On drivers' reasoning about traffic signs: the case of qualitative location

Ana Hernando[1], Antonio Lucas-Alba[1*], Maria Teresa Blanch[2], Andrés Sebastián Lombas[1], Alberto Arbaiza[3], Hans Remeijn[4], Gunilla Thyni[5], Gilberto Tognoni[6]

[1]*Departament of Psychology and Sociology, Universidad de Zaragoza, C/Ciudad Escolar s/n, 44003, Teruel, Spain (lucalba@unizar.es)*

[2]*Faculty of Psychology. Universidad Católica de Valencia, Valencia, Spain*

[3]*Dirección General de Tráfico, Madrid, Spain.*

[4]*Rijkswaterstaat, Delft, The Netherlands*

[5]*Trafikverket, Stockholm, Sweden*

[6]*SINA S.p.A., Verona, Italy*

Abstract

This article explores the most appropriate arrangement (vertical, horizontal) and the frame of reference adopted by drivers (intrinsic, relative) as determinants of the comprehension of new traffic messages (e.g., congestion before arriving to Milan). Two specific cases for location (event-before-city, event-after-city) were tested following two layouts: H (horizontal, left-right) and V (vertical, bottom-up). Four comprehension tests carried out between 2006 and 2013 with 10,099 drivers in four countries (Italy, Netherlands, Spain, Sweden) were analyzed in a 2 (case: Before vs. After) x 2 (disposition: H, V) x 4 (Country) between-subject design. The comprehension of the V variants (78.1%) exceeded the comprehension of the H variants (54.1%) in all the countries in the "before" case. In no country did the V or H variants come close to functional understanding in the "after" case. The results provided evidence of the preferred model and relative frame of reference as determinants of message understanding.

Keywords: frame of reference, preferred mental models, reasoning, traffic sign comprehension

General Audience Summary

A controversial aspect that arises from the use of different traffic signaling devices is that drivers often have to understand messages they are seeing for the very first time. This paper analyzes the results of a series of empirical studies carried out with the aim of internationalizing variable message signs (VMS) by substituting key words (e.g., prepositions) for abstract graphic signs (e.g., an arrow). Faced with novel

elements in a traffic message about which drivers must conclude something in real time, they have no choice but to reason. This article explores the most appropriate arrangement (vertical, horizontal) and the frame of reference adopted by drivers (intrinsic, relative) as determinants of the comprehension of novel and complex VMS (e.g., congestion before arriving to Milan). Our study focuses on the design variants tested to inform drivers about two cases for location (event-before-city and event-after-city), following two basic layouts: H (horizontal, left-right) and V (vertical, bottom-up). Four comprehension tests carried out between 2006 and 2013 with 10,099 drivers in four countries (Italy, Netherlands, Spain, Sweden) were analyzed in a 2 (case: Before vs. After) x 2 (disposition: H, V) x 4 (Country) between-subject design. The comprehension of the V variants (78.1%) exceeded the comprehension of the H variants (54.1%) in all the countries in the “before” case. However, in no country did the V or H variants come close to functional understanding in the "after" case. The results provided evidence of the preferred model and relative frame of reference as determinants of message understanding. Although it is not realistic to expect national or international drivers to memorize all possible traffic messages, it is feasible to understand how their prior knowledge and preferences modulate their conclusions to design more functional traffic messages.

On drivers’ reasoning about traffic signs: the case of qualitative location

Simple painted road signs are rightly assumed to be part of the driver's long-term memory (Ben-Bassat, 2019; Crundall & Underwood, 2001). But as road signs become complex, our understanding of the cognitive processes involved also becomes complex. Some researchers then adopt a pragmatic perspective, exploring the particular demographics of drivers who understand certain messages (Ben-Bassat & Shinar, 2015; Ng & Chan, 2008) and obey them (Ben-Elia & Shiftan, 2010), or proposing the adoption of text messages (Roca, Insa & Tejero, 2018; Shinar & Vogelzang, 2013). Focusing on the ergonomic principles of traffic signs (Ben-Bassat, 2019; Ben-Bassat & Shinar, 2006; Jamson & Mrozek, 2017) recent studies point to a relevant fact: faced with a complex or novel situation (e.g., road signs placed in an ambiguous traffic context), long-term memory may not be sufficient, so drivers need to reason before concluding on their meaning (Castro, Moreno-Ríos, Tomay & Vargas, 2008; Vargas, Moreno-Ríos, Castro & Underwood, 2011).

This paper addresses the evaluation of complex electronic traffic signs, providing more empirical data to identify suitable formulas for international signage. More specifically, this article analyzes the results of comprehension tests carried out with Variable Message Signs (VMS) designed to inform European drivers about variable events (congestion, roadworks, wind, or snow) located qualitatively, that is, by reference to a city placed before or after such events. One of the basic difficulties of this goal arose with the VMS template that many European countries adopted in the 1980s and 1990s, originally designed to combine a pictogram with words and phrases from a certain language (Italian, French, Swedish, etc.) (COST30 BIS, 1985; Ellenberg & Fabre, 1995). In the specific case under study, the goal was to replace a preposition (before, after) with a language-independent element that could be displayed in a VMS. Due to its simplicity, the possibility of adapting it to a 5x7 pixel matrix (see Fig. 1) and its versatility, the arrow was the most obvious choice. Arrows are 'meaningful graphic forms' that encourage people to interpret causal and functional aspects in a diagram (Tversky, 2005; Tversky, Zacks, Lee & Heiser, 2000), capturing a large variety of semantics with their simple shape.

Clearly, the contextual versatility of arrows was both a strength and a weakness. Drivers infer on the fly the possible meanings of the arrows from their immediate visual context, taking into account the surrounding elements and their reciprocal congruence (Di Stasi, Megías, Cándido, Mandonado & Catena, 2012), forming an "arrow diagram" (Kurata & Egenhoffer, 2005). For example, by looking at the fourth sign in the upper right corner (Fig. 1), some drivers may infer that congestion occurs after the city, while other drivers may think that it occurs when heading into the city (i.e. before). So this paper answers this basic question: how should the three main elements of the VMS (pictogram, arrow, city name) be combined in the available VMS template for drivers to understand their meaning?

1.1. Drivers' reasoning the way forward

The answer to this question must integrate two determinants of drivers' understanding: novelty and the contextual interpretation of the arrow. As these VMS did not take advantage of the well-learned structure and layout of painted signs, drivers participating in these studies had to infer their meaning. To discuss how drivers reason about VMS we are adopting the preferred mental models theory (PMMT, Ragni, Fangmeier, Webber & Knauff, 2007; Ragni & Knauff, 2013) a theoretical variant of the mental models theory (MMT, Johnson-Laird, 1983; 2006). The PMMT assumptions better fit the objective of this study, a case of spatial relational reasoning with a set of ambiguous premises about which drivers must reach one fundamental conclusion (the VMS meaning). Unlike the MMT, the PMMT states that, in most cases, people "construct just a single, simple and typical model" (Ragni & Knauff, 2013, p. 564), the preferred model, and ignores the rest, unless we explicitly ask them to consider alternatives (requirement that we do not raise). Figure 1 shows the four basic configurations analyzed. We will start by assuming the construction of a simple one-dimensional model from a basic spatial array (e.g., a row of *n* cells) representing the road line. We have to locate an object (LO) by reference to a referent object (RO) and we would usually locate the RO first. However, since the starting point is the construction of an incremental model, we will assume that individuals tend to prefer to change this role in the first premise, placing the LO first (Oberauer & Wilhelm, 2000; Ragni & Knauff, 2013). This yields one preferred model for each premise (Fig. 1). If participants have a preferred initial representation based on the event (LO is placed first), the following predictions can be tested empirically: (1) If the interpretation of the arrow is that it represents the event with respect to the city, then both messages (before/after) should obtain similar comprehension rates regardless of the layout displayed (horizontal, left-right, or vertical, bottom-up). (2) However, previous studies (Ulrich & Maienborn, 2010; Ulrich et al., 2011) suggest that representing the scene would be easier if the event is "before" the city. RO (city) is added after LO (event) and people who read and write from left to right represent elements in this way (Ragni et al., 2007). Conversely, the "after" condition will be more difficult as it requires participants to reverse the sequence to place the tokens (city, then event). (3) Similarly, some studies predict a temporal order of the sequence from behind (past) to front (future) (e.g., Fuhrman & Boroditsky, 2010; Rinaldi, Vecchi, Fantino, Merabet & Cattaneo, 2018). If participants prefer to represent the sequence of events according to the bottom-up timeline, the same preference as in (2) follows for "before", which maintains the time sequence of the event-then-city, but not for "after" which requires reversing the order of placement of the elements.

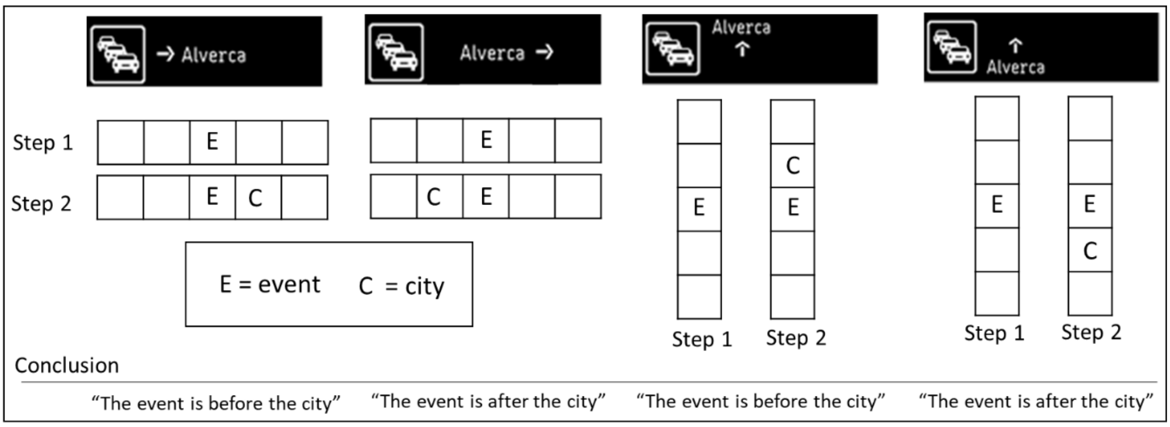


Figure 1. Plausible preferred mental models for "event before/after city"

1.2. The drivers' point of view

A complementary determinant of drivers' reasoning that we explore in this analysis is the potential synergy between the drivers' frame of reference and the VMS frame of reference (Johnson-Laird, 2006; Levinson, 2003). On the one hand, most road signs make sense from a relative frame of reference; the driver is the observer to whom road signs should make sense in the road environment: "[From here where you are] 50 km to arrive at [city]"; "To the Airport take the next exit to [the/your] right", and the like. On the other hand, arrows are asymmetric devices (tail-body-head) that both impact and are nuanced by near elements in diagrams, maps, or panels, configuring an intrinsic frame of reference on the fly (e.g., VMS on Fig. 1). Intrinsic frames of reference involve "an object-centered coordinate system, where the coordinates are determined by the 'inherent features' or facets of the object to be used as the ground or relatum." (Levinson, 2003, p. 41). Figure 2 exemplifies this by showing the variations of the dangerous curve pictogram over 63 years (Krampen, 1983).

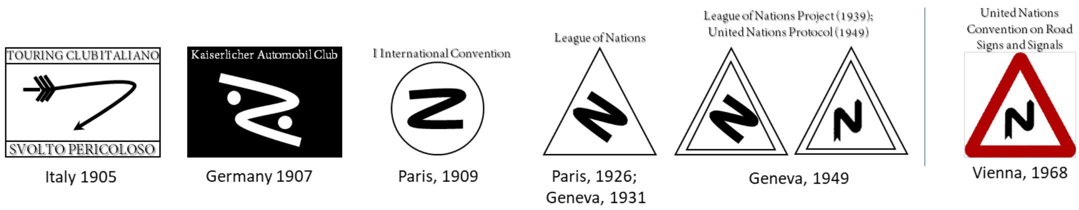


Figure 2. Dangerous curve traffic signs 1905-1968 (redone in digital format by the authors).

Drivers are likely to have a preferred spatial array in traffic (e.g. bottom-up vertical arrangement) in which LO and RO are placed when building a preferred mental model. Our last prediction is (4) that this preference will favor the modeling and understanding of the messages in the vertical layout that adopt a relative frame of reference compared to other designs.

2. Goals of the present work

The aim of this work is to determine how drivers' reasoning strategies and most preeminent frame of reference impinge upon comprehension of novel VMS messages. Recently, Hernando et al. (2022) confirmed predictions 4-6 by comparing the vertical, bottom-up and the horizontal, left-right axis, so now we focus on the vertical axis, including the top-down order. Another important goal is to determine which alternatives are more robust, by testing and verifying how functional the explicit versus generic arrows are in the VMS templates under study. Besides Between 2006 and 2013, a series of comprehension tests were carried out within the framework of EasyWay, the European Union program for the implementation of Intelligent Transport Systems (https://www.its-platform.eu/). Most European national road agencies displayed word-dependent messages in VMS, and the aim of these studies was to explore ways to improve common understanding among (European) drivers. This article analyzes the results of comprehension tests performed with messages designed to report on variable events (congestion, roadworks, wind, or snow) located before or after a city. Studies were carried out in 2006, 2010, 2011, and 2013, with Dutch (NL), Italian (IT), Spanish (ES), and Swedish (SE) samples.

3. Method

3.1. Participants

A total of 10,099 drivers responded to comprehension tests dealing with the variants shown in Fig. 3. The sample included 2938 Dutch, 1988 Italian, 2510 Spanish, and 2663 Swedish drivers, distributed among the different test editions and sign variants (Table 2). The resulting grid included 60 cells (4 countries x 15 signs), with an average number of 168.32 participants per cell (SD = 74.81, MIN =43, MAX = 370). Averaged percentages of main demographics and variables concerning driving experience per country across the four studies (2006-2013) are shown in Table 1. However, the five-level age classification shown in Table 1 has only been adopted since 2010; in 2006, a three-level age classification was used: 18 to 30 years (ES: 45.39; IT: 21.71; NL: 34.39; SE: 28.09; Average: 32.40), 31 to 50 years (ES: 46.64; IT: 63.46; NL: 52.59; SE: 54.14; Average: 54.21), and more than 50 years (ES: 7.97; IT: 14.82; NL: 13.01; SE: 17.77; Average: 13.39). The N.A. (Not Applicable) response was intended for drivers who did not currently drive very much, regardless of how long they have had a driving license. In general, the drivers in our sample were predominantly male, middle-aged (26-45 years), with university studies, long driving experience (> 15 years), and annual mileage (> 20,000 km/year), they drove frequently on motorways, and were familiar with VMS.

Variable	Level	ES	IT	NL	SE	TOTAL
Gender	Male	76.58	75.41	80.48	72.94	76.35
	Female	23.42	24.59	19.52	27.06	23.65
Age	18-25	22.05	7.88	10.95	14.17	13.76
	26-35	36.24	21.79	23.29	24.08	26.35
	36-45	25.99	34.75	23.29	24.30	27.08
	46-55	11.74	22.32	16.38	19.36	17.45

	>55	3.98	13.27	26.11	18.10	15.36
Education	Elementary	9.86	0.46	1.51	5.54	4.34
	Vocational	22.39	34.53	55.71	12.83	31.36
	Secondary	20.34	27.66	16.89	29.72	23.65
	University	47.41	37.36	25.90	51.92	40.65
Driving	N.A.	5.68	1.79	1.70	3.44	3.15
experience	< 5 years	18.39	6.10	8.12	11.32	10.98
	5-15 years	35.44	21.96	29.76	26.56	28.43
	> 15 years	40.50	70.16	60.43	58.68	57.44
Annual	N.A.	6.10	2.10	1.55	4.98	3.68
Mileage	< 10,000 km	18.64	17.01	14.31	26.16	19.03
	10-20,000 km	35.70	37.14	31.70	35.22	34.94
	> 20,000km	39.57	43.76	52.45	33.65	42.36
Motorways	N.A.	2.10	1.30	1.03	2.04	1.62
	Never	2.65	1.59	0.30	2.38	1.73
	Occasionally	16.07	22.92	10.14	12.12	15.31
	Often	79.18	74.21	88.54	83.46	81.35
Familiar	Yes	93.12	94.84	89.80	81.24	89.75
With VMS?	No	6.89	5.16	10.21	18.77	10.25

Table 1. Main demographics and driving experience across countries

3.2. Procedure

The modus operandi of the four studies was very similar, through electronic tests posted on public websites (usually official traffic administrations, but also drivers' associations or highway companies). Participats who accessed these websites in their national languages could see a banner inviting them to collaborate in a set of studies to improve understanding of road signs in Europe. The design and structure of the test followed ISO-9186 (2001; 2007) recommendations for computer tests. Typically, subjects had access to the test simply by clicking on the invitation banner and reading a brief explanation of the context of the test and its purpose (i.e., to check European drivers' understanding of road signs shown in VMS). Participants were then invited to fill in demographics, such as age, gender or driving experience (Table 1), and then read an explanation of the task context: "On each page of this test, there is a variable message sign (VMS). Look at each VMS and write in the box below it what you think that variable message sign means. Write ‘I don't know’ if you cannot assign a meaning to the VMS. An example is given on the next page”. Participants then read “This is an example. Context: on a motorway or dual carriageway” and then saw a sign showing a standard traffic situation (e.g., “caution, road works”) for about 8 s. Below, this sentence was shown “What do you think this variable message sign means?” and then a self-informed response followed: “I am driving towards a dangerous road section due to road works” (the pictogram in the example was not displayed later in the test). Clicking on the “continue” box led to the first stimulus of the set. Drivers were required to pay attention to the sign first, then type in their response. Participants were invited to respond to a set of 8 consecutive signs (each study included 6 sets, totaling 48 signs, most of them not considered here) and took an average of 567.7 s (SD = 1003.6), about 10 minutes, to perform the whole task.

The document ISO-9186 (2001; 2007) establishes that the answers of the participants must be assigned independently by the judges who compare them with one of seven basic categories: 1) the correct understanding of the sign is true (the judge estimates that the probability of a correct understanding is greater than 80%), 2) the correct understanding of the sign is very likely (between 66 and 80%), 3) the correct understanding of the sign is likely (between 50 and 60%), 4) the understanding is the opposite of what is expected, 5) any other answer, 6) the answer "I don't know", 7) no answer. In our studies, this basic structure was assumed, although with some modifications. On the one hand, only the safe understanding of the sign (that is, the phrase that describes the meaning of the message or a very similar one) was considered correct (coded as category 1). On the other hand, along with categories 4-7, different categories of misunderstanding were also considered, reflecting alternative possibilities in the context of locating variable events (Fig. 5). The objective of this adaptation in the correction procedure was to learn from the different types of incorrect answers to improve the design in successive studies. All the countries involved assumed the same correction criteria and the same response categories.

Instruments adopted to present the signs differed between the first three studies (2006; 2010; 2011) and the 2013 study. The first three studies presented the messages to participant drivers in a static fashion. In the 2013 study, a web-based driving simulator was used to display the messages. Participants were not asked (and could not) drive. Using the simulator only meant that the message presentation was dynamic. The basic scenario placed the driver in a car moving towards a VMS gantry in the right lane of a two-lane motorway (drivers could not select the speed nor change lane or direction). A reading window of about 8 s was set as the driver approached the VMS at 90 km/h. Therefore, the time available for participants to read the message was set to be the same for the two types of presentation (static, simulator). The driver passed under the gantry (simulator), or the message was removed from the screen (static), and then the same basic question "What do you think this variable message sign means?" appeared on screen. Then a writing box appeared allowing participants to type in their answers (without time restriction).

3.3. Materials

The VMS template adopted in all studies (a 32x32-pixel matrix plus three rows with 12 to 16 5x7-pixel matrices) is shown in Fig. 3 (bottom-right). In line with the objective of internationalizing the messages, a basic rule followed in the studies was that all the elements displayed in the VMS must be known (e.g., the congestion pictogram), understandable (e.g., an arrow) or easily inferred (a city name). Other words (e.g., prepositions, conjunctions, and the like) were not allowed. Some tests performed in parallel in 10 European countries (ES, IT, NL and SE included) determined that these danger warning pictograms used in these studies obtained average comprehension rates above 89.8% with or without the red triangle (Lucas-Alba et al., 2011). Showing the same silhouette without the red triangle prompted significantly fewer mentions of the word “danger” or “caution”, and allowed extra space for the pictogram. This makes sense when the event is far away (i.e., not so dangerous), the most prevalent case for qualitative location (e.g., German dWiSta panels show a congestion silhouette without triangle, Hartz & Schmidt, 2005). Therefore, after 2010, most tests displayed messages without the red triangle. On the other hand, given that some VMS had 8x11 pixel matrices, being able to show lowercase, this possibility was later included in our stimuli. The city names were the same for all participants in the 2006 test, using place names belonging to a non-participating third country (Germany and Austria). However, local examples were introduced to rule out potential problems with unknown place names (although later results would confirm that this was not the case, see table 2), therefore in the 2010, 2011 and 2013 studies all countries used the same message templates but with city names in their territories. Last but not least, exploring the qualitative location of variable events also had practical implications. Most road operators in different countries avoid reporting changing events such as congestion, wind, or fog by using a quantitative location (e.g., "congestion in 15 km"), especially if there is only one VMS available before the critical road section (Arbaiza & Lucas-Alba, 2012).

The first attempts to identify an international location set for VMS (2006, 2010) explored the basic top-down, left to right horizontal (H) layout (the standard parsing for official languages in Europe; see Bergen & Chan, 2005; Spalek & Hammad, 2005). The four "before" variants tested are shown in Fig. 3 (first row): horizontal-dangerous congestion, horizontal-congestion, horizontal-wind, and horizontal-snow. Only "after" variant tested (Fig. 3).

A complementary approach was explored in the 2011 and 2013 tests: the vertical (V), bottom-up layout. The four "before" V variants tested are shown in Fig. 3: vertical-congestion, vertical-roadworks, vertical-wind, and vertical-snow. The six "after" V variants tested are shown also in Fig. 3: vertical-congestion12 (12=in the *first and second rows*), vertical-congestion23 (23=in the *second and third rows*), vertical-roadworks23, vertical-wind23, vertical-congestion(e) (city name *enclosed*), and vertical-congestion-sa (*special arrow*).

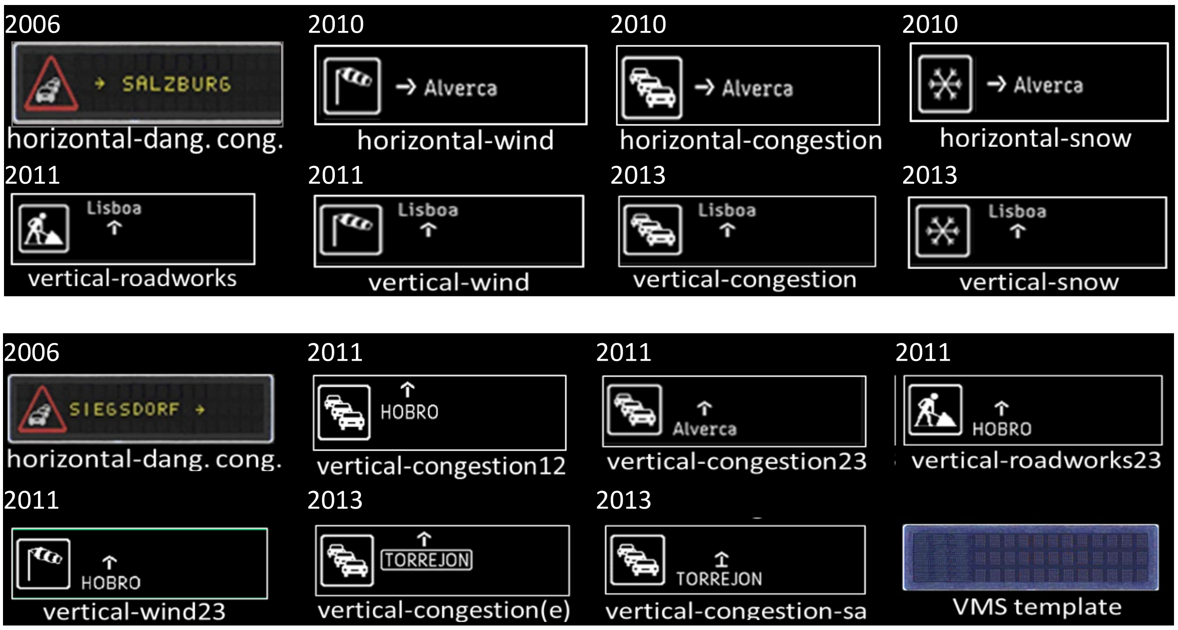


Figure 3. Horizontal and Vertical layout design variants for *before* and *after* explored in 2006-2013.

4. Results

4.1. Descriptive statistics

Table 2 presents the weighed average comprehension rate (1 = 100% comprehension) of the eight "before" and the seven "after" messages per country, with the corresponding Horizontal and Vertical layout variants. Pictogram variation, however, was not part of our analysis and was only meant to explore responses to qualitative location under differing events. In bold total comprehension rates per country under Horizontal and Vertical layout formats.

Case	layout (H: horizontal, V: vertical)	ES	IT	NL	SE
BEFORE	H, dangerous congestion	*.568*	*.665*	*.785*	*.506*
(n=5076)	H, wind	*.277*	*.419*	*.797*	*.306*
	H, congestion	*.202*	*.233*	*.811*	*.425*
	H, snow	*.213*	*.395*	*.777*	*.351*
	Total H, before (n=2253)	**.422**	**.535**	**.792**	**.415**
	V, snow	*.879*	*.880*	*.853*	*.733*
	V, road work	*.641*	*.901*	*.921*	*.888*
	V, wind	*.812*	*.842*	*.845*	*.712*
	V, congestion	*.597*	*.766*	*.726*	*.764*
	Total V, before (n=2823)	**.700**	**.833**	**.817**	**.775**
AFTER	H, dangerous congestion	*.069*	*.284*	*.450*	*.156*
(n=5023)	**Total H, after (n=476)**	**.069**	**.284**	**.450**	**.156**
	V, congestion-sa	*.555*	*.366*	*.403*	*.508*
	V, road work23	*.387*	*.374*	*.298*	*.374*
	V, wind23	*.257*	*.262*	*.216*	*.347*
	V, congestion(e)	*.024*	*.140*	*.067*	*.042*
	V, congestion23	*.033*	*.100*	*.019*	*.080*
	V, congestion12	*.035*	*.081*	*.000*	*.071*
	Total V, after (n=4547)	**.216**	**.213**	**.151**	**.222**

Table 2. Weighed average comprehension rates of "before" and "after" Horizontal and Vertical layout variants per country.

4.2. Inferential statistics

A between-subject ANOVA was carried on for a 2 (Case, before/after) x 2 (Disposition, horizontal/vertical) x 4 (Country, ES / IT / NL / SE) design. Overall, the "before" messages (M = .661) yielded better comprehension rates than the "after" messages (M = .220), $F_{(1,\ 10083)}$ = 1296.06, p = .0001,

$\eta_p^2 = .114$ (Table 2). Also the V variants obtained better comprehension rates ($M = .491$) than the H variants ($M = .390$), $F_{(1,\ 10083)} = 67.49$, $p = .0001$, $\eta_p^2 = .007$. Both factors yielded a significant interaction, $F_{(1,\ 10083)} = 129.91$, $p = .0001$, $\eta_p^2 = .013$. Comprehension rates for the V ($M = .201$) and H variants ($M = .240$) were similar ($F_{(1,\ 10083)} = 3.36$, $p = .067$, $\eta_p^2 = .000$) when the event was located after the city; however, when the event was located before the city, the comprehension of the V dispositions (M = .781) was significantly better than that of the H dispositions (M = .541; $F_{(1,\ 10083)} = 391.17$, $p = .001$, $\eta_p^2 = .037$).

Comprehension also differed in terms of the country of origin, $F_{(3,\ 10083)} = 59.14$, $p = .0001$, $\eta_p^2 = .017$. Bonferroni post hoc comparisons showed a linear order: ES ($M = .352$), then SE ($M = .392$), IT ($M = .466$), and NL ($M = .553$; all differences were significant at $p < .05$). Country presented an interaction with Case, $F_{(3,\ 1083)} = 3.48$, $p = .015$, $\eta_p^2 = .001$: comprehension for "before" was significantly higher than comprehension for "after" in every country (ES, M (before) = .561, M (after) = .143; $F_{(1,\ 10083)} = 409.27$, $p = .0001$, $\eta_p^2 = .039$; IT, M (before) = .684, M (after) = .248; $F_{(1,\ 10083)} = 210.86$, $p = .0001$, $\eta_p^2 = .020$; NL, M (before) = .805, M (after) = .300; $F_{(1,\ 10083)} = 433.25$, $p = .0001$, $\eta_p^2 = .041$; SE, M (before) = .595, M (after) = .189; $F_{(1,\ 10083)} = 338.81$, $p = .0001$, $\eta_p^2 = .033$). Although IT and NL fared a bit better than the rest (Table 2), the average worse (ES: $M = .143$) and better (NL: $M = .300$) comprehension rates for "after" differed little yet significantly ($F_{(1,\ 10083)} = 12.12$, $p = .0001$, $\eta_p^2 = .004$). Comprehension rates for "before" were significantly better, but countries differed significantly from each other (except for SE and ES, $p = .209$): the lowest rate (ES: $M = .561$) was comparatively lower than the highest one (NL: $M = .805$: $F_{(1,\ 10083)} = 94.23$, p = .0001, $\eta_p^2 = .027$). In all cases, Bonferroni adjustment was applied for multiple comparisons. Country also presented an interaction with Disposition, $F_{(3,\ 10083)} = 50.08$, $p = .0001$, $\eta_p^2 = .015$. All countries achieved better averaged comprehension results for V than for H (ES, M (V) = .458, M (H) = .245; $F_{(1,\ 10083)} = 105.683$, $p = .0001$, $\eta_p^2 = .010$; IT, M (V) = .523, M (H) = .409; $F_{(1,\ 10083)} = 14.371$, $p = .0001$, $\eta_p^2 = .001$; SE, M (V) = .499, M (H) = .285; $F_{(1,\ 10083)} = 93.312$, $p = .0001$, $\eta_p^2 = .009$), except for NL (V: M = .484; H: M = .621; $F_{(1,\ 10083)} = 31.95$, $p = .0001$, $\eta_p^2 = .003$; Table 2). Finally, the Case, Disposition and Country factors yielded a significant interaction, $F_{(3,\ 10083)} = 5.12$, $p = .002$, $\eta_p^2 = .002$. Participants from all countries reached high comprehension rates with the V set in the "before" case (ES, M (V) = .700, M (H) = .422; $F_{(1,\ 10083)} = 141.58$, $p = .0001$, $\eta_p^2 = .014$; IT, M (V) = .833, M (H) = .535, $F_{(1,\ 10083)} = 107.18$, $p = .0001$, $\eta_p^2 = .011$; SE, M (V) = .775, M (H) = .415, $F_{(1,\ 10083)} = 254.66$, $p = .0001$, $\eta_p^2 = .025$); except for Dutch drivers who obtained sufficient comprehension rates in the "before" case in both V ($M = .817$) and H ($M = .792$; $F_{(1,\ 10083)} = 1.34$, $p = .248$, $\eta_p^2 = .000$) cases (Fig. 4).

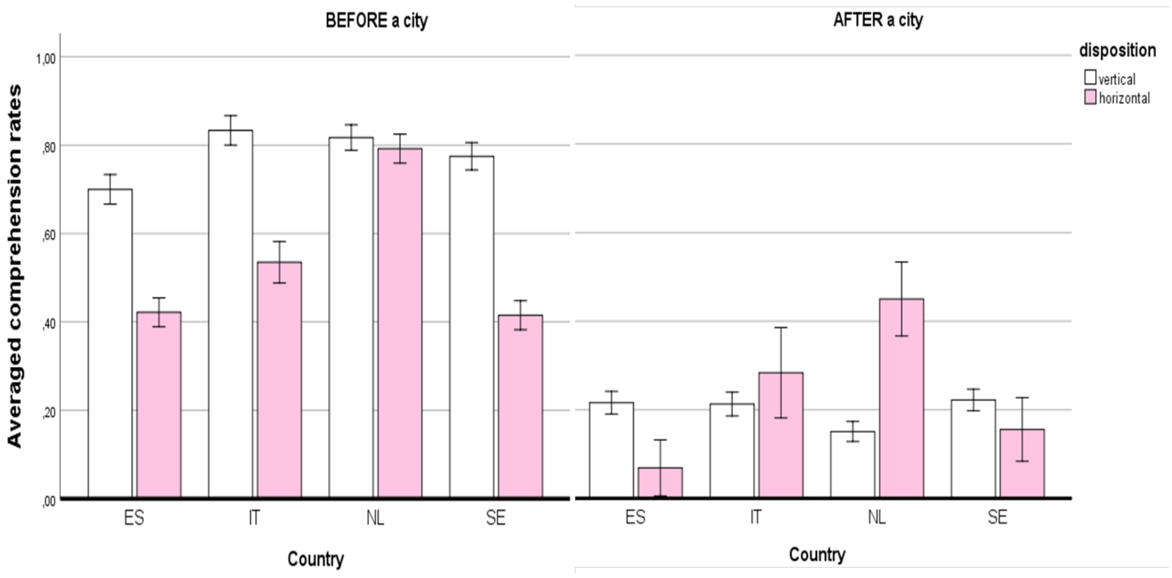
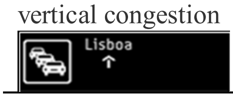


Figure 4. Event-before-city (left) and event-after-city (right): comprehension rates per country for V and H dispositions.

4.3. Qualitative analysis: most frequent errors in answers to the "before" case.

The representations that emerged from drivers' correct and incorrect responses in the "before" case are depicted in Fig. 5. We will examine the incorrect answers qualitatively, assuming a rule of thumb, that is, considering that the answers that accumulate at least 5% of answers are sufficiently representative. We will focus first on the less fortunate horizontal (H) variants (weighed averages). The vertical (V) variants constrained the inferential mechanism quantitatively (fewer incorrect answers), and qualitatively: the most frequent wrong answer was E (Fig. 5), hardly present with H variants. Tables 3 and 4 show the descriptive results, incorrect answers and their representation in Fig. 5.

Sign (event-*before*-city)	Incorrect answers	M (%)	S.D.	Fig. 5.
horizontal dang. cong. 	Congestion at the exit to Salzburg	10.9	6.3	A
	Departure/recommended direction to Salzburg	9.2	5.6	B
horizontal congestion 	Congestion at the exit to [city]	21.5	16.7	A
	Route recommend by the following exit to avoid congestion	12.3	10.9	C
	Recommended exit/direction to [city]	10.1	2.6	B
	Congestion in [city]	7.5	3.5	D
horizontal wind 	Wind in [city]	31.5	11.3	D
	Wind at the exit of [city]	11.1	6.7	A
	Danger (no location mentioned)	9.0	11.5	E
horizontal snow 	Snow in [city]	35.7	12.6	D
	Snow at the exit to [city]	8.8	7.5	A
	Danger	18.0	12.6	E
vertical congestion 	Specific recommendation to [city]	6.0	5.2	B
	Congestion (no location mentioned)	5.9	4.3	E

vertical roadwork 	Roadwork (no location mentioned)	4.7	4.5	E
vertical wind 	Wind (no location mentioned)	8.7	6.1	E
vertical snow 	Snow (no location mentioned)	8.1	6.8	E

Table 3. Descriptive results for incorrect answers to the 'before' cases.

4.4. Qualitative analysis: most frequent errors in answers to the "after" case.

The representations of drivers' correct and incorrect answers in the "after" case are depicted in Fig. 5. Again, only sufficiently representative incorrect responses (≥ 5%) are examined. We will focus first on the horizontal (H) variant (weighed averages), which presented a rich panoply of wrong answers. Bottom-up vertical (V) variants were also unsuccessful, although to a varying degree (see table 4).

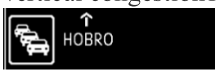
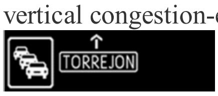
Sign (event-*after*-city)	Incorrect answers	M (%)	S.D.	Fig. 5
horizontal dang. cong-	Congestion (no location mentioned)	24.4	11.2	E
	Congestion at the entrance to Siegsdorf	13.7	13.4	D
	Alternative exit to Siegsdorf available	12.1	11.1	B
	Congestion before Siegsdorf	9.6	6.8	G
	Alternative exit to Siegsdorf congested	8.1	5.6	A
vertical congestion12	Congestion before [city]	59.7	15.1	G
	Go straight to get to [city]	9.7	3.5	F
	Congestion (no location mentioned)	6.3	1.8	E
	Congestion at the exit from the [city]	4.7	3.2	A
vertical congestion 23	Congestion before [city]	55.5	18.5	G
	Go straight to get to [city]	8.4	7.3	F
	Congestion (no location mentioned)	7.7	3.3	E
	Congestion at the exit form the [city]	5.5	3.8	A
vertical wind23	Wind before [city]	52.6	12.8	G
	Wind (no location mentioned)	12.3	4.3	E
	Wind near of [city]	4.7	2.2	G
vertical roadworks 23	Works before [city]	42.9	10.8	G
	Works near [city]	6.6	2.7	D
	Road works (no location mentioned)	5.8	3.9	E
vertical congestion-e	Congestion going to [city]	56.3	10.6	G
	Congestion at the exit from [city]	14.7	4.8	A
	Congestion (no location mentioned)	7.8	4.4	E
	Rerouting recommendation formula	6.0	2.7	B
vertical congestion-sa	Congestion going to [city]	35.5	10.4	G
	Congestion at the exit from [city]	9.6	1.9	A

Table 4. Descriptive results for incorrect answers to the 'after cases.

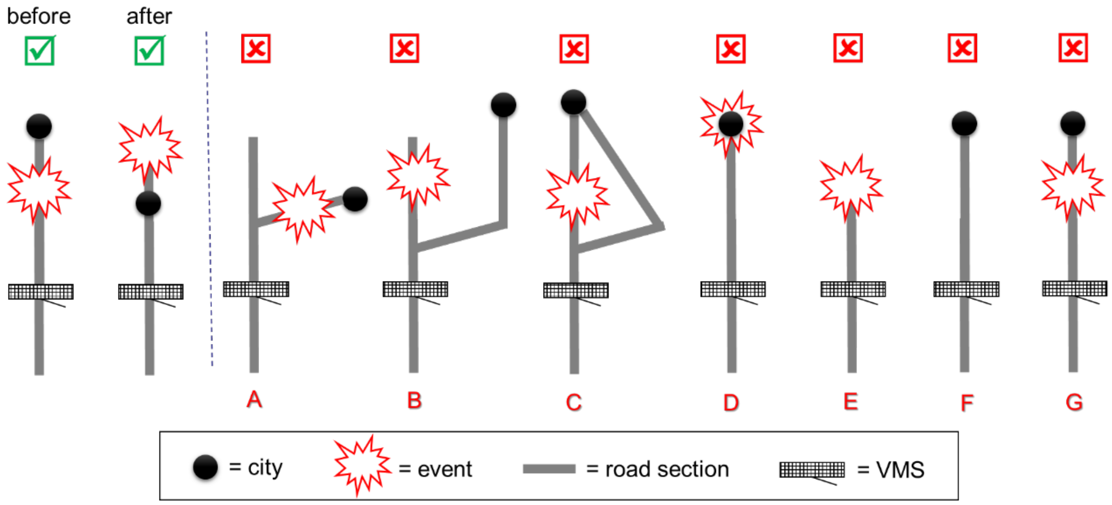


Figure 5. Correct vs. incorrect model representations after participants' answers.

5. Discussion

Clearly, not the first, but the second and third predictions were confirmed. Results for "before" (above 50%) were better than for "after" (around 20%). However, while H vs V did not differ with "after", V yielded better results with "before" (confirming the fourth prediction), while H yielded better results with before than V with after. Also, the "before" and "after" errors differed in nature: errors with before were less frequent and mostly involved incomplete information (e.g., indicating there was a congestion, but not where). Errors with after were more frequent and problematic because people understood the opposite of the intended meaning (e.g., Fig. 5 G). The obvious recommendation is to avoid using "after" information with the event-city-arrow triad: most drivers will understand otherwise (except about 20% who will be correct).

According to the PMMT, representing the scene would be easier if the event is before the city, because the after condition requires participants to reverse the LO-RO sequence to place the tokens. The second prediction aligns with this statement (see Jahn, Knauf & Johnson-Laird, 2007, Ulrich & Maienborn, 2010; Ulrich et al., 2011). However, participants interpreted more correctly the V (predictions 3 and 4) than the H set (prediction 2). The third prediction aligns with findings on embodied cognition that expect a temporal order of the sequence from behind (past) to front (future) (Fuhrman & Boroditsky, 2010; Rinaldi et al., 2018), favoring the adoption of a PMM based on a bottom-up timeline. The location of objects in the front-back dimension is fast, because it maintains a bodily and functional asymmetry that is not so marked in the left-right dimension (De Vega, 2002; Franklin & Tversky, 1990). Generic metaphors in language also present the future in front of the Ego (Núñez & Sweetser, 2006). Finally, the results with V "before" fit with Posner's (1980) function of orientation and anterior attention networks (Posner & Dehaene, 1994).

A better comprehension of V “before”, compared to H “before”, also fits with the fourth prediction: VMS consistent with the drivers' relative frame of reference (Levinson, 2003; Johnson-Laird, 2006; Fig. 2) facilitated model representation. Most drivers did not adopt the intrinsic diagrams perspective: an arrow pointing right was seen as a compulsory, advised, or recommended exit or diversion route (Fig. 5 A-C), and placing an upward-pointing arrow above a city name was interpreted as the next location on route (Fig. 5 D, G). The arrow pointing up departs from an empty space that drivers may identify with their own position: if we fold that plane forward 90°, it will coincide with the projection of the driving plane (road), where down is near (driver, base of the arrow) and up is far (location, tip of the arrow). This rotation is common in traffic signs. However, drivers would need to make a 90° turn counterclockwise and then a 90° turn forward to model "before" from a relative frame of reference. But interpreting the arrow as "exit" (Fig. 5, A-C) would fit directly into the relative frame of reference, and a frontal rotation of 90° would yield a plausible iconic model for the sign. Concluding that the arrow represented an exit could reflect a modulation of former knowledge (Johnson-Laird & Byrne, 2002), and would require less cognitive effort.

Regarding prior knowledge, most drivers would not expect information about open-ended events that occur somewhere after a section of the road (the arrow is pointing to "nothing" in the after condition-something a little odd). Classical experiments on indeterminate location (Bransford, Barclay, & Franks, 1972; Ehrlich & Johnson-Laird, 1982; Mani & Johnson-Laird, 1982) would predict drivers' reluctance to delve into this possibility as well (indeterminate locations block the elaboration of iconic mental models). Considering the deictic character of the arrow (Eco, 1996) drivers likely converted "after" into "towards there", a very successful wrong answer, hovering around 50% (Fig. 5-G).

All in all, the V-messages “before” successfully restricted the models of possibilities, favoring the preferred model, and achieved acceptable comprehension ratings from drivers from four different countries. Predictions 3 and 4 present a number of interesting theoretical synergies that could be the subject of future research. One last note to address the good performance of the Dutch drivers in H "before". Although the VMSs studied were not part of the Dutch catalog of signs, Dutch drivers were used to reading relatively complex messages with small arrows interspersed in the text and with small pictograms (Fig. 6), a not so common practice in the rest of the participating countries (Blanch, Lucas-Alba & Messina, 2011). This may have contributed to more flexible representational mechanisms both for the V and H variants under the “before” case. Clearly, this speculation would require further study.

Fig. 6. Dynamic Route Information Panels (DRIP) in The Netherlands (after Blanch et al., 2011).

5.1. Limitations, extensions and future work

Future studies should make an additional effort controlling variability. The changes in VMS and display formats tried to improve studies with large international samples. Message presentation (static in 2006-2011 vs. web-based simulator in 2013) provided the same reading window (8 s), and the simulator did not allow drivers to manipulate lane change or speed. Although studies focused on location formulations and pictograms enjoyed similarly high comprehension rates (Lucas-Alba et al., 2011), variants were unevenly distributed across studies. None of these factors above showed a statistically significant effect on comprehension, but all these methodological issues should be considered as a limitation to be overcome in further studies. Finally, not all samples were the same and not all signs had the same number of responses. However, all countries provided data from large samples that included participants of different characteristics (sex, age, driving experience, education). Future studies could benefit from greater methodological and procedural control, as do laboratory studies with smaller samples.

Author Contributions

Acknowledgments

We thank the reviewers for their comments, which really helped improve this article. We thank Professor Cándida Castro (University of Granada) and Dr. Tamar Ben-Bassat (Tamar Graphic Design) for helping us improving the paper with their comments. We thank Federico Fernández (then DGT traffic deputy director) and Professor Luis Montoro (University of Valencia) for supporting the research on VMS harmonization. Support came from Dirección General de Tráfico, Spain (grant SPIP 2014-01345).

References

Arbaiza, A., & Lucas-Alba, A. (Coords.). (2012). *VMS-DG01. Principles of VMS Message Design. Version 2.0, December 2012*. http://www.easyway-its.eu/

Ben-Bassat, T. (2019). Are ergonomically designed road signs more easily learned?. *Applied ergonomics*, *78*, 137-147. https://doi.org/10.1016/j.apergo.2019.02.009

Ben-Bassat, T., & Shinar, D. (2006). Ergonomic guidelines for traffic sign design increase sign comprehension. *Human factors*, *48*(1), 182-195. https://doi.org/10.1518/001872006776412298

Ben-Bassat, T., & Shinar, D. (2015). The effect of context and drivers' age on highway traffic signs comprehension. *Transportation research part F: traffic psychology and behaviour*, *33*, 117-127. https://doi.org/10.1016/j.trf.2015.07.009

Ben-Elia, E., & Shiftan, Y. (2010). Which road do I take? A learning-based model of route-choice behavior with real-time information. *Transportation Research Part A: Policy and Practice*, *44*(4), 249-264. https://doi.org/10.1016/j.tra.2010.01.007

Bergen, B., & Chan, T. (2005). Writing Direction Influences Spatial Cognition. Proceedings of the Annual Meeting of the Cognitive Science Society, 27. Retrieved from https://escholarship.org/uc/item/4tt0k00j

Blanch, Mª. T., Lucas-Alba, A., & Messina, C. (2011). *ESG-4 Mare Nostrum: The Working Book.* Madrid: Dirección General de Tráfico.

Bransford, J.D., Barclay, J.R., & Franks, J.J. (1972). Sentence memory: A constructive versus interpretive approach. *Cognitive Psychology*, *3* (2), 193-209. https://doi.org/10.1016/0010-0285(72)90003-5

Carroll, D. W. (2008). *Psychology of Language (5th Ed).* Thomson Wadsworth.

Castro, C., Moreno-Ríos, S., Tornay, F., & Vargas, C. (2008). Mental representations of obligatory and prohibitory traffic signs. *Acta psychologica*, *129*(1), 8-17. https://doi.org/10.1016/j.actpsy.2008.03.016

COST 30 bis. (1985). *Electronic traffic aids on major roads.* Luxembourg: Commission of the European Communities.

Crundall, D., & Underwood, G. (2001). The priming function of road signs. *Transportation Research Part F: Traffic Psychology and Behaviour*, *4*(3), 187–200. https://doi.org/10.1016/S1369-8478(01)00023-7

De Vega, M. (2002). Del significado simbólico al significado corpóreo. *Estudios de Psicología*, *23*(2), 153-174. https://doi.org/10.1174/02109390260050012

Di Stasi, L. L., Megías, A., Cándido, A., Maldonado, A., & Catena, A. (2012). Congruent visual information improves traffic signage. *Transportation research part F: traffic psychology and behaviour*, *15*(4), 438-444. https://doi.org/10.1016/j.trf.2012.03.006

Eco, U. (1996). *Tratado de semiótica general*. Barcelona: Lumen.

Ehrlich, K., & Johnson-Laird, P.N. (1982). Spatial descriptions and referential continuity. *Journal of Verbal Learning and Verbal Behavior*, *21* (3), 296-306. https://doi.org/10.1016/S0022-5371(82)90626-0

Ellenberg, M., & Fabre, P. (1995). Melyssa Corridor. Mediterranean Lyon-Sttutgart Site for ATT. Transport Telematics Project V2040. D10. Commission of the European Communities - R&D Programme Telematics Systems in the Area of Transport (DRIVE II).

Franklin, N., & Tversky, B. (1990). Searching imagined environments. *Journal of Experimental Psychology: General*, *119*, 63-76. https://doi.org/10.1016/j.trf.2012.03.006

Fuhrman, O., & Boroditsky, L. (2010). Cross-cultural differences in mental representations of time: Evidence from an implicit Nonlinguistic task. Cognitive Science, 34(8), 1430-1451. https://doi.org/10.1111/j.1551-6709.2010.01105.x

Hartz, B., & Schmidt, M. (2005). *Dynamische Wegweiser mit integrierten Stauinformationen (dWiSta)*. Forschung Strassenbau und Strassenverkehrstechnik. Heft 924 (2005). Bundesministerium für Verkehr, Bau- und Wohnungswesen, Abteilung Strassenbau, Strassenverkehr. Bonn, Germany.

ISO-9186 (2001). Graphical symbols: test methods—part 1: methods for testing comprehensibility. Geneva: International Organization for Standardization.

ISO-9186 (2007). Graphical symbols: test methods—part 1: methods for testing comprehensibility. Geneva: International Organization for Standardization.

Jahn, G., Knauff, M., & Johnson-Laird, P. N. (2007). Preferred mental models in reasoning about spatial relations. *Memory & cognition*, *35*(8), 2075-2087.

Jamson, S., & Mrozek, M. (2017). Is three the magic number? The role of ergonomic principles in cross country comprehension of road traffic signs. *Ergonomics*, *60*(7), 1024-1031. https://doi.org/10.1080/00140139.2016.1245874

Johnson-Laird, P. N. (1983). *Mental models. Towards a cognitive science of language, inference and consciousness*. Harvard University Press.

Johnson-Laird, P. N. (2006). *How we reason*. Oxford University Press.

Johnson-Laird, P. N., & Byrne, R. M. (2002). Conditionals: a theory of meaning, pragmatics, and inference. *Psychological review*, *109*(4), 646. https://doi.org/10.1037/0033-295X.109.4.646

Krampen, M. 1983. Icons on the road. *Semiotica*, *43* (1/2), 1-203.

Kurata, Y., & Egenhofer, M. J. (2005). Structure and Semantics of Arrow Diagrams. In: Cohn A.G., Mark D.M. (eds) Spatial Information Theory. COSIT 2005. Lecture Notes in Computer Science, 3693. Springer, Berlin, Heidelberg. https://doi.org/10.1007/11556114_15

Levinson, S. (2003). *Space in Language and Cognition: Explorations in Cognitive Diversity*. Cambridge University Press.

Lucas-Alba, A., Arbaiza, A., Blanch, Mª.T., Desnouailles, Ch., Evans, D., Kjemtrup, K., Martins, R., Papoutsitakis, I., Remeijn, H., Thyni, G., Tognoni, G., & Zorin, U. (2011). "Upgrading cultural infrastructures: Notes on the strategy concerning variable message signs on the European scene," 2011 14th International IEEE Conference on Intelligent Transportation Systems (ITSC), Washington, DC, 2011, 1886-1891, https://doir.org/10.1109/ITSC.2011.6083065.

Mani, K., & Johnson-Laird, P.N. (1982). The mental representation of spatial descriptions. *Memory & Cognition* 10, 181–187. https://doi.org/10.3758/BF03209220

Ng, A. W., & Chan, A. H. (2008). The effects of driver factors and sign design features on the comprehensibility of traffic signs. *Journal of safety research*, *39*(3), 321-328. https://doi.org/10.1016/j.jsr.2008.02.031

Núñez, R. E., & Sweetser, E. (2006). With the future behind them: Convergent evidence from Aymara language and gesture in the crosslinguistic comparison of spatial construals of time. *Cognitive science*, *30*(3), 401-450. https://doi.org/10.1207/s15516709cog0000_62

Oberauer, K., & Wilhelm, O. (2000). Effects of directionality in deductive reasoning: I. The comprehension of single relational premises. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(6), 1702. https://doi.org/10.1037/0278-7393.26.6.1702

Posner, M. I. (1980). Orienting of attention. *Quarterly journal of experimental psychology*, *32*(1), 3-25. https://doi.org/10.1080/00335558008248231

Posner, M. I., & Dehaene, S. (1994). Attentional networks. *Trends in neurosciences*, *17*(2), 75-79. https://doi.org/10.1016/0166-2236(94)90078-7

Ragni, M., Fangmeier, T., Webber, L., & Knauff, M. (2007). Preferred mental models: How and why they are so important in human spatial reasoning. In Barkowsky et al., (Eds): *Spatial Cognition V, LNAI 4387*, 175-190. Heidelberg:Springer-Verlag Berlin.

Ragni, M., & Knauff, M. (2013). A theory and a computational model of spatial reasoning with preferred mental models. *Psychological review*, *120*(3), 561. https://doi.org/10.1037/a0032460

Reinolsmann, N., Alhajyaseen, W., Brijs, T., Ross, V., Timmermans, C., Pirdavani, A., Hussain, Q., & Brijs, K. (2019). Investigating the Impacts of Graphical Route Information Panel Layouts on Drivers' Comprehension and Response Time. *Arabian Journal for Science and Engineering*, *44*(10), 8439-8450. https://doi.org/10.1007/s13369-019-04014-2

Rinaldi, L., Vecchi, T., Fantino, M., Merabet, L. B., & Cattaneo, Z. (2018). The ego-moving metaphor of time relies on visual experience: No representation of time along the sagittal space in the blind. Journal of Experimental Psychology: General, 147(3), 444-450. https://doi.org/10.1037/xge0000373

Roca, J., Insa, B., & Tejero, P. (2018). Legibility of text and pictograms in variable message signs: can single-word messages outperform pictograms? *Human factors*, *60*(3), 384-396. https://doi.org/10.1177/0018720817751623

Spalek, T.M., & Hammad, S. (2005). The Left-to-Right Bias in Inhibition of Return Is Due to the Direction of Reading. *Psychological Science, 16*(1), 15-18. https://doi.org/10.1111/j.0956-7976.2005.00774.x

Shinar, D., & Vogelzang, M. (2013). Comprehension of traffic signs with symbolic versus text displays. *Transportation research part F: traffic psychology and behaviour*, *18*, 72-82. https://doi.org/10.1016/j.trf.2012.12.012

Tversky, B. (2005). Visuospatial Reasoning. In K.J. Holyoak & R.G. Morrison (Eds). *The Cambridge Handbook of Thinking and Reasoning*. Cambridge University Press, 209-240.

Tversky, B., Zacks, J., Lee, P., & Heiser, J. (2000). Lines, Blobs, Crosses and Arrows: Diagrammatic Communication with Schematic Figures. In: Anderson M., Cheng P., & Haarslev, V. (eds). *Theory and Application of Diagrams*., 221-230. Springer. https://doi.org/10.1007/3-540-44590-0_21

Ulrich, R., Eikmeier, V., De la Vega, I., Ruiz Fernández, S., Alex-Ruf, S., & Maienborn, C. (2011). With the past behind and the future ahead: Back-to-front representation of past and future sentences. Memory & Cognition, 40(3), 483-495. https://doi.org/10.3758/s13421-011-0162-4

Ulrich, R., & Maienborn, C. (2010). Left-right coding of past and future in language: The mental timeline during sentence processing. Cognition, 117(2), 126-138. https://doi.org/10.1016/j.cognition.2010.08.001

UNECE (1968/1995). Convention on Road Signs and Signals, E/CONF.56/17/Rev.1/Amend.1. Retrieved from: http://www.unece.org/trans/main/welcwp1.html

Vargas, C., Moreno-Rios, S., Castro, C., & Underwood, G. (2011). Encoding time and signs exposure time in the representation of diagrammatic deontic meanings. *Acta psychologica*, 137(1), 106-114. https://doi.org/10.1016/j.actpsy.2011.03.006