ms, 95 % CI [245, 397]) to Distance 6 ($M = 621$ ms, $SD = 409$ ms, 95 % CI [399, 843]). These increasing fluency differences might in turn have led to more frequent use of the fast FH, and thus to the observed distance effect in decision times. As such, the distance effect for two known cities may be compatible with the toolbox approach.

For recognition cases, we found that the mean fluency decreased with the recognized city's subjective rank position, from Rank 1 ($M = 1,104$ ms, $SD = 534$ ms, 95 % CI [981, 1,227]) to Rank 6 ($M = 1,733$ ms, $SD = 750$ ms, 95 % CI [1,340, 2,126]). In other words, subjectively smaller cities were recognized more slowly than larger ones (Hertwig et al., 2008; Herzog & Hertwig, in press; Marewski & Schooler, 2011). However, when analyzed as a function of the distance in subjective rank orders, pairs of different distances showed nearly the same mean fluency for the recognized city, from Distance 1 ($M = 1,136$ ms, $SD = 554$ ms, 95 % CI [1,012, 1,259]) to Distance 10 ($M = 1,083$ ms, $SD = 611$ ms, 95 % CI [943, 1,223]). Thus, fluency cannot account for the observed distance effect in recognition cases, and thus, the effect is incompatible with the toolbox view.

Finally, we also split the deviation and inconsistency rates reported above according to type of pair. We found for both measures that the observed distance effect persisted for each type of pair (see Table 1).

In sum, we found that decision times, deviation rates (in relation to subjective orders), and inconsistencies (in repeated inferences for the same pair of cities) decreased with increasing rank difference of the cities in a pair. These effects were quite large. Most importantly, and replicating our previous findings reported above, the distance effect was not only observed for knowledge cases (with both cities known, as in the Brown & Tan, 2011, study), but also for recognition cases (with only one city known), and even for

guessing cases (with neither city known). Assumptions derived from the toolbox heuristics (RH, FH, and TTB) could explain the distance effect in our data only for knowledge cases, but not for either recognition or guessing cases.

General discussion

In this article, we extended and tested Brown and Tan's (2011) assumption that decision makers base their probabilistic inferences in paired comparisons on subjective linear orders of the tested objects. This contrasts to the toolbox approach (Gigerenzer et al., 1999), which assumes a repertoire of specialized heuristics for different types of comparisons.

We reanalyzed a previously published data set (Hilbig & Pohl, 2009, Exp. 3) and presented a new experiment to test the proposed role of linear orders: (1) The central result is the distance effect in decision times that was found in the reanalyzed data set, as well as in our new experiment: The larger the subjective distance between two cities, the faster the decision. This distance effect, moreover, was present for all three types of pairs (recognition, knowledge, and guessing). (2) The distance effect was further corroborated by alternative dependent measures, namely deviation rates and inconsistencies: Both declined with increasing subjective distance, and again for all types of pairs. In sum, these observations are well aligned with the predictions from a linear-order perspective. (3) As expected from the semantic congruity effect (Banks et al., 1976), the question format (larger? vs. smaller?) had a main effect on decision times, with shorter decision times for the "larger?" question (see also Hilbig, Scholl, & Pohl, 2010; McCloy et al., 2010). Nevertheless, the distance effects were comparable in magnitude for both types of questions (as Banks et al., 1976, had also found).

The toolbox approach, with its assumption of multiple strategies that fit to different situations, can easily explain the distance effect in knowledge cases, either through the fluency heuristic (FH; Hertwig et al., 2008; Schooler & Hertwig, 2005) or the take-the-best heuristic (TTB; Gigerenzer & Goldstein, 1996). Brown and Tan (2011) interpreted their distance effect as incompatible with the TTB, because the same highly valid cue could have been used in pairs of different distances.

A potential explanation of the distance effect in recognition cases could be derived from a recent suggestion that decision makers use fluency as a cue not only in knowledge cases, but also in recognition cases (Marewski et al., 2010). However, in our experiment, the observed fluencies for the recognized city in pairs of different distances were highly similar and did not increase with distance. Thus, weighting the recognition cue by fluency cannot account for the observed distance effect in recognition cases. Consequently, both the original RH and recent extensions fail to account for this distance effect.

Finally, in guessing cases, with two unknown cities, participants should have nothing left but to guess. However, the

data show that these cases were affected by subjective distance, too, suggesting that participants possess (or infer) some knowledge for subjectively ordering the "unknown" cities. That these subjective orders were not totally random, but at least partially built on valid knowledge, was shown by the positive rank-order correlations with the true order. To explain the distance effect from a toolbox viewpoint, one would need to argue that guessing cases were handled just like knowledge cases (see above), albeit with less knowledge.

The most problematic finding for the linear-order perspective, though it is well aligned with the toolbox assumptions, was that recognition cases were answered faster than either knowledge or guessing cases (Pachur & Hertwig, 2006). However, research has shown that the overall main effect of type of pair may actually be misleading, as it is not generally true that recognition cases yield faster decisions than knowledge cases (Hilbig & Pohl, 2009). Nonetheless, if linear-order representations include and treat all objects equally, irrespective of their recognition status, decision times should only depend on distance, not on type of pair. As such, our findings cannot be fully accounted for by the linear-order perspective.

Overall, neither the toolbox approach nor the linear-order perspective is fully compatible with all observations. A possible remedy to these theoretically inconsistent findings would be to consider an alternative approach, namely the difference in evidence (cf. Hilbig & Pohl, 2009; Newell, 2005). To derive predictions for different types of pairs, assume, merely for demonstration, that the amount of evidence is mapped on a scale from 0 to 4: Unrecognized city names, for example, will elicit no (0) or only very limited evidence (1) regarding their size. Thus, on average, the difference in evidence between two unrecognized cities would typically be small, making an inference rather difficult. Recognized city names, on the other hand, may elicit more diverse amounts of evidence, ranging from mere recognition without any further knowledge (2), over one or more probabilistic cues (3), up to full criterion knowledge (4; cf. Hilbig et al., 2009). Correspondingly, pairs of cities that differ more strongly in terms of the evidence that speaks to their criterion value will be answered more quickly than pairs with less differing evidence (see Hilbig & Pohl, 2009). Following this approach, the mean difference in evidence will typically be larger for recognition cases (with exactly one object recognized) than for guessing or knowledge cases (no matter how the evidence scale is defined), so that decision times should be fastest for recognition cases, thus explaining the main effect of type of pair. Further assuming that knowledge is typically correlated with the criterion (i.e., that people know more about larger cities than about smaller ones), it is clear that the mean difference in evidence increases with increasing distance of the two objects on the underlying continuum, thus also accounting for the distance effect.

As such, the difference in evidence (or, conversely, the degree of conflict) between two options determines decision times—as predicted by evidence accumulation (e.g., Busemeyer & Townsend, 1993; Newell, Collins, & Lee, 2007) and network models (e.g., Glöckner & Betsch, 2008a, b) of decision making: The larger the evidential difference, the faster the decision. This explanation should be tested by assessing (or manipulating) the subjective evidence for the objects in a set, and thereby predicting decision times. This method was not incorporated here, but appears to be a promising route for future research (see Pachur et al., 2008).

Author note This research was supported by a grant from the University of Mannheim and by Grant No. Er 224/2-1 from the German Science Foundation (DFG). We thank Tina Tanz for her help in programming and running the reported experiment, and two anonymous reviewers for their constructive comments on an earlier version of this article.

References

- Ayal, S., & Hochman, G. (2009). Ignorance or integration: The cognitive processes underlying choice behavior. *Journal of Behavioral Decision Making*, 22, 455–474.
- Banks, W. P. (1977). Encoding and processing of symbolic information in comparative judgments. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 11, pp. 101–160). New York: Academic Press.
- Banks, W. P., Fuji, M., & Kayra-Stuart, F. (1976). Semantic congruity effects in comparative judgments of magnitudes of digits. *Journal of Experimental Psychology: Human Perception and Performance*, 2, 435–447.
- Bröder, A. (2000). Assessing the empirical validity of the “Take-the-best” heuristic as a model of human probabilistic inference. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26, 1332–1346.
- Bröder, A., & Gaissmaier, W. (2007). Sequential processing of cues in memory-based multiattribute decisions. *Psychonomic Bulletin & Review*, 14, 895–900.
- Brown, N. R., & Tan, S. (2011). Magnitude comparison revisited: An alternative approach to binary choice under uncertainty. *Psychonomic Bulletin & Review*, 18, 392–398.
- Busemeyer, J. R., & Townsend, J. T. (1993). Decision field theory: A dynamic-cognitive approach to decision making in an uncertain environment. *Psychological Review*, 100, 432–459.
- Erdfelder, E., Küpper-Tetzl, C. E., & Mattern, S. D. (2011). Threshold models of recognition and the recognition heuristic. *Judgment and Decision Making*, 6, 7–22.
- Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review*, 103, 650–669.
- Gigerenzer, G., Hoffrage, U., & Kleinbölting, H. (1991). Probabilistic mental models: A Brunswikian theory of confidence. *Psychological Review*, 98, 506–528.
- Gigerenzer, G., Todd, P. M., & the ABC Research Group. (1999). *Simple heuristics that make us smart*. New York: Oxford University Press.
- Glöckner, A., & Betsch, T. (2008a). Modeling option and strategy choices with connectionist networks: Towards an integrative model of automatic and deliberate decision making. *Judgment and Decision Making*, 3, 215–228.
- Glöckner, A., & Betsch, T. (2008b). Multiple-reason decision making based on automatic processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34, 1055–1075.
- Goldstein, D. G., & Gigerenzer, G. (1999). The recognition heuristic: How ignorance makes us smart. In G. Gigerenzer, P. M. Todd, & the ABC Research Group (Eds.), *Simple heuristics that make us smart* (pp. 37–58). New York: Oxford University Press.
- Goldstein, D. G., & Gigerenzer, G. (2002). Models of ecological rationality: The recognition heuristic. *Psychological Review*, 109, 75–90.
- Hertwig, R., Herzog, S. M., Schooler, L. J., & Reimer, T. (2008). Fluency heuristic: A model of how the mind exploits a by-product of information retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34, 1191–1206.
- Herzog, S. M., & Hertwig, R. (in press). The ecological validity of fluency. In C. Unkelbach & R. Greifeneder (Eds.), *The experience of thinking: How the fluency of mental processes influences cognition and behaviour*. Hove: Psychology Press.
- Hilbig, B. E. (2010). Reconsidering “evidence” for fast-and-frugal heuristics. *Psychonomic Bulletin & Review*, 17, 923–930.
- Hilbig, B. E., Erdfelder, E., & Pohl, R. F. (2010a). One-reason decision-making unveiled: A measurement model of the recognition heuristic. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36, 123–134.
- Hilbig, B. E., Erdfelder, E., & Pohl, R. F. (2011). Fluent, fast, and frugal? A formal model evaluation of the interplay between memory, fluency, and comparative judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37, 827–839.
- Hilbig, B. E., & Pohl, R. F. (2009). Ignorance- versus evidence-based decision making: A decision time analysis of the recognition heuristic. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35, 1296–1305.
- Hilbig, B. E., Pohl, R. F., & Bröder, A. (2009). Criterion knowledge: A moderator of using the recognition heuristic? *Journal of Behavioral Decision Making*, 22, 510–522.
- Hilbig, B. E., Scholl, S. G., & Pohl, R. F. (2010b). Think or blink: Is the recognition heuristic an ‘intuitive’ strategy? *Judgment and Decision Making*, 5, 300–309.
- Marewski, J. N., Gaissmaier, W., Schooler, L. J., Goldstein, D. G., & Gigerenzer, G. (2010). From recognition to decisions: Extending and testing recognition-based models for multialternative inference. *Psychonomic Bulletin & Review*, 17, 287–309.
- Marewski, J. N., Pohl, R. F., & Vitouch, O. (2011). Recognition-based judgments and decisions: What we have learned (so far). *Judgment and Decision Making*, 6, 359–380.
- Marewski, J. N., & Schooler, L. J. (2011). Cognitive niches: An ecological model of strategy selection. *Psychological Review*, 118, 393–437.
- McCloy, R., Beaman, C. P., Frosch, C. A., & Goddard, K. (2010). Fast and frugal framing effects? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36, 1043–1052.
- Moyer, R. S., & Bayer, R. H. (1976). Mental comparison and the symbolic distance effect. *Cognitive Psychology*, 8, 228–246.
- Moyer, R. S., & Dumais, S. T. (1978). Mental comparison. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 12, pp. 117–155). New York: Academic Press.
- Newell, B. R. (2005). Re-visions of rationality? *Trends in Cognitive Sciences*, 9, 11–15.
- Newell, B. R., Collins, P., & Lee, M. D. (2007). Adjusting the spanner: Testing an evidence accumulation model of decision making. In D. McNamara & G. Trafton (Eds.), *Proceedings of the 29th Annual Conference of the Cognitive Science Society* (pp. 533–538). Austin, TX: Cognitive Science Society.
- Pachur, T., Bröder, A., & Marewski, J. (2008). The recognition heuristic in memory-based inference: Is recognition a non-compensatory cue? *Journal of Behavioral Decision Making*, 21, 183–210.

- Pachur, T., & Hertwig, R. (2006). On the psychology of the recognition heuristic: Retrieval primacy as a key determinant of its use. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*, 983–1002.
- Parkman, J. M. (1971). Temporal aspects of digit and letter inequality judgments. *Journal of Experimental Psychology*, *91*, 191–205.
- Pohl, R. F. (2011). Recognition information in inferential decision making: An overview of the debate. *Judgment and Decision Making*, *6*, 423–438.
- Schooler, L. J., & Hertwig, R. (2005). How forgetting aids heuristic inference. *Psychological Review*, *112*, 610–628.
- Smithson, M. (2010). When less is more in the recognition heuristic. *Judgment and Decision Making*, *5*, 230–244.