Redefining the Congruity Effect in Comparative Judgments: A Review of the Theories and a Further Test

JERWEN JOU, MARIA SEVILLA MATOS, MIKAYLA A. MARTINEZ,
FRANCISCO J. SIERRA, CLAIR GUZMAN, and AMANDA R. HUT
University of Texas–Rio Grande Valley

This article reviews major theories about the congruity effect (CE) in comparative judgments and reports 2 experiments further testing a newly proposed theory. The congruity effect is the phenomenon that in a comparative judgment, when the comparative in the instruction ("choose larger" or "choose smaller") matches the magnitude of the 2 items being compared ("choose larger" when the two items are large, "choose smaller" when the two items are small), the choice is faster than when they mismatch. For four decades this has been the belief of researchers. A recent study (Jou, Escamilla, Torres, Ortiz, & Salazar, 2018) has shown that the instruction comparative plays no necessary role in generating the congruity effect. Instead, the item closer to the end of a continuum of an attribute tends to be retrieved from memory faster than the item farther from an end regardless of whether there is an instruction. The assumptions of the newly proposed theory, called the serial position–based distinctiveness account, were further tested in this study. The test results were affirmative. They showed that the congruity effect comes from the serial position–based distinctiveness of the items, and thus the decades-old misunderstanding of the role of the instruction in producing the CE is corrected.

KEYWORDS: congruity effects, comparative judgments, absolute judgments, serial order memory, global distinctiveness models, local distinctiveness models

Using serial order information is a ubiquitous cognitive activity in everyday life, including during speech processing and production (Crowder, 1976; Restle & Brown, 1970). How people learn and use order information has interested researchers since Ebbinghaus (1885/1964), for more than a century (see Crowder & Greene, 2000; and Marshuetz, 2005 for a review). The two most commonly used retrieval paradigms in serial order memory research can be broadly classified as retrieval of single serial items or determination of the relative order or magnitudes between a pair of items. The former paradigm includes serial recall (e.g., Bhatarah, Ward, & Tan, 2008; Nairne, 1988) and absolute identification or judgments (e.g., Jou, 2019; Kent & Lamberts, 2005; Neath, Brown, McCormack, Chater, & Freeman, 2006). In an absolute identification task, subjects first learn an array of stimuli that systematically vary along a certain attribute dimension (e.g., of size, length, loudness, intelligence), with each stimulus associated with a magnitude label, typically a number. The identification or judgment task consists of having subjects respond to the pre-

American Journal of Psychology
Summer 2020, Vol. 133, No. 2 pp. 221–239 •

sented stimulus by producing the magnitude value associated with the presented stimulus (Neath et al., 2006) or respond "yes" or "no" to a stimulus item paired with a magnitude value to indicate whether they match or mismatch (Jou, 2019; Jou et al., 2018).

The typical method used in the pair retrieval paradigm is comparative (or relative) judgments. In a typical episodic comparative judgment task, people first learn an ordering of stimuli along a magnitude continuum of an attribute. They are then asked to choose, from a pair of items sampled from the continuum, the one possessing a greater or lesser amount of the attribute (Banks, 1977; Banks & White, 1985; Birnbaum & Jou, 1990; Foltz, Poltrock, & Potts, 1984; Henderson & Well, 1985; Jou, 2010; Marschark, 1983; Shoben, Cech, Schwanenflugel, & Sailor, 1989). There are three robust effects from the judgment: the serial position effect, the distance effect, and the congruity effect (CE). The serial position effect is the phenomenon that pairs located at more extreme positions of the continuum are compared faster and more accurately than pairs located at or near the middle of the continuum (Banks, 1977; Birnbaum & Jou, 1990; Jou, 2010; Murdock, 1960; Shoben, Cech, et al., 1989). The distance effect refers to the fact that the more the two items differ from each other on the attribute dimension, the faster and more accurate the comparison is (Moyer & Bayer, 1976; Moyer & Landauer, 1967).

The CE, which is the focus of the present article, refers to the finding that choosing the larger of a large-sized pair of items is faster and more accurate than choosing the smaller of the pair, and choosing the smaller of a small-sized pair is faster and more accurate than choosing the larger of the pair (Banks, 1977; Banks, Fujii, & Kayra-Stuart, 1976; Jou et al., 2018; Marschark & Paivio, 1979). The CE can be regarded as a more general phenomenon of the polarity correspondence effect (Proctor & Cho, 2006). It is a statistical crossover interaction between the choice instruction (choosing "greater" vs. "lesser" item) and the serial position of the items on the attribute continuum. In this interaction, the response time magnitude order of the two choices is reversed at the opposite sides of the continuum (see Figures 1 and 2 for an illustration of the CE).

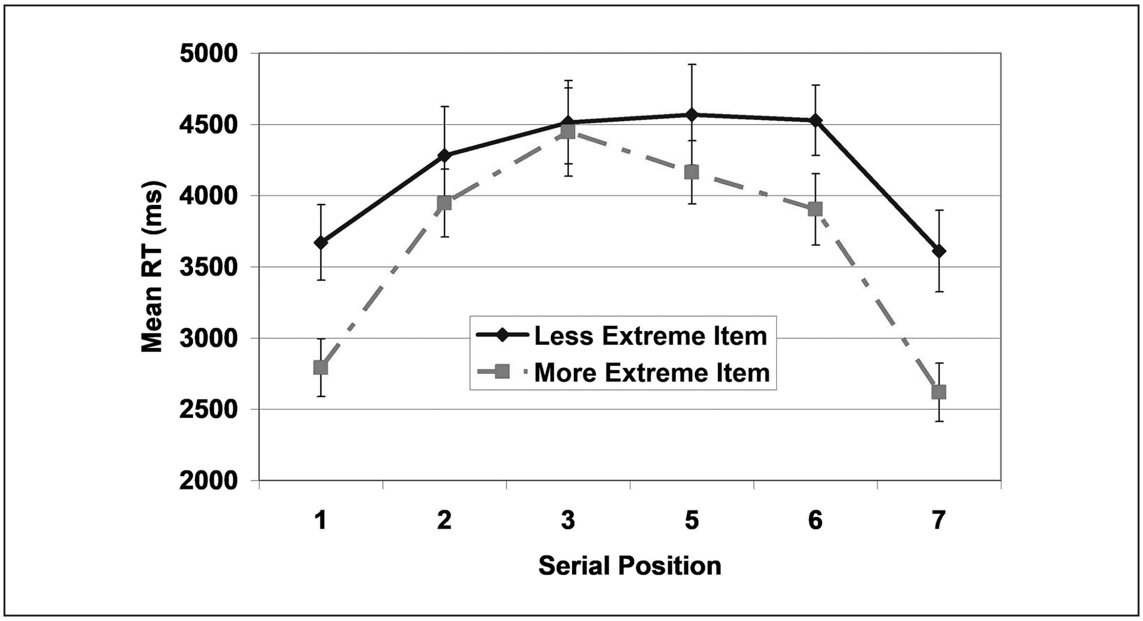

FIGURE 1. Mean response times (RTs) of the freely chosen pair members as a function of the relative positional extremity of the chosen items and serial position. The black curve connects the mean RTs of the freely chosen less extreme items, and the pink curve connects the mean RTs of the freely chosen more extreme items. The data are from pairs with 1-step interitem distance. The serial position is the lower of the two ranks. Error bars represent ±1 standard error of mean. From Jou et al. (2018), *Journal of Memory and Language,* Copyright 2018 by Elsevier. Reprinted by permission

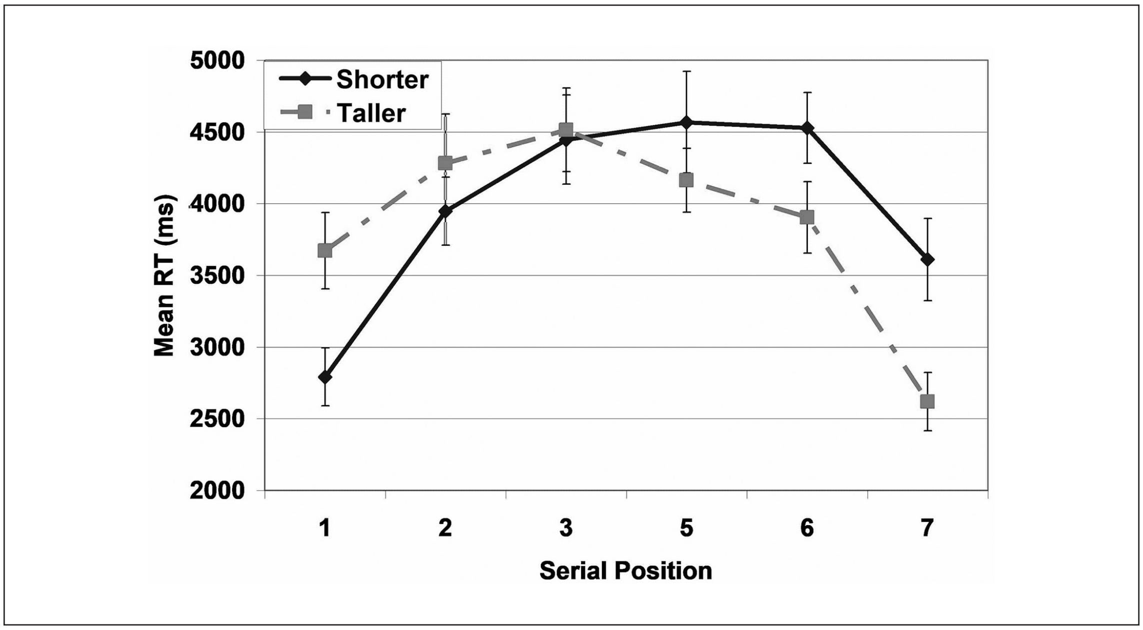

FIGURE 2. Mean response times (RTs) as a function of the freely chosen pair member's relative magnitude ("taller" vs. "shorter") and the serial position. The data points are the same as in Figure 1. The black curve connects the mean RTs of the freely chosen shorter and the pink curve connects those of the freely chosen taller people of the pairs. The serial position is the lower of the two ranks. Error bars represent ±1 standard error of mean. From Jou et al. (2018), *Journal of Memory and Language,* Copyright 2018 by Elsevier. Reprinted by permission

The CE was first reported by Shipley and colleagues (Shipley, Coffin, & Hadsell, 1945; Shipley, Norris, & Roberts, 1946). They found that subjects chose the more preferred color from a pair of highly liked colors faster than they chose the less preferred, and vice versa, for a pair of colors rated low in preference. After Shipley et al., Audley and Wallis (1964) found that the selection of the brighter of two bright lights was faster than of two dark lights, and vice versa. From the 1970s onward, congruity effects have been reported in memorial and perceptual comparative judgments on objects and animal sizes (Banks & Flora, 1977; Cech, 1995; Cech, Shoben, & Love, 1990; Leth-Steensen, Petrusic, & Shaki, 2014), magnitude of numbers (Banks et al., 1976; Duncan & McFarland, 1980; Foltz et al., 1984), and other dimensions such as loudness (Banks & Root, 1979), time, quality, temperature (Holyoak & Walker, 1976), and likeability of people (Birnbaum & Jou, 1990), among others.

The main purpose of this article is to review the major extant theories of CEs, reveal the problems these theories have in explaining many findings from comparative judgments, and argue that now may be the time to make a shift in the conceptualization of what the CE is. Jou et al. (2018) recently proposed a theory of CE, the serial position–based distinctiveness (SPBD) account, that seems to be able to provide a coherent, integrative explanation for the seemingly divergent findings from different comparative judgment studies that the traditional CE theories fail to explain. However, the SPBD account has not been compared with the alternative accounts of CE. In this article, we review six extant theories of CE and indicate why the SPBD is the preferred model. In addition, we conducted two experiments to further test some assumptions of the SPBD account of the CE.

Major Theories of CEs

We briefly describe six major theories below. They are the semantic coding (Banks, 1977), the expectancy (Marschark & Paivio, 1979), the evidence accrual (Petrusic, 1992), the reference point (Holyoak, 1978; Marks, 1972), the biased start point (Birnbaum & Jou, 1990), and the recently proposed SPBD theory (Jou et al., 2018).

SEMANTIC CODING THEORY VERSUS EXPECTANCY THEORY.

One very influential theory is the semantic coding theory (Banks, Clark, & Lucy, 1975; Banks, Fujii, & Kayra-Stuart, 1976; see Banks, 1977 for a review). The semantic coding theory holds that the comparison process uses categorical linguistic or semantic codes and that both the instruction comparative and the magnitude of the stimulus items must be translated into a binary ("large" vs. "small") linguistic code before the comparison can be made. Moreover, the two codes have to be matched before the selection of the target can be executed. Thus, if the two items in a pair are both large, the larger one will be coded as "large +" and the smaller one as "large." If both items in a pair are small, the smaller one will be coded as "small +" and the larger one as "small." However, if the instruction is to choose the larger, but the two items are small (coded as "small +/ small"), the latter code must be transformed into its equivalent "large" code of "large/large +" to match the comparative code. A fast response in the code-matched condition versus a required recoding in the code mismatched condition generates the CE. Because of the great influence of this theory, the CE is widely called the semantic congruity effect[1] in the comparative judgment literature.

The second major theory is the expectancy theory (Marschark, 1983; Marschark & Paivio, 1979, 1981), which is a serious competitor for the semantic coding theory. This theory argues that the instruction comparative creates an expectation for the magnitude range of the stimulus items. When the comparative is "choose larger" and the stimulus items are actually large, the expectation is met, resulting in a faster choice. However, when the expectation created is "large" and the stimuli are small, the expectation is contradicted, thus slowing down the response (Banks, White, Sturgill, & Mermelstein, 1983; Marschark, 1983; Marschark & Paivio, 1979). The same mechanism applies to "choose smaller" instruction.

In the 1970s and 1980s, the semantic coding and expectancy theories were contesting with each other and as a result produced many published articles (Banks & White, 1985; Banks et al., 1983; Cech, 1989; Holyoak & Mah, 1981; Howard, 1983; Marschark, 1983; Marschark & Paivio, 1979, 1981; Shoben, Sailor, & Wang, 1989). Proponents of the semantic coding and expectancy models agreed that the two models could be tested by presenting the stimulus pair *before* the instruction. If the expectancy created by the instruction causes the CE, then when the stimuli are presented first, there will be no expectation about the magnitude range of the stimuli, and therefore no CE should be observed. Indeed, Marschark and Paivio (1979) did not observe a CE when they presented the stimuli first. However, the proponents of the semantic coding model and other skeptics of the expectancy theory showed that the CE persisted in some contexts when the stimulus pair was presented before the instruction (Banks & Flora, 1977; Banks et al., 1983; Banks & White, 1985; Holyoak & Mah, 1981; Shoben, Sailor, et al., 1989).

Marschark (1983) defended the expectation theory by suggesting that the expectation could go either way, from the comparative to the stimuli or from the stimuli to the comparative, and that this would be especially the case when the attribute was a salient one for the objects, such as size for a pair of elephant versus whale, or ferocity for a pair of lion versus wolf. In this debate, each side criticized the methods of the other side's experiments for why the results did not turn out as their theories predicted. Overall, however, the semantic coding theory seemed to be favored by more researchers (Cech, 1989; Holyoak & Mah, 1981; Shoben, Sailor, et al., 1989).

EVIDENCE ACCRUAL THEORY.

This model characterizes the decision process in a comparative judgment as an evidence accruing process, or repeated sampling of information from the stimuli until the information in favor of one item has met a criterion, at which point that item is chosen. This theory attributes the CE to different judgment criteria used under different comparison conditions, with higher (more conservative) criteria adopted in more difficult conditions (instruction–stimulus incongruence or other contextual circumstances making the discrimination more difficult) and lower (less conservative) criteria in easier comparison conditions (Petrusic, 1992; Petrusic, Shaki, & Leth-Steensen, 2008; Shaki, Leth-Steensen, & Petrusic, 2006). According to the theory, when a higher criterion is used, more evidence must be accrued before a decision, which adds to the response time (RT) but makes the response more accurate. When a lower criterion is used, less evidence is needed for a decision, which speeds up the decision making but increases the error rate. When the comparative in the instruction and the stimulus magnitude are not congruent, the comparison is more difficult than when they are congruent.

This model predicts that the CE will be larger under the more difficult than under the easier comparison condition. Thus, Shaki et al. (2006) found that the CE was larger in a "larger" and "smaller" mixed instruction presentation condition than in a blocked presentation condition because of the greater judgment difficulty created by the mixed instruction presentation. In another study, Petrusic et al. (2008) used consonant–vowel–consonant arbitrary nonsense syllables to represent "choose larger" or "choose smaller" in order to increase the memory demands of the task. They predicted a greater CE under that coded instruction condition than under the conventional instruction condition, which was what they found.

REFERENCE POINT THEORIES.

These theories are analog continuum models based on the assumptions that the comparative in the instruction establishes a reference point at either the low or high end of the continuum, with the items near either end of the continuum being easier to discriminate than items far from the reference points. There are two specific versions of this theory. One version (Holyoak, 1978; Holyoak & Mah, 1982; Jamieson & Petrusic, 1975) posits that the RT for choosing a target derives from comparing against a criterion ratio of 1.0 the ratio of the two distances between each of the items and the reference point. The more the ratio is removed from 1.0, the faster the response. When the two items are judged from the nearer end of the continuum, the ratio of the two distances is more different from 1.0 than when they are judged from the farther end, hence speeding up the decision and vice versa.

Another version (Marks, 1972) of the theory posits that the judgment RT is a function of the magnitude difference between the two stimuli and the stimulus magnitude discriminal dispersion (variance). The discriminal dispersion decreases as the stimuli move from the middle of the continuum toward an end of the continuum. Assume that the two comparison items are located nearer the low end of the continuum but that the instruction is "Indicate which is larger." In this case, the reference point will be set at the high end of the continuum. Under that circumstance, the comparison will be less accurate and slower because of the larger stimulus variance. In contrast, if the instruction is "Indicate which is smaller," the reference point will be set at the low end, thus reducing the stimulus variance and making the choice faster and more accurate.

BIASED START POINT THEORY.

This theory (Birnbaum & Jou, 1990) derives from Link's (1978, 1991) random walk model, in which subjects start the walk or sampling information for each item somewhere between the two decision boundaries, one for each of the two choices. According to this theory, when one of the two items is near an end of the continuum, the random walk starts from a biased point (nearer one bound), and if that bias is consistent with the polarity of the comparative, the random walk time is shorter than if the comparative specifies the opposite polarity. This concept of a biased start point in the walk also explains the end effects of the serial order memory. However, it further postulates that when the two items are near the middle of the continuum, the start point of the random walk will be unbiased and the RT will be slow and little affected by the polarity of the comparative.

SPBD ACCOUNT.

The SPBD account proposes that the pair mate located nearer an end of the series and therefore having a greater positional distinctiveness is retrieved from memory faster than the one farther from an end, regardless of what the instruction specifies or whether there is even an instruction. Jou et al. (2018) have recently demonstrated that the instruction plays no necessary role in the generation of the CE. Instead, the CE originates from the relative item distinctiveness derived from the serial position of an item in the serial order memory (Murdock, 1960). In the comparative judgment task of their first experiment, they made a manipulation in which no traditional instruction was given that directs subjects to choose the larger or smaller of the two items. Instead, they told subjects that they might choose whichever item they preferred if they thought the choice would make their response faster and easier. After the choice was made, subjects needed to report whether the chosen item was the taller or shorter person. They found that two thirds of the time, subjects preferred choosing the item closer to an end of the series (the positionally more extreme item) and that they chose the positionally more extreme item faster and reported its relative magnitude more accurately than they did the positionally less extreme item. They called this finding the *item extremity effect*. In the comparative judgment task of their second experiment, they put a question mark ("?") under a randomly chosen name from the pair of names for subjects to indicate whether the queried name was the taller or shorter person. Again, the response was faster and more accurate when the queried item was the positionally more extreme of the two items. In the data of both experiments, when the chosen "taller" mean RT points were connected by a line against the serial positions of the pairs and the "shorter" points were connected by another line, the traditional crossover interaction emerged (see Figures 1 and 2).

According to Jou et al. (2018), the CE is actually a serial position effect within the pair of items. In addition, the serial position effect cannot be changed or influenced by the comparative in the instruction. Thus, the comparative in the instruction is neither necessary nor causal in generation of the CE. It is a coincidental factor. The instruction's only function is to distribute half of responses to the fast and the other half to the slow "track." Metaphorically, what determines the response speed is the track, and the instruction plays no part in laying down the tracks.

Jou et al. (2018) directly adopted Murdock's (1960) stimulus serial position distinctiveness theory as the basis of their CE account. According to Murdock's theory, the two end items of a sequentially related array are the most distinctive because, first, they have neighbors only on one side, whereas all other items have neighbors on both sides, and second, the middle item's magnitude is the average of all items and hence the most similar to all other items, and therefore the least distinctive of all. An item's distinctiveness increases progressively as its position moves from the middle point toward an end of the series. Murdock (1960) quantified the distinctiveness of each serial position by computing the total distance of a position from the rest of the positions in a series. Assume the series contains eight items ranked on an 8-point ordinal scale, and the distance between one item and its immediate neighbor is 1. In Table 1, the entry represents the distance in ordinal steps between two serial items, and the rightmost column indicates the sum of all the distances (i.e., the total distinctiveness) between a particular item and all other items.

In Murdock's original quantification of distinctiveness, he used log distance. Here we used the raw distance measure because we are not fitting the model quantitatively to our data, and in addition the

TABLE 1. Interitem Distance Measured in the Number-of-Steps Difference Between Two Serial Order Items (Murdock, 1960).

Item serial position	1	2	3	4	5	6	7	8	Total distinctiveness
1		1	2	3	4	5	6	7	28
2	1		1	2	3	4	5	6	22
3	2	1		1	2	3	4	5	18
4	3	2	1		1	2	3	4	16
5	4	3	2	1		1	2	3	16
6	5	4	3	2	1		1	2	18
7	6	5	4	3	2	1		1	22
8	7	6	5	4	3	2	1		28

Note. Distinctiveness of a serial order item is defined as the sum of distances between that item and all the other items according to Murdock's (1960) theory. A number in the table is the step distance between an item and another. From Jou et al. (2018), *Journal of Memory and Language*. Copyright 2018 by Elsevier. Reprinted by permission.

two measures yield the same ordinal rankings in distinctiveness, which only are relevant to the present purpose.

PROCESSES GENERATING THE CE ACCORDING TO THE SPBD ACCOUNT.

Simply put, the relative extremity relation between the two items determines the CE. It is suggested that when subjects see the pair of stimulus items, the more distinctive one is automatically activated and retrieved before the less extreme one. If the designated choice in the instruction coincides with this automatically first-retrieved item, the response is made fast. If the designated choice happens to be the less extreme of the two items, the more activated item has to be inhibited, and a reselection of the less extreme item has to be made. This process takes extra time. Note that if the instruction generates a comparison code (Banks, 1977), an expectation (Marschark & Paivio, 1979), or a reference point (Holyoak, 1978) as the first step in generating a CE, then when it is removed, there should be no CE. But the CE arises as usual under those conditions.

The distinctiveness strictly defined by serial positions is referred to as global distinctiveness, and the model based on that concept is known as the global distinctiveness model (Murdock, 1960; Neath & Brown, 2006). One corollary from the formulation of the global distinctiveness theory is that the middle item can never outperform its neighbors in memory because it is the least distinctive of all items. One challenge to this idea and hence to the SPBD account is from a rival to the global distinctiveness model known as the local distinctiveness model (Neath & Brown, 2006; Neath et al., 2006). According to Neath and colleagues, the crucial assumption that the midseries item can never outperform its neighbors is a shortcoming of the global distinctiveness model. They argued that the midseries item does not always have to be the least distinctive item. For example, they demonstrated that the midseries item can be made more distinctive and memorable than its neighbors by isolating it from its neighbors, thus creating a von Restorff effect (von Restorff, 1933, cited in Hunt, 1995) for the midseries item. In one study using a tone absolute identification task, Neath et al. (2006) isolated the midseries tone by increasing its frequency difference from its immediate neighbors beyond the regular interitem tone interval, and as a result they dramatically increased the absolute identification performance for the midseries tone, changing the otherwise *U*-shaped serial position function into a *W*-shaped function with a prominent middle peak. They pointed out that the global distinctiveness theory could not account for the effect of isolating the midseries item, or the local distinctiveness effect (Neath & Brown, 2006; Neath et al., 2006).

When each item in a serial array is learned in a homogeneous way, the global and local distinctiveness models make the same prediction for the memory performance on the order. The challenge for both the global distinctiveness model and the SPBD account is whether the created local distinctiveness will change the regular CE pattern in the comparative judgment. In other words, will the relative RT magnitudes of choosing one or the other item in a pair still follow the item extremity principle when the midseries item is made especially distinctive?

To address this question, Jou et al. (2018) created a local distinctiveness in the order by strengthening the association of name number 4, a midseries item in an eight-item list and associated with one of the two longest RTs, with its serial position information through repeated presentations of the name–serial position pair. In addition, they tested that series both with an absolute judgment (an item–serial position association recognition task) and with a comparative judgment task. The results showed that although in the absolute judgment task, the strengthened middle item produced a deep dip in the middle of the otherwise inverted *U*-shaped RT function (consistent with Neath et al.'s 2006 finding), it produced no significant effect in the comparative judgment task either on the serial position curve or on the CE. Therefore, they concluded that the CE seems to be determined by the serial position–based global distinctiveness of the items and is not affected by the local distinctiveness.

The reason Jou et al. (2018) gave for why the local distinctiveness had differential effects on the comparative and the absolute judgments is that an isolation-based distinctiveness makes one single item especially distinctive but does not enhance serial position information of the other items. In an absolute judgment, only the magnitude information of an individual item is probed. In a comparative judgment, on the other hand, one must have the magnitude information of both (or all) items to make the correct choice. That is why the enhanced memory strength of one single item is not sufficient to produce an effect. Therefore, only serial position–based, global distinctiveness determines the result pattern of a comparative judgment.

In a recent article, Jou (2019) suggested that the two types of judgment tasks tap into two different types of information: the item-specific and the relational information (Greene, Thapar, & Westerman, 1998; Hunt & Einstein, 1981; Hunt & McDaniel, 1993; Mulligan, 1999, 2000, 2001). Specifically, an absolute judgment is sensitive to the item-specific information that is enhanced in the isolation manipulation, whereas a comparative judgment relies heavily on the between-item, relative magnitude information that connects one item with another. Hence, in the comparative judgment, the crucial information is where an item stands in the ordering in relation to all the other items. The individual item's distinctive characteristics become irrelevant information.

Assessments of the Theories

The semantic coding theory has been very influential since the 1970s. Because it assumes that the origin of CE involves using linguistic codes, the CE has been called the semantic congruity effect by most researchers in this area. Its competitor, expectancy theory, was shown to be unsubstantiated in most investigations (Cech, 1989; Holyoak & Mah, 1981; Howard, 1983; Shoben et al., 1989). The major problems with expectancy theory are that the CE persisted (a) when the stimulus pair was presented before the instruction (Banks & Flora, 1977; Banks et al., 1983; Holyoak & Mah, 1981; Shoben et al., 1989), (b) when the stimulus pair was presented simultaneously with the instruction (Jou et al., 2018, Experiment 3 in supplementary materials, and Experiment 4; but see also Marschark & Paivio, 1979), (c) when the instruction was presented in a blocked mode (Banks & White, 1985, Experiment 2; Cech, 1995, Experiment 1; Jou et al., 2018, Experiment 3; Shaki et al., 2006), and (d) when subjects were explicitly informed of the stimulus magnitude range and supposed to be prepared for the range of upcoming stimuli in each trial (Howard, 1983). In response to these critiques, Marschark (1983) extended the expectancy to one that can go both in the instruction-to-stimuli and the stimuli-to-instruction directions. However, by arguing that the expectancy could go both ways, he weakened the theory by making it almost indistinguishable from the semantic coding theory because the latter theory holds that the CE obtains in both presentation modes. Since then, the semantic coding theory has not been seriously challenged (but see Petrusic & Baranski, 1989). In fact, several studies expanded this theory to cover new related findings (Cech, 1989, 1995; Cech & Shoben, 1985).

The aforementioned theories that ascribe an important role to the comparative in causing the CE have difficulty explaining why the CE persists in a blocked instruction presentation mode (e.g., Audley & Wallis, 1964; Banks & White, 1985, Experiment 2; Jou et al., 2018, Experiments 3 and 4; Shaki et al., 2006). There is no obvious reason for subjects to constantly create a "large" code from the instruction, for example, or to expect a large stimulus range, or to set the reference point invariably at the large end of the continuum throughout all the trials. Both Banks and Flora (1977) and Marschark and Paivio (1979) mentioned that CE in a blocked instruction presentation condition was a problem for their theories and attributed it to some special strategies possibly used by subjects.

As noted, the evidence accrual, as well as the reference point, theory holds that under more difficult judgment conditions, the magnitude of CE should increase. Shaki et al. (2006) did obtain a larger CE in a mixed instruction presentation (which was more difficult) than in a blocked presentation condition (an easier condition), and Petrusic et al. (2008) obtained a larger CE in a consonant–vowel–consonant coded instruction (using nonsense syllables to represent choosing "larger" or "smaller" comparatives, a more difficult condition) than in a traditional instruction condition. However, the higher error rate in the more difficult than in the less difficult condition contradicted their prediction of a lower error rate to be associated with a higher judgment criterion adopted in the more difficult condition (Shaki et al., 2006). The larger CE in the mixed instruction was also susceptible to an alternative interpretation as a task-switching cost incurred in the more difficult mixed instruction condition (Jou, 2014; Los, 1996; Rogers

& Monsell, 1995), although the authors rejected that interpretation.

The biased start point account (Birnbaum & Jou, 1990) states that a biased start point in the random walk shortens the distance of the walk and therefore the RT. The major claim of this theory that the item in the pair that is closer to the high or low end of the continuum creates a bias toward a quick selection of that item is consistent with the main idea of the SPBD account, that there is a tendency to choose the more extreme of the two items. The theory does not specify what cognitive process the instruction comparative generates to produce the CE, as the other theories do. However, it does indicate that if the pair contains two items from the middle of the continuum, the random walk start point is unbiased; therefore, the specification of the comparative has little effect on the selection time of either item, thus generating a small or no CE. This assumption is corroborated by empirical data in Figures 1 and 2. The biased start point theory is akin to the SPBD account in two aspects. It posits that both the end effect and the CE are derived from the same bias factor and that the closer the pair is to an end, the larger the bias and hence the larger the CE.

The most notable difference between the SPBD and other theories is that the SPBD account claims that the CE is essentially a serial position effect within a pair of items and that the comparative in the instruction plays no necessary or causal role in the generation of the CE. The relative difficulty of the two items is created independently of the comparative. The only function of the comparative is to direct which item to choose, the "easier" or the "harder" item. It contributes nothing in the making of the easier and harder items, as verified by Jou et al. (2018). Importantly, this new idea provides a coherent account for a few seemingly inconsistent findings about CE in the literature and a new perspective to reexamine an old issue. We will return to this point later.

One criticism of the SPBD account is that the local distinctiveness created in Jou et al.'s (2018) study was through strengthening the item–serial position association, instead of isolating the middle item from its neighbors as others have done (e.g., Neath et al., 2006), and that that might be why the manipulation did not produce an effect on the comparative judgment task. In the following we report two experiments conducted to address that criticism.

Further Testing of the SPBD Account

The local distinctiveness theory (Neath & Brown, 2006; Neath et al., 2006) is a serious rival to the global distinctiveness theory as a model of order memory. The crucial point of difference between the two theories is that, according to the local distinctiveness model, a local serial item can be isolated from its neighbors and gain an unusual advantage in memory over the other items on the list, and as a result it can alter the normal shape of the bowed serial position function. As noted earlier, repeated presentations of a midseries item in Jou et al. (2018) did not produce an effect on the serial position curve or the CE in the comparative judgment task. However, according to some researchers in this area, increasing the memory strength of an item may not be the most effective way of isolating an item from the rest of the items. Thus, one potentially outstanding issue for the SPBD theory is that the repeated presentation method used in Jou at al.'s 2018 study was not the standard isolation procedure and therefore not very effective. To address this issue, in two experiments we doubled the interval rank distance between the rank-4 person and his two neighbors in an absolute and a comparative judgment task on the memorized order.

EXPERIMENT 1

In Experiment 1, we used the same absolute judgment task as in Jou et al.'s (2018) study, in which subjects first learned a list of eight people's names ordered by height and then took an absolute judgment test. In the control condition, the eight names were ranked 1 through 8: 1, 2, 3, 4, 5, 6, 7, 8, with 1 presenting the shortest and 8 the tallest person's height rank. In the experimental condition, one of the midseries items (the original item number 4) was changed to 5, and the person following it was changed to 7, thus producing this series: 1, 2, 3, 5, 7, 8, 9, 10. In this rank coding, number 5 stands out by being more separated from its neighbors than the rest of the group. The results of this experiment were compared with those of Experiment 2, where a comparative judgment task was used.

METHOD

Subjects

Fifty undergraduate psychology students, 34 women and 16 men, participated in the control condition for course credit. Their mean age was 22.08 years, ranging from 18 to 47. Everybody performed at least at the .70 accuracy level. Eighty-one undergraduate psychology students, 59 women and 22 men, participated in the experimental condition.[2] Their mean age was 20.44 years, ranging from 18 to 40. The data for 8 of them did not meet the minimal criterion of .70 accuracy, and they were excluded from analysis, leaving 73 subjects in the analysis.

Design, Materials, and Procedure

The eight names used were eight 4-letter common English male names: John, Mike, Paul, Carl, Bill, Jeff, Rich, and Dave. In the control condition, each name was randomly assigned a height rank numbered from 1 to 8, with 1 representing the shortest and 8 the tallest. In the experimental condition, the rank numbers used were 1, 2, *3*, *5*, 7, 8, 9, 10. Subjects learned the sequence by studying the name–rank pairs displayed in the center of a computer monitor, one pair at a time, for 2.5 s each, with the name on top and the number below it separated by a blank space. The interstimulus interval was 0.5 s. A default minimum four rounds of learning was set with an option of taking additional rounds of learning if the subject so desired. The eight name–number pairs were presented in a new random order for each round of display.

Subjects were told that after they had learned the order, they would take two memory tests with the nature of the memory tests unspecified at that point. They were asked to test themselves by trying to recall the names sequentially at the end of each round of study. At the end of the fourth round of presentation and of any round after the fourth round, a question “Ready for the memory tests?” appeared. They pressed the “y” key to go to the tests if they were ready and the “n” key to continue the learning process otherwise. They took two memory tests, a recall test followed by the absolute judgment test. Before the tests began (when they could still return to learning), subjects were told that they should be at least 90% accurate in the memory tests to pass the test, or they would be asked to repeat the experiment (although the rule was not enforced) to discourage insufficient learning and careless responses.

The purpose of the recall test was to check the extent to which subjects remembered the order. At each trial in the test, a randomly selected name was displayed along with an underlined space displayed on its right-hand side into which subjects could type a number to indicate that name’s height rank. Each name appeared only once in the test. After the recall test, the absolute judgment test began.

For the absolute judgment test, eight correctly matched pairs were made, and another eight incorrectly matched pairs were created by scrambling the names and numbers into mismatched pairs. A round of test consisted of eight correctly and eight incorrectly matched pairs, making a total of 16 test pairs. In a round of test, each of the 16 test pairs was presented once, and the 16 test items were presented in a new random order for each subject at each round of testing. The test round was repeated four times, making a total of 64 trials for the whole test. The names and numbers of the incorrectly matched pairs were randomly re-paired across the test rounds. The “z” and the “/” were designated as two response keys, and the mapping of the two keys to the “yes” and “no” responses was counterbalanced across subjects. Subjects were asked to respond as quickly and as accurately as they could. There was a short break in the middle of the 64 trials.

RESULTS AND DISCUSSION

Subjects in the control condition repeated the learning cycle 6.00 times on average. Their recall test overall accuracy was .956. The overall accuracy of their absolute judgment was .970.[3] Nobody performed below the .70 minimum accuracy level. The error data were not included in the analyses. RTs greater than the mean plus three standard deviations were excluded from analysis. These RT outliers made up 2.5% of the total control condition data. Subjects in the experimental condition took an average of 6.70 rounds of learning. Their overall recall accuracy was .904. Their absolute judgment overall accuracy was .972. The outliers excluded made up 3.4% of the total data in the experimental condition. The data from the mismatched pairs were not used in the analyses.

The mean RTs as a function of serial position and condition (control vs. experimental) are presented in Figure 3.

A visual inspection of the figure indicated that there was a noticeable dip in the middle of the RT

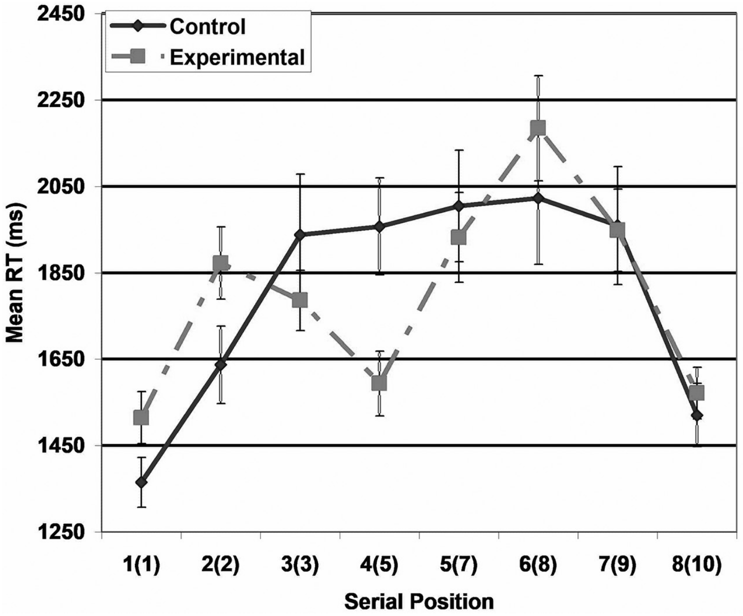

FIGURE 3. Mean response times (RTs) of the absolute judgments as a function of the magnitude coding condition and serial position, Experiment 1. The numbers in parentheses on the *x*-axis scale markers are the experimental condition rank codes

curve for the experimental but not for the control condition. An analysis of variance (ANOVA) with condition as a between-subject and the ordinal serial position as a within-subject factor indicated that the main effect of condition (mean of experimental = 1,800 ms, mean of control = 1,855 ms) was not significant, $F < 1$. The main effect of serial position was significant, $F(7, 932) = 23.48$, $MSE = 450{,}925$, $p < .0001$, $\eta_p^2 = .150$. The condition by serial position interaction was also significant, $F(7, 932) = 2.61$, $MSE = 450{,}925$, $p = .011$, $\eta_p^2 = .020$.

The two RT curves in Figure 3 crossed each other at more than one location, and therefore the significant interaction could have derived from a combination of these differences in the shapes of the two functions. To confirm the visual observation that the experimental curve revealed a double bowing whereas the control curve had only one bowing, an orthogonal quartic trend test (which detects a double bowing) was applied to these two curves. For the experimental curve, the quartic trend was significant, $F(1, 590) = 16.84$, $MSE = 466{,}085$, $p < .0001$, $\eta_p^2 = .028$. For the control curve, this trend was not significant, $F(1, 342) = 1.53$, $MSE = 426{,}144$, $p = .217$, $\eta_p^2 = .004$, thus confirming the visual observation.

These results are consistent with the findings by Jou et al. (2018), Jou (2019), and Neath et al. (2006). When a midseries term is made more salient by either isolating it from its neighbors with a greater interval distance (Neath et al., 2006), reinforcing the memory of that item (Jou et al., 2018), or making the item more prominent in memory in a previous learning episode (Jou, 2019), the item produced an isolation or local distinctiveness effect in an absolute judgment task in the form of a faster response to that item.

However, the more important goal is to determine whether a greater interval distance between the midseries item and its neighbors has an effect on the CE of a comparative judgment. Experiment 2 investigated this question.

EXPERIMENT 2

Jou and Aldridge (1999) suggested that the interval-level distance is an irrelevant factor in a memorial comparative judgments in which the judgment is assumed to be based on an ordinal scale. Similarly, Henderson and Well (1985) showed that learning an ordering of circles (in which each circle was represented by a color name) with large and small size intervals between them produced the same results in a symbolic comparative judgment as did learning a series of circles with equal-sized intervals or as simply learning an arbitrary ordering. Also, Potts (1974) did not find a performance difference in discrimination between adjacent pair members separated by variable distances. Based on these findings from previous studies, we predicted that we would be more likely to obtain the same result than not. The only difference between this experiment and Experiment 1 was that in this experiment, subjects took a comparative judgment instead of an absolute judgment test.

METHOD

Subjects

Seventy-four undergraduate psychology students, 42 women and 32 men, participated in the control condition for course credit. Their mean age was 20.84 years, ranging from 18 to 34. The data for four subjects fell below the minimal criterion of .70 correct, and they were excluded from analysis, leaving 70 subjects in the analysis. Seventy-nine undergraduate psychology students, 52 women and 27 men, participated in the experimental condition. Their mean age was 19.83 years, ranging from 18 to 33. Four of them scored below the minimum criterion of .70 accuracy and were excluded from the analysis, leaving 75 subjects in the analysis.

Design, Materials, and Procedure

The eight names, the manipulation of the rank numbers, the studying procedure, and the recall test were the same as in Experiment 1. The test following the recall test was a comparative judgment test. The test name pairs were constructed by pairing each rank with all other ranks, yielding 56 pairs (an 8 × 7 matrix), with half (28) of the pairs having the name of the lower rank on the left and the name of the higher rank on the right side (forward pairs), and the other half (28 pairs) with the order of the two names reversed (backward pairs). These 56 pairs of names were divided into two subsets of 28 pairs each. In each subset, 14 pairs were forward and 14 were backward pairs. Thus, each pair in a subset was unique; that is, if the forward version of a pair was selected for inclusion in Subset 1, the backward version of that pair would be assigned to Subset 2 and vice versa. Therefore, the two subsets counterbalanced the forward and backward orders of the two names.

The order of presenting the two subsets within the 56-trial block was counterbalanced across subjects. The test pairs within the 28-trial subset were presented in a new random order for each subject at each round of presentation. The 56-trial block was tested for two rounds, with the presentation order of the two subsets within the block counterbalanced across the two repeated blocks. Thus, the total number of trials was 112. Half of the trials were presented with a "choose the taller person" and the other half with "choose the shorter person" instruction displayed at the middle location on top of the test pair. The "taller" and "shorter" instructions were randomly mixed together in a block.

The instruction before the test told the subjects that a pair of names representing one taller and one shorter person would be displayed in the center of the screen on a line with one name on the left and one on the right side, separated by five spaces, and that their task was to choose the taller or shorter of the two people as specified by the instruction. The two response keys were the "z" key on the left side and the "/" key on the right side. Subjects pressed the key corresponding to the side on which the instruction-specified target person appeared. There was a short break between the two 56-trial blocks.

RESULTS AND DISCUSSION

In the control condition, subjects took an average of 5.57 learning cycles. The overall recall accuracy was .931. The overall accuracy in the comparative judgment task was .948. The outliers made up 2.5% of the data. In the experimental condition, subjects took an average of 6.21 rounds of learning. The overall recall accuracy was .915. The overall comparative judgment accuracy was .927. The outliers made up 3.2% of the data. The error data and outliers were not included in the analysis.

An ANOVA with condition (experimental vs. control) as a between-subject factor and test cycle (cycle 1 vs. cycle 2) as a within-subject factor showed no significant condition effect (mean of experimental condition = 3,216 ms; mean of control condition = 3,322 ms), $F < 1$. The main effect of test cycle (mean of cycle 1 = 3,591 ms; mean of cycle 2 = 2,945 ms) was significant, $F(1, 142) = 154.42$, $MSE = 194{,}477$, p

< .0001, $\eta_p^2 = .521$. The condition × test cycle interaction was not significant, $F < 1$.

The mean RTs of the pairs with the split size of 1 ordinal step of the control condition are presented as a function of the ordinal rank (i.e., the lower of the two serial positions of the two items) in Figure 4 and those of the experimental condition in Figure 5.[4]

A visual comparison between these two figures indicated that the overall patterns of these two sets of curves appeared similar, although the experimental "choose the shorter" curve showed a slight dip at position 3. An ANOVA with choice and serial position (the smaller of the two ranks in a pair) as two within-subject factors was conducted to verify the presence

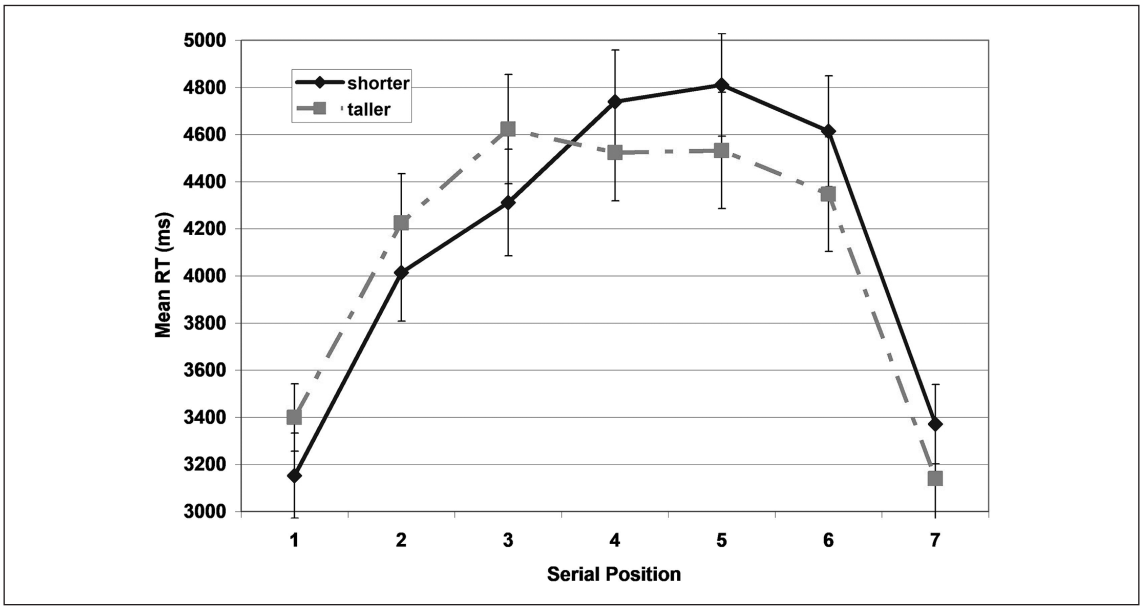

FIGURE 4. Mean response times (RTs) of comparative judgments of the control condition as a function of choice and serial position, Experiment 2

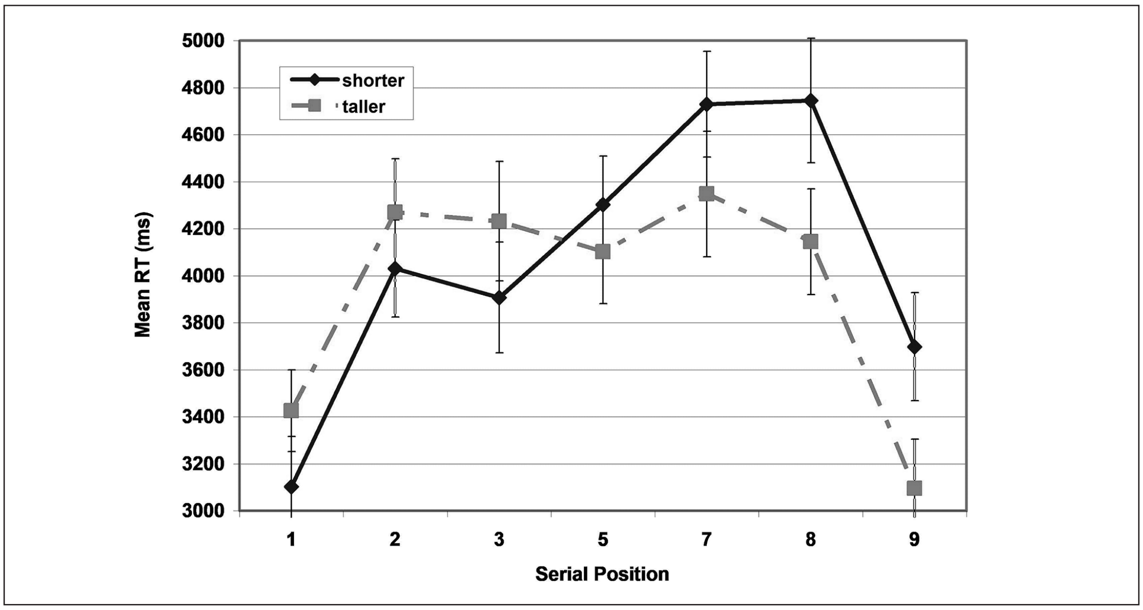

FIGURE 5. Mean response times (RTs) of comparative judgments of the experimental condition as a function of choice and serial position, Experiment 2

of the CE (an interaction between the choice and the serial position variables) for the experimental and the control condition, respectively.[5] For the experimental condition, the choice main effect (mean of "choose taller" = 3,988 ms, mean of "choose shorter" = 4,146 ms) was marginally significant, $F(1, 73) = 2.83$, $MSE = 2{,}138{,}821$, $p = .097$, $\eta_p^2 = .037$. The serial position effect was significant, $F(6, 429) = 13.40$, $MSE = 2{,}569{,}357$, $p < .0001$, $\eta_p^2 = .158$, as was the choice × serial position interaction, $F(6, 386) = 2.95$, $MSE = 1{,}770{,}578$, $p = .008$, $\eta_p^2 = .044$.

The corresponding ANOVA for the control condition data showed that the choice main effect (mean of "choose taller" = 4,152 ms, mean of "choose shorter" = 4,157 ms) was not significant, $F \approx 0$. The main effect of serial position was significant, $F(6, 409) = 19.91$, $MSE = 2{,}695{,}624$, $p < .0001$, $\eta_p^2 = .163$, as was the choice × serial position interaction, $F(6, 388) = 2.68$, $MSE = 1{,}596{,}951$, $p = .015$, $\eta_p^2 = .040$. Thus, both sets of data showed a significant CE.

The next ANOVA using condition as a between-subject and serial position as a within-subject factor compared across the two conditions the two serial positions involving the isolated midseries term in the experimental condition, that is, the ordinal rank 4 term recoded as rank 5 in the experimental condition. The two serial positions involving that ordinal rank were the original ordinal serial positions 3 (forming the pair 3–4) and 4 (forming the pair 4–5) for the control and positions 3 (forming the pair 3–5) and the re-coded 5 (forming the pair 5–7) for the experimental condition. The condition main effect (mean of experimental condition = 3,507 ms, mean of control condition = 3,655 ms) was not significant, $F < 1$. Neither the serial position main effect (mean of ordinal rank 3 = 3,627 ms, mean of ordinal rank 4 = 3,531 ms), $F(1, 142) = 1.95$, $MSE = 346{,}292$, $p = .165$, $\eta_p^2 = .014$, nor the serial position × condition interaction was significant, $F < 1$. More important, the RT relative magnitude orders of the two items in each pair, which defined the CE, were the same across the experimental and control conditions.

Although the isolated items did not change the crossover interaction pattern from the control counterparts, did the two pairs containing the isolated item produce lower RTs than their two neighboring pairs? To find that out, an ANOVA with choice and serial position as two factors compared the two target pairs with their two neighbors (covering the middle four pairs in Figure 5). The choice main effect was not significant, $F < 1$. The serial position effect was not significant, either, $F(3, 212) = 1.57$, $MSE = 2{,}244{,}020$, $p = .198$, $\eta_p^2 = .022$. The choice × serial position interaction was marginally significant, $F(3, 182) = 2.48$, $MSE = 1{,}673{,}279$, $p = .063$, $\eta_p^2 = .040$. Although the interaction was marginally significant, the evidence of the target pairs producing lower RT than their neighbors was weak at best. The crucial point here is that the isolated item (rank number 5 item) was not chosen faster than the pair mate rank number 3 item (which is closer to the low endpoint than item 5 to the high endpoint) and therefore did not change the CE pattern.

Thus, it seems that the item extremity principle worked as usual despite the fact that one of the two items in the midseries pairs was isolated. Thus, isolating the midseries item by increasing the interval distance between it and its neighbors, like increasing the memory strength of this item (Jou et al., 2018), did not alter the CE pattern relative to not isolating the item. One question is whether the lack of an effect was caused by the manipulation being weak. At any rate, the manipulation produced a significant effect in Experiment 1. Therefore, the discrepancy in the results between the two experiments should have resulted from the different natures of the memory tests (Greene et al., 1998) more than from a weak manipulation. Jou (2019) suggested that an absolute judgment task taps more into item-specific information than relational information, whereas a comparative judgment task measures primarily between-item, relational information that is indexed by the global distinctiveness of the serial positions.

Jou (2019) further suggested that the comparative judgment task actually works to suppress the local distinctiveness created in a previous learning episode. To demonstrate this, he had subjects perform the comparative judgment task for three repeated rounds. With each repeated round of testing, the RTs of all serial positions decreased significantly except for the previously isolated midseries position. As a result, the residual local distinctiveness effect was all but completely "worn off" at the end of the third round of testing. This unequal RT decrease across serial positions over the repeated rounds of testing was not found in the absolute judgment condition (where the RTs of all positions dropped equally over the repeated rounds of testing).

GENERAL DISCUSSION

To summarize, the two experiments reported here answered the lingering question of whether Jou et al.'s (2018) results that favored the global distinctiveness over the local distinctiveness theory were due to the fact that the isolating method used in that study was not actually isolating the middle item from its neighbors. The results of these two experiments were consistent with Jou et al.'s (2018) outcome, and the results of Experiment 2 were in accord with Henderson and Well's (1985) and Potts's (1974) conclusion that interval level distance differences had no effect on comparative judgments. Thus, these two experiments resolved an unsettled issue in Jou et al.'s (2018) study and thereby reaffirmed the SPBD account of CE.

Lastly, a related question to be answered is whether any serial order memory models should be held accountable for data obtained from both the single-item retrieval and the pair comparison paradigms. Both the global and the local distinctiveness models were formulated based on the single-item retrieval paradigm, primarily serial recall. The original authors did not test their models with the comparative judgment task. However, because comparative judgments are a widely used index of serial order memory, any serial order memory models should take comparative judgment data into consideration. Although the global distinctiveness theory fails to explain the local distinctiveness phenomenon, it is actually designed to provide a powerful account for the unidimensionally structured linear order memory tapped by comparative judgments. In that paradigm, only the relative order held among the items counts; individual salient characteristics are irrelevant. Overall, it may be fair to say that each of the two order memory models is constrained by the specific order information measured in each paradigm.

The SPBD account can provide a coherent explanation for several divergent and puzzling findings in the area of the CE research. For this reason, we believe that the SPBD account provides a fresh, possibly more accurate notion of what the CE is about. We describe those findings below.

CE Obtained in Blocked Instruction Presentation Mode

As noted, the CE remains when instruction is presented in the blocked form. Because in that design the comparative is the same in a block of trials, it obviously loses any cuing or predictive function for the magnitude of the stimulus items. So it is not clear why the CE remains (as hinted by Marschark & Paivio, 1979). According to the SPBD account, the item relative extremity determines the CE, and the comparative cannot influence it. That is why the CE remains under the blocked instruction condition.

CE Found in Stimuli Presented Before and Simultaneously with the Instruction

The CE resisted instruction presentation manipulation in studies that attempted to eliminate it. As noted, when the stimulus pair was presented before or simultaneously with the instruction, the CE did not go away. This is expected because the CE arises independently of the instruction.

CE Arose When the Right-Side Item's Relative Magnitude Was Queried

The CE occurred as usual when subjects were asked to always indicate the relative magnitude of the right-hand side item in a pair (Cech, 1989). Although Cech used that finding as evidence against the expectancy hypothesis and in favor of the semantic coding model, he did hint that the occurrence of the CE might not require specification of a comparative in the instruction.

Multiple CEs

In one study (Cech et al., 1990), after subjects learned a list of 12 items, from the smallest flea to largest elephant, they took a comparative judgment test that was based on six continuous items of the learned longer list (e.g., animal number 1 to 6 or animal number 7 to 12). Cech et al. (1990) found that they could obtain multiple congruity effects in that each half of the original longer list produced a CE. The authors argued that this finding was inconsistent with the expectancy model, but it lent support to an expanded semantic coding model as the stimulus magnitude was assumed to have been re-coded in the partial lists. From the SPBD perspective, although subjects learned a longer list, the partial list the comparative judgment test actually used was the task-relevant list.

The list actually tested, not originally learned, defines which item in a pair is the more extreme of the two items.

CE Magnitude Is Larger for Easier Comparisons

Unlike what the evidence accrual model asserts, the CE is typically larger when the comparison is easier rather than more difficult. Figures 1 and 2 show that the CE is the largest at the two ends of the inverted U-shaped curve, where the responses are the fastest, and is the smallest at the middle point of the series, where the responses are the slowest. This follows directly from Murdock's (1960) theory. According to Murdock's (1960) serial position distinctiveness theory, the two end items are the most distinctive, and the middle item is the least distinctive. The total distinctiveness of the serial items shows that it increases from the middle items toward the two end items of the series. Not only does this explain the bow-shaped serial position function of serial order memory, but it also explains why the CE increases toward the two ends of a series. The larger the difference in distinctiveness between the two paired items, the larger the RT difference is for selecting one versus the other item. For example, the total distinctiveness of item 1 in Table 1 is 28, and that of item 2 is 22, so the difference is 6. The total distinctiveness of item 2 is 22, and that of item 3 is 18, and the difference is 4; and the difference between items 3 and 4 is 2. Thus, because the distinctiveness difference between the two items in a pair increases toward the two endpoints of the series, so does the response time difference and hence the size of the CE. This principle also holds for pairs with a larger than one-step interitem distance (see Jou et al., 2018). This finding directly contradicts a major assertion of the evidence accrual theory: that the CE increases as the discrimination becomes more difficult.

CE Is Variable (Unstable) in Perceptual Comparative Judgments

According to some researchers, there are fundamental differences in judgment processes between perceptual and memorial comparisons (Banks, Mermeltein, & Yu, 1982; Marschark & Pavio, 1979, 1981).

The CE in perceptual comparative judgments was either nonexistent or inconsistently found (Banks, Mermelstein, & Yu, 1982; Marschark & Paivio, 1981; see Petrusic & Baranski, 1989 for limiting conditions for perceptual CE, i.e., perceptual CE occurs only for confusable stimuli). According to Banks et al. (1982) and Marschark and Paivio (1981), in the cases where the CE was found in perceptual comparisons, some type of memorial, symbolic representations of the stimuli were created or incidentally introduced. For example, when a small number of stimuli were repeatedly presented, the procedure created a symbolic memory representation of the stimuli (e.g., in Audley & Wallis, 1964 study, only two stimulus pairs were used in comparison, and each was judged 24 times) (Marschark & Paivio, 1981) and could partially or completely transform the perceptual comparison into a memorial comparison. According to the SPBD idea, the CE results from differential retrieving speeds of the two stimuli from memory. If the stimuli are not represented in memory, naturally they do not need to be retrieved from memory, hence the absence of the CE.

CE in Bilateral Pairs

The traditional definition of the CE as choosing the larger (rather than the smaller) of two large items or the smaller (rather than the larger) of two small items faster and more accurately is actually a special case of the new definition of the CE as choosing the positionally more extreme item faster and more accurately than the positionally less extreme item. For a pair of unilateral items that falls on one side of the array (assuming the series is divided at the middle point into a low half and high half), the prediction of the SPBD account and that of the traditional concept of the CE coincide. That is, the larger of the two large items or the smaller of the two small items is also the more extreme of the two items. However, for the bilateral pairs in which one item falls on the small side and the other on the large side, the traditional concept of CE makes no prediction. For example, assume there are 10 serial order items, and a probe pair contains item 3 and 9. The SPBD theory predicts that item 9 will be chosen faster than item 3, but it would be a big stretch to call both items "large" because 3 is quite small on a 10-point scale. We have not seen a CE theory that attempts to deal with the question of why item 9 is chosen faster than item 3, perhaps with the exception of the biased start point theory. Of course, when a pair contains an end item and non-end item, the end item (the most extreme item) will be chosen faster than the non-end item, which is referred to as the end anchor effect (Banks, 1977; Birnbaum & Jou, 1990; Jou, 2010). The idea of an item extremity effect in comparative judgments gives a conceptually coherent explanation to the relative speed of choosing one versus the other item in unilateral or bilateral pairs and pairs containing an end item.

Summary and Conclusions

The new findings from Jou et al. (2018) have fundamentally changed the concept of what the CE in comparative judgments is. That is, the comparative in the instruction plays no necessary or causal role, such as generating a comparative magnitude code or an expectation of the magnitude of the stimuli, or setting up a reference point, as has been believed since the 1970s. The same CE will arise whether there is an instruction or not and regardless of how it is presented. The CE comes from the SPBD of the serial order items. It is a manifestation of the serial position effect within a pair of serial order items. As noted, the congruent or incongruent relation between the comparative in the instruction and the relative positional extremity of the two stimulus items is coincidental rather than causal in the sense that the comparative cannot change the fact that choosing the positionally more extreme item is faster. However, if the comparative happens to designate the more extreme item, one gets a faster response than if it designates the less extreme item.

The new perspective on the CE helps us understand in a coherent manner many divergent, puzzling findings related to the issue of the CE and broadens the previous definition of the CE as a facilitation in choosing the larger (or smaller) of two large (or small) things to a facilitation in choosing the more extreme of the two serially ordered objects. Finally, the two experiments in this study reaffirmed the conclusion reached in Jou et al.'s (2018) study that the serial position functions and the CE in comparative judgments are based on global distinctiveness of the items, not local distinctiveness of the serial items.

NOTES

We are grateful to Alice Healy for her very careful review of a draft of this article and constructive comments, to an anonymous reviewer for the beneficial critiques, and to Robert Proctor for his very helpful editorial feedback, all of which made this a much better article.

Address correspondence about this article to Jerwen Jou, Department of Psychological Science, University of Texas–Rio Grande Valley, 1201 West University Drive, Edinburg, TX 78539-2999 (e-mail: jerwen.jou@utrgv.edu).

1. In the literature, the congruity effect is commonly referred to as the semantic congruity effect. This term derived from an influential CE theory, known as the semantic coding theory of congruity effects (Banks, 1977). As we demonstrate later in the article, the conceptualization regarding the semantic aspect of the effect in that theory is questionable. Therefore, we no longer follow this naming tradition in this article.

2. Toward the end of the semester, we encountered great difficulty in obtaining enough subjects for our lab-based study because of a change in the subject pool participation rule regarding the proportion of time students could spend on online studies compared with lab-based studies. Because the results of the control conditions with their standard serial learning materials were highly predictable compared with the experimental condition with its changed learning materials, and therefore with the results much less certain, we put more subjects in the experimental condition to give more data for a solid conclusion.

3. The recall overall accuracy was calculated with everyone included. The judgment overall accuracy was calculated after the subjects failing to meet the .70 accuracy criterion were removed. We found that subjects who did poorly in the recall tests often did better in the judgment tests. This might be due to the removal of low-accuracy subjects from judgment data analyses and probably also to the recognition tests being easier than recall tests.

4. Although the two items in pairs 3–5 and 5–7 had an interval split size of 2, their ordinal rank split size was 1 step.

5. Tests for pairs of larger split sizes produced similar patterns of results. The figures and test results are available upon request.

REFERENCES

Audley, R., & Wallis, C. (1964). Response instructions and the speed of relative judgments: I. Some experiments on brightness discrimination. *British Journal of Psychology, 55,* 59–73.

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. 101–159). New York, NY: Academic Press.

Banks, W. P., Clark, H. H., & Lucy, P. (1975). The locus of the semantic congruity effect in comparative judgments. *Journal of Experimental Psychology: Human Perception and Performance, 1,* 35–47.

Banks, W. P., & Flora, J. (1977). Semantic and perceptual processes in symbolic comparison. *Journal of Experi-*

mental Psychology: Human Perception and Performance, 3, 278–290.

Banks, W. P., Fujii, 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.

Banks, W. P., Mermelstein, R., & Yu, H. K. (1982). Discriminations among perceptual and symbolic stimuli. *Memory & Cognition, 10,* 265–278.

Banks, W. P., & Root, M. (1979). Semantic congruity effects in judgment of loudness. *Perception & Psychophysics, 26,* 133–142.

Banks, W. P., & White, H. (1985). Semantic congruity and expectancy as separate processes. *Memory & Cognition, 13,* 485–493.

Banks, W. P., White, H., Sturgill, W., & Mermelstein, R. (1983). Semantic congruity and expectancy in symbolic judgments. *Journal of Experimental Psychology: Human Perception and Performance, 9,* 560–582.

Bhatarah, P., Ward, G., & Tan, L. (2008). Examining the relationship between free recall and immediate serial recall: The serial nature of recall and the effect of test expectancy. *Memory & Cognition, 36,* 20–34.

Birnbaum, M. H., & Jou, J. (1990). A theory of comparative response time and "difference" judgments. *Cognitive Psychology, 22,* 184–210.

Cech, C. G. (1989). Congruity and the expectancy hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 1129–1133.

Cech, C. G. (1995). Is congruity due to encoding? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 1275–1288.

Cech, C. G., & Shoben, E. J. (1985). Context effects in symbolic magnitude comparisons. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11,* 299–315.

Cech, C. G., Shoben, E. J., & Love, M. (1990). Multiple congruity effects in judgments of magnitude. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 1142–1152.

Crowder, R. G. (1976). *Principles of learning and memory.* Hillsdale, NJ: Erlbaum.

Crowder, R. G., & Greene, R. L. (2000). Serial learning: Cognition and behavior. In E. Tulving & F. I. M. Craik (Eds.), *The Oxford handbook of memory* (pp. 125–135). New York, NY: Oxford University Press.

Duncan, E. M., & McFarland, C. E. Jr. (1980). Isolating the effects of symbolic distance and semantic congruity in comparative judgments: An additive-factors analysis. *Memory & Cognition, 8,* 612–622.

Ebbinghaus, H. (1964). *Memory: A contribution to experimental psychology* (H. A. Ruger & C. E. Bussenius, Trans.). New York: Dover. (Original work published 1885)

Foltz, G. S., Poltrock, S. E., & Potts, G. R. (1984). Mental comparison of size and magnitude: Size congruity effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 442–453.

Greene, R. L., Thapar, A., & Westerman, D. L. (1998). Effects of generation on memory of order. *Journal of Memory and Language, 38,* 255–264.

Henderson, J. M., & Well, A. D. (1985). Symbolic comparisons with and without perceptual referents: Is interval information used? *Memory & Cognition, 13,* 176–182.

Holyoak, K. J. (1978). Comparative judgments with numerical reference points. *Cognitive Psychology, 10,* 203–243.

Holyoak, K. J., & Mah, W. A. (1981). Semantic congruity in symbolic comparisons: Evidence against an expectancy hypothesis. *Memory & Cognition, 9,* 197–204.

Holyoak, K. J., & Mah, W. A. (1982). Cognitive reference point in judgments of symbolic magnitude. *Cognitive Psychology, 14,* 328–352.

Holyoak, K. J., & Walker, J. H. (1976). Subjective magnitude information in semantic orderings. *Journal of Verbal Learning and Verbal Behavior, 15,* 287–299.

Howard, R. W. (1983). The semantic congruity effect: Some tests of the expectancy hypothesis. *Acta Psychologica, 53,* 205–216.

Hunt, R. R. (1995). The subtlety of distinctiveness: What von Restorff really did. *Psychonomic Bulletin & Review, 2,* 105–112.

Hunt, R. R., & Einstein, G. O. (1981). Relational and item-specific information in memory. *Journal of Verbal Learning and Verbal Behavior, 20,* 497–514.

Hunt, R. R., & McDaniel, M. A. (1993). The enigma of organization and distinctiveness. *Journal of Memory and Language, 32,* 421–445.

Jamieson, D. G., & Petrusic, W. M. (1975). Relational judgments with remembered stimuli. *Perception & Psychophysics, 18,* 373–378.

Jou, J. (2010). The serial position, distance, and congruity effects of reference point setting in comparative judgments. *American Journal of Psychology, 123,* 127–136.

Jou, J. (2014). Task-switching cost and repetition priming: Two overlooked confounds in the fixed-set procedure of the Sternberg paradigm and how they affect memory set-size effect. *Quarterly Journal of Experimental Psychology, 67,* 1871–1894.

Jou, J. (2019). Combining two separate series into a single ordering: Testing the local and global distinctiveness theories with absolute and relative judgments. *Consciousness and Cognition, 72,* 19–30.

Jou, J., & Aldridge, J. W. (1999). Memory representation of alphabetic position and interval information. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 680–701.

Jou, J., Escamilla, E. E., Torres, A. U., Ortiz, A., & Salazar, P. (2018). Where does the congruity effect come from in memorial comparative judgments? A serial-position–based distinctiveness account. *Journal of Memory and Language, 103,* 127–150.

Kent, C., & Lamberts, K. (2005). An exemplar account of the bow and set-size effects in absolute identification. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 289–305.

Leth-Steensen, C., Petrusic, W. M., & Shaki, S. (2014). Enhancing semantic congruity effects with category-contingent comparative judgments. *Frontiers in Psychology, 5,* Article 1199, 1–7.

Link, S. W. (1978). The relative judgment theory analysis of response time deadline experiments. In N. J. Castellan & F. Restle (Eds.), *Cognitive theory* (Vol. 3). Hillsdale, NJ: Erlbaum.

Link, S. W. (1991). The psychological design of psychological experiments. In H. G. Geisler (Ed.), *Psychophysical explorations of mental structures* (pp. 15–35). Toronto, ON: Hogrefe & Huber.

Los, S. A. (1996). On the origin of mixing costs: Exploring information processing in pure and mixed blocks of trials. *Acta Psychologica, 94,* 145–188.

Marks, D. F. (1972). Relative judgment: A phenomenon and a theory. *Perception & Psychophysics, 11,* 156–160.

Marschark, M. (1983). Semantic congruity in symbolic comparisons: Salience, expectancy, and associative priming. *Memory & Cognition, 11,* 192–199.

Marschark, M., & Paivio, A. (1979). Semantic congruity and lexical marking in symbolic comparisons: An expectancy hypothesis. *Memory & Cognition, 7,* 175–184.

Marschark, M., & Paivio, A. (1981). Congruity and the perceptual comparison task. *Journal of Experimental Psychology: Human Perception and Performance, 7,* 290–308.

Marshuetz, C. (2005). Order information in working memory: An integrative review of evidence from brain and behavior. *Psychological Bulletin, 131,* 323–339.

Moyer, R. S., & Bayer, R. H. (1976). Mental comparison and the symbolic distance effect. *Cognitive Psychology, 8,* 228–246.

Moyer, R. S., & Landauer, T. K. (1967). The time required for judgments of numerical inequality. *Nature, 215,* 1519–1520.

Mulligan, N. W. (1999). The effects of perceptual interference at encoding on organization and order: Investigating the roles of item-specific and relational information. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 54–69.

Mulligan, N. W. (2000). Perceptual interference at encoding enhances item-specific encoding and disrupts relational encoding: Evidence from multiple recall tests. *Memory & Cognition, 28,* 539–546.

Mulligan, N. W. (2001). Word frequency and memory: Effects on absolute versus relative order memory and on item memory versus order memory. *Memory & Cognition, 29,* 977–985.

Murdock, B. B. Jr. (1960). The distinctiveness of stimuli. *Psychological Review, 67,* 16–31.

Nairne, J. S. (1988). A framework for interpreting recency effects in immediate serial recall. *Memory & Cognition, 16,* 343–352.

Neath, I., & Brown, G. D. A. (2006). Making distinctiveness models of memory distinct. In J. S. Nairne (Ed.), *The foundation of remembering: Essays in honor of Henry L. Roediger, III* (pp. 125–140). New York, NY: Psychology Press.

Neath, I., Brown, G. D. A., McCormack, T., Chater, N., & Freeman, R. (2006). Distinctiveness models of memory and absolute identification: Evidence for local, not global, effects. *Quarterly Journal of Experimental Psychology, 59,* 121–135.

Petrusic, W. M. (1992). Semantic CEs and theories of the comparison process. *Journal of Experimental Psychology: Human Perception and Performance, 18,* 962–986.

Petrusic, W. M., & Baranski, J. V. (1989). Semantic congruity effects in perceptual comparison. *Perception & Psychophysics, 45,* 439–452.

Petrusic, W. M., Shaki, S., & Leth-Steensen, C. (2008). Remembered instructions with symbolic and perceptual comparisons. *Perception & Psychophysics, 70,* 179–189.

Potts, G. R. (1974). Incorporating quantitative information into a linear ordering. *Memory & Cognition, 3,* 553–538.

Proctor, R. W., & Cho, Y. S. (2006). Polarity correspondence: A general principle for performance of speeded binary classification tasks. *Psychological Bulletin, 132,* 416–442.

Restle, F., & Brown, E. R. (1970). Organization of serial pattern learning. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 4, pp. 249–331). New York, NY: Academic Press.

Rogers, R. D., & Monsell, S. (1995). Costs of a predictable switch between simple cognitive tasks. *Journal of Experimental Psychology: General, 124,* 207–231.

Shaki, S., Leth-Steensen, C., & Petrusic, W. M. (2006). Effects of instruction presentation mode in comparative judgments. *Memory & Cognition, 34,* 196–206.

Shipley, W. C., Coffin, J. I., & Hadsell, K. C. (1945). Affective distance and other factors determining reaction time in judgments of color preference. *Journal of Experimental Psychology, 35,* 206–215.

Shipley, W. C., Norris, E. D., & Roberts, M. L. (1946). The effect of changed polarity of set on decision time of affective judgments. *Journal of Experimental Psychology, 36,* 237–243.

Shoben, E. J., Cech, C. G., Schwanenflugel, P. J., & Sailor, K. M. (1989). Serial position effects in comparative judgments. *Journal of Experimental Psychology: Human Perception and Performance, 15,* 273–286.

Shoben, E. J., Sailor, K. M., & Wang, M. Y. (1989). The role of expectancy in comparative judgments. *Memory & Cognition, 17,* 18–26.