THE UNIVERSITY of EDINBURGH

Edinburgh Research Explorer

The preference for approximation

Citation for published version:
Solt, S, Cummins, C & Palmovi, M 2017, 'The preference for approximation', *International Review of Pragmatics*, vol. 9, no. 1, pp. 1-21. https://doi.org/10.1163/18773109-00901010

Digital Object Identifier (DOI):
10.1163/18773109-00901010

Document Version:
Peer reviewed version

Published In:
International Review of Pragmatics

The Preference for Approximation

Stephanie Solt[1]
ZAS Berlin
Schützenstr. 18
10117 Berlin, Germany
stephanie.solt@gmail.com

Chris Cummins
School of Philosophy, Psychology and Language Sciences
University of Edinburgh
Dugald Stewart Building
3 Charles Street
Edinburgh, EH8 9AD
United Kingdom
c.r.cummins@gmail.com

Marijan Palmović
Department of Speech and Language Pathology
University of Zagreb
University Campus Borongaj
Borongajska cesta 83f
10000 Zagreb, Croatia
marijan.palmovic@erf.hr

[1] Corresponding author

The Preference for Approximation

Abstract. A curious fact about the communication of numerical information is that speakers often choose to use approximate or rounded expressions, even when more precise information is available (for instance, reporting the time as 'three thirty' when one's watch reads 3:27). It has been proposed that this tendency towards rounding is driven by a desire to reduce hearers' processing costs, a specific claim being that rounded values produce the same cognitive effect at less cognitive effort than non-round values (Van der Henst et al. 2002). To date, however, the posited processing advantage for roundness has not been experimentally substantiated. Focusing on the domain of temporal expressions, we report on two experiments that demonstrate that rounded clock times are easier to remember and manipulate than their non-round counterparts, a finding that supports the role of processing considerations in numerical expression choice. We further find a role for domain-specific granularity of measurement.

Key words: number, clock time, approximation, imprecision, granularity

1. Introduction

It is typically assumed that it is better to be precise than imprecise. But with regards to the communication of numerical information, a variety of evidence points instead to a speaker and hearer preference for imprecision or approximation over precision.

Round numbers such as 100 and 40 tend to be interpreted approximately, while non-round numbers such as 99 and 43 are interpreted precisely (Krifka 2002, 2007, 2009). Dehaene and Mehler (1992) present data from numerous languages showing that round numbers are used much more frequently than non-round numbers of similar magnitude, a pattern they attribute to their use in conveying approximate quantities. Thus on a broad scale, approximation seems to be the preferred option. In fact, in many contexts the use of a non-round number strikes us as comically over-precise, as in Krifka's (2007) example of a road sign near the Zurich airport which alerts drivers to the presence of a stop sign 103 meters ahead.

Speakers round even when they have more precise information available. Van der Henst, Carles & Sperber (2002) demonstrate in a series of experiments that when answering the question “What time is it?”, speakers show a marked tendency to give answers rounded to the nearest 5-minute mark (e.g. 3:10 instead of 3:08). This is true even for individuals using digital watches, for whom rounding represents an additional effort over reporting the precise time displayed (see also Van der Henst & Sperber 2012 for further discussion, and Hayashi 2005 and Gibbs & Bryant 2008 for similar findings).

In some cases the preference for approximation might be attributed to speaker uncertainty (e.g., as to the accuracy of one’s wristwatch) or the possibility of measurement error (e.g., regarding the distance to a sign), but this is not always the case. In written materials such as the reporting of survey results, it is common to find an approximate expression of number or proportion side by side with an exactly percentage or value, as in the following example (cf. Williams & Power 2009 for similar examples).

(1) Six in ten Americans (59%) read the bible at least occasionally.

Apparently, reporting values or proportions in approximate or coarse-grained terms serves some communicative purpose that is not met by the communication of exact percentages. Rounding is furthermore not limited to the popular press or informal discourse: scientists communicating their results to their peers also make frequent use of approximations (Dubois 1987).

Why is there such a tendency towards rounding or approximation, even when this represents extra effort on the part of the speaker (or writer)? An intuitive answer is that rounding in some way makes the job of comprehension easier for the hearer. Van der Henst, Carles & Sperber (2002) propose just such an explanation for their findings, framed within the theory of communication known as Relevance Theory (Sperber & Wilson 1987, 1995), a central tenet of which is that speakers aim to produce utterances that are maximally relevant from the perspective of the hearer. Relevance in turn is a function of cognitive effect and processing cost: the greater the positive cognitive effect of an utterance and the lower the effort required to process it, the more relevant it is. With regards to telling the time, Van der Henst and colleagues propose that a rounded answer produces the same cognitive benefit as a precise one, but with a lower processing load. Specifically: "Suppose you have an appointment at 3:30 p.m. and it is 3:08. Being told 'It is 3:10' is likely to be optimally relevant: the two-minute departure from the exact time is unlikely to have any consequences, and the rounded answer is **<u>easier to process</u>**." (p. 464; emphasis added).

Van der Henst et al. support the proposal that rounding represents a hearer-oriented strategy via further experiments that show that speakers are sensitive to the level of precision relevant to the hearer. The simple but clever methodology used in their studies was to approach individuals in public places such as train stations, and ask them the time. If respondents did not round at all, we would predict on purely statistical grounds that 20% of their answers would be 'round', i.e. multiples of 5. In fact, round answers were much more frequent than this (98% among analogue watch wearers, 66% among those with digital watches), indicating that rounding is frequent. But when the requestor was perceived to require a more precise answer – when he claimed to be setting his watch, or to have a meeting in a few minutes – the frequency of rounded answers declined.

In work in Relevance Theory, the clock time example has been adopted as a prime case study for how the interplay between cognitive effect and processing cost that determines relevance can be experimentally substantiated (see e.g. Wilson & Sperber 2004, Zhang 2005, Sperber & Wilson 2008, Gibbs & Bryant 2008). Furthermore, the notion of a processing advantage for roundness is not inherently tied to a relevance theoretic account of language use. In seeking to explain the association between round numerical form and approximate interpretation, Krifka (2002, 2007) considers (though ultimately rejects) an account couched in the framework of Bidirectional Optimality Theory (Blutner 2000), which draws on two pragmatic principles, one favouring simple expressions over complex ones, the second favouring approximate interpretations over precise ones. The latter principle, he suggests, might be motivated in that the more coarse-grained representation of quantitative information might be “cognitively less costly” (Krifka 2007, p. 112). From a different perspective, Kao et al. (2014) provide experimental evidence that round numbers are more likely than ‘sharp’ ones to be interpreted precisely, and show that this can be successfully modelled within a Rational Speech Act framework – but only by assuming a higher processing cost for non-round numbers.

Given that the posited processing advantage plays a crucial role in diverse theoretical accounts of round number use and interpretation, it is thus perhaps surprising that this has the status of an unverified assumption. None of the above-mentioned authors offer any direct evidence that round numbers are in fact easier to process, nor is there a large body of existing research to this point. In the one directly relevant study we are aware of, Mason, Healy & Marmie (1996) found that subjects’ memory for numbers produced as answers to mathematical problems was better for round numbers (e.g. 11,000) than non-round numbers (11,365), even when tested only on the first two digits. This is consistent with the hypothesized processing advantage for roundness, but hardly conclusive. The first objective of the present research is thus to provide a firmer empirical basis for existing theoretical accounts by directly testing the processing of round versus non-round numerical expressions.

Should the hypothesized processing advantage be experimentally substantiated, we are left with further questions that are relevant to the choice between theoretical approaches. Specifically, what aspect of a linguistic form, or of its meaning, results in it requiring less effort to process? This is closely related to the question of which numerical expressions are so advantaged, and whether this is a categorical or graded phenomenon.

One immediately appealing possibility is that the processing advantage derives from frequency. As noted above, round numbers are used more frequently than their non-round counterparts; more frequent expressions are well known to incur lower processing loads (e.g. Oldfield & Wingfield 1965, Balota & Chumbley 1984, and much later work). A second possibility is that the crucial property is simplicity. Round numbers are privileged in their representation in Arabic numerals (the roundest in particular ending in 0). Perhaps more importantly, their verbal forms are on average briefer than those of non-round numbers (e.g. 'one hundred' – 3 syllables vs. 'one hundred and three' – 5 syllables; Krifka 2007). Shorter expressions are also advantaged on some processing tasks (e.g. memory span; Baddeley et al. 1975).

However, there is reason to think that neither of these potential explanations is entirely satisfactory. An account based on frequency runs the risk of circularity: if round numbers are easier to process because they are encountered more frequently, what is the cause of their greater frequency, if not that they are in some way easier for speakers and/or hearers to deal with? With regards to brevity or simplicity, Krifka (2007) observes that rounder numbers are not necessarily shorter or simpler than their less round counterparts: *forty five*, for example, is no shorter than *forty three*, and *one hundred* is actually longer than *ninety*.

In addressing the question of which numerical forms allow approximate interpretations, Krifka proposes that the crucial factor is not roundness itself but rather scale granularity. The results of measurement can be reported with respect to scales that differ in how coarse- or fine-grained they are, that is, in the density of their scale points. For example, relative to a scale whose basic unit is the mile, the distance between Amsterdam and Milan might be reported as 514 miles; but relative to a coarser-grained scale – say, one based on units of 100 miles -- that same distance might be reported as 500 miles. According to Krifka, typical scale granularity levels are based on powers of ten (10-20-30-…; 100-200-300-…; etc.) as well as the results of applying to such scales an operation of halving (5-10-15-…) or doubling (20-40-60-…). But in certain domains other structures are observed. A prime example is the measurement of time, for which Krifka proposes that possible granularity levels include a scale that counts hours, one that counts half hours, one that counts quarter hours (i.e. that counts in increments of 15 minutes), one that counts in 5 minute increments, and one that counts in minutes. The last three of these are represented in Fig. 1. Speaking precisely, i.e., relative to the 1-minute-granularity scale (a) in Fig. 1, the time displayed might be described as 2:41. But speaking more approximately, that same time might be described as 2:40 (5-minute granularity level (b)) or even 2:45 (15-minute level (c)).

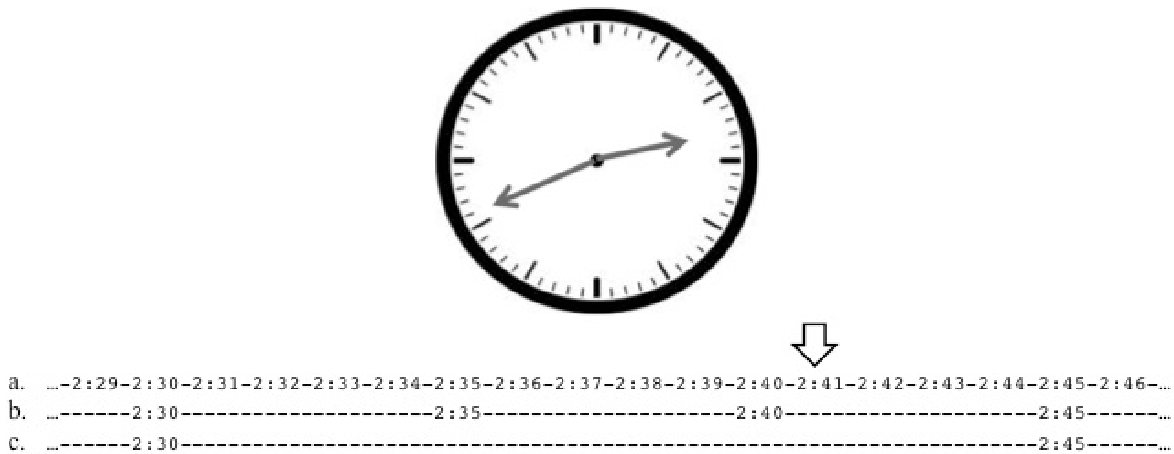

Figure 1.

Relating Krifka's theory to the above discussion of the processing of numerical expressions, the posited advantage for round numbers might be reconstrued as an advantage for values that occur on coarser-grained scales (e.g. 2:45) over those that occur only on scales of finer granularity (e.g. 2:41). Seen this way, it is forms that allow approximate interpretation, or perhaps approximate interpretation itself, that incurs lower processing costs.

Importantly, a theory of round number use based on scale granularity makes different predictions regarding which forms will be favoured relative to a theory based on roundness as a purely mathematical property of numbers. Jansen & Pollmann (2001) operationally define roundness in terms of divisibility properties, specifically whether a number can be expressed as a single digit multiple of a power of 10, 5 times a power of 10, 2 times a power of 10, and/or 2.5 times a power of 10. The more of these properties a number has, the rounder it is. For example, 100 can be expressed in all of these ways (1×10^2; $2\times(5\times10^1)$; $5\times(2\times10^1)$; $4\times(2.5\times10^1)$), and as such is maximally round. By contrast, 40 is less round, in that it can be expressed in only three of these ways (4×10^1; $8\times(5\times10^0)$; $2\times(2\times10^1)$), and 45 is even less round, being expressible in just one of these ways ($9\times(5\times10^0)$). Finally, a number like 43 has none of these divisibility properties, and is thus non-round. If ease of processing is correlated with roundness measured in this way, we would predict the ranking of processing costs shown in (2a) (where < is to be understood as 'easier to process'). But suppose the values in question in fact correspond to times in minutes. On Krifka's account, 45 minutes occurs on a coarser scale than 40 minutes, which in turn occurs on a coarser scale than 43 minutes; if it is scale granularity in this sense that drives processing ease, then we predict instead the ranking in (2b).

(2) a. 40 < 45 < 43

b. 45 min < 40 min < 43 min

The second objective of the present research is thus to shed light on the source of the posited processing advantage for round numbers, by investigating the extent to which domain-specific scale granularity has an effect above and beyond that of roundness as a mathematical property.

2. Experiments

We report on two experiments designed to investigate the following questions:

1) Is there a processing advantage for 'rounder' numerical expressions?
2) To the extent that such an advantage can be demonstrated, is there evidence for the role of domain-specific granularity of measurement?

We focus on the domain of clock times, both because this domain has been the subject of prior research on rounding, and because it offers the potential to tease apart numerical roundness and scale granularity.

None of the accounts positing a processing advantage for roundness makes specific predictions as to which aspect(s) of processing are impacted. Lacking any such theoretically motivated guidance, we have chosen to utilize two experimental tasks that simulate ordinary, everyday tasks that speakers might need to carry out with numerical information, more specifically clock times. The first is a memory task, parallel to the task of remembering a set of times (for example, train departure times). The second is a novel "clock math" task, parallel to everyday reasoning about temporal differences.

2.1 Experiment 1

Our first experiment was a short-term memory task, employing the widely used Sternberg paradigm (Sternberg 1966).

2.1.1 Participants

Participants were 34 native German speakers (mean age 26.6, 25 female), students at the University of Bielefeld and Humboldt University Berlin, or recruited by word of mouth in the Berlin area. Data from an additional 4 participants were excluded due to low accuracy or indications that they misunderstood the task. Participants were paid 5 euros for participation.

2.1.2 Materials

Stimuli for the experiment consisted of sequences of 3, 4 or 5 clock times, described to participants as departure times for trains, followed by a probe time. Participants' task was to decide whether or not the probe occurred in the sequence.

The primary experimental manipulation involved the granularity of the scale on which the time expression occurs. Three granularity levels were tested:[2]

- Coarse, i.e. 15-minute granularity	Examples: 3:15, 8:30
- Medium, i.e. 5-minute granularity	Examples: 3:10, 8:25
- Fine, i.e. 1-minute granularity	Examples: 3:21, 8:36

Each stimulus item included times of a single granularity level (coarse, medium or fine); the probe, whether correct or incorrect, also had the same granularity level as the items in the sequence.

Stimuli were constructed as follows. First, for each sequence length (3, 4 or 5 times), 10 item templates were constructed, each consisting of an increasing sequence of hour values with placeholders for the minute values, plus a probe of the same form (see Figure 2(a)). Hour values ranged from 2 to 9 (neither 2-digit hour values nor 1 as an hour value were used). 5 item templates were correct (probe in sequence), and 5 incorrect (probe not in sequence); incorrect items were such that the hour value and the minute value each occurred separately in the sequence. The placeholders were then replaced by minute values at each of the three granularity levels. Here, a potential confound is that at the coarse granularity level, there are only 4 possible minute values (:00, :15, :30, :45), compared to 12 values at the medium granularity level (8 if coarse-grained values are excluded) and 60 at the fine level (48 if coarse/medium values are excluded). To manage this and ensure greater parallelism

[2] Note that times at the coarse granularity level also occur on scales of medium and fine granularity, and likewise those of medium granularity also occur on the scale of fine granularity. In what follows we associate each time with the coarsest scale on which it occurs.

between the three conditions, 3 minute values were selected for each granularity level to form the basis of the stimuli:

- Coarse :15 :30 :45
- Medium :10 :25 :40
- Fine :21 :36 :51

These values were plugged into the item templates to create stimuli items (see Figure 2(b)). This yielded 90 items in total (3 sequence lengths x 10 item templates x 3 granularity levels), with the items in the three granularity conditions parallel in structure.

(a) Item template	(b) Stimuli items
2:xx 6:yy 8:zz Probe: 2:yy	**Coarse:** 2:45 6:30 8:15 Probe: 2:30
	Medium: 2:10 6:25 8:40 Probe: 2:25
	Coarse: 2:36 6:51 8:21 Probe: 2:51

Figure 2.

2.1.3 Procedure

The task was administered in German on a PC running E-Prime. In each trial, participants saw a fixation cross for 2 seconds, followed by the sequence of times (each shown for 2 seconds), followed by a pause of 0.5 seconds, and then the probe time. Participants pressed one of two keys to indicate whether or not the probe occurred in the sequence. Sequence lengths were presented in blocks (3 value, then 4 value, then 5 value), with breaks between blocks. Order of trials was randomized within blocks. There were 3 practice trials at the beginning of the experiment. Participants' responses and reaction times were recorded.

2.1.4 Predictions

If round clock times are favoured in terms of short-term memory relative to non-round times, we predict greater accuracy and faster reaction times for the coarse and medium conditions versus the fine condition, since both coarse and medium items are multiples of 5 (i.e. correspond to the clock times characterized as round by Van der Henst et al. 2002).

If the granularity of scales specific to time measurement plays a role in addition to or independently of that of roundness, such that items occurring on coarser-grained scales are favoured over those occurring on finer-grained scales, we further predict greater accuracy and faster reaction times for the coarse condition relative to the medium condition. If, on the other hand, it is roundness as a mathematical property of numbers that is the crucial factor driving ease of processing, we predict no such effect. In fact, we might even expect better results for medium versus coarse, since the minute places in the medium condition (10, 25 and 40) are in Jansen & Pollmann's sense rounder on average than those in the coarse condition (15, 30, 45).

2.1.5 Results

Trials with reaction times less than 500 msec or greater than 10,000 msec were removed prior to analysis; this resulted in the removal of 10 trials (0.4% of total).

Results for percentage correct are shown in Table 1. A generalized linear mixed effects model was fitted to the data, using the lme4 package (Bates et al., 2014) in R (R Core Team, 2015), with response (correct or incorrect) as dependent variable, granularity and sequence length as fixed effects, and random intercepts for subject and item template. Granularity was Helmert contrast coded (1: coarse vs. medium; 2: coarse+medium vs. fine); this coding scheme is suited to the predictions being tested, in that it provides a comparison between round (i.e. coarse+medium) and non-round items, and within the round category between coarse and medium items. The effect of sequence length was significant ($z = -2.021$, $p < 0.05$). More importantly, the first granularity contrast was not significant ($z = -1.219$ $p = 0.22$), indicating no reliable difference between coarse and medium granularity. The second granularity contrast was, however, significant ($z = -3.489$, $p < 0.001$), such that respondents were less accurate on fine items than on coarse/medium items. Finally, there was a significant interaction between the second granularity component and sequence length (z =

3.051, $p < 0.01$), in that the effect of fine versus coarse/medium granularity was attenuated at longer sequence lengths (which were presented later in the course of the experiment). Testing via ANOVA established that the model as described is superior to simpler models (in particular one without the interaction term) and is not improved by the inclusion of additional predictor variables.

	Granularity			
Length	**Coarse**	**Medium**	**Fine**	**Total**
3	87.8	84.7	79.0	83.8
4	80.0	80.1	73.5	78.1
5	73.4	75.4	77.6	75.7
Total	80.6	80.3	76.7	79.2

Table 1.

Results for reaction time on correctly answered trials are shown Table 2, and the log-transformed reaction times are displayed in Figure 3.[3] A linear mixed effects model was fitted using lme4 and lmerTest (Kuznetsova et al. 2015), with log-transformed reaction time as dependent variable, granularity, sequence length and presence of probe in sequence as fixed effects, and random intercepts for subject and item template. Granularity was Helmert contrast coded as above. The effect of sequence length was significant ($t = 6.452$, $p < 0.001$), with longer sequences eliciting longer reaction times. There was also a significant effect of correct response ($t = -6.343$, $p < 0.001$), such that items in which the probe occurred in the sequence (i.e., for which the correct response was “yes”) elicited shorter reaction times than those in which the probe did not occur in the sequence (i.e., for which the correct response was “no”). Once again the first granularity contrast was not significant ($t = -0.409$, $p = 0.68$), indicating no difference in reaction time between coarse and medium, but the second contrast was significant ($t = 3.323$, $p < 0.001$), indicating that reaction times were longer for fine items than for coarse/medium items. The model as described was superior to ones with

[3] Here and in Experiment 2, reaction times have been log-transformed prior to statistical analysis to better approximate a normal distribution.

fewer predictor variables, and was not significantly improved by the inclusion of additional predictor variables (in particular interaction terms).

	Granularity			
Length	**Coarse**	**Medium**	**Fine**	**Total**
3	1563.7 (727.3)	1567.2 (737.8)	1785.3 (999.8)	1634.4 (831.0)
4	1936.7 (1048.4)	1931.5 (1099.9)	2019.1 (1148.3)	1960.8 (1097.6)
5	2245.0 (1425.6)	2179.9 (1386.9)	2209.4 (1234.0)	2211.2 (1347.9)
Total	1895.5 (1118.8)	1880.6 (1121.8)	2003.5 (1142.8)	1925.4 (1128.5)

Table 2.

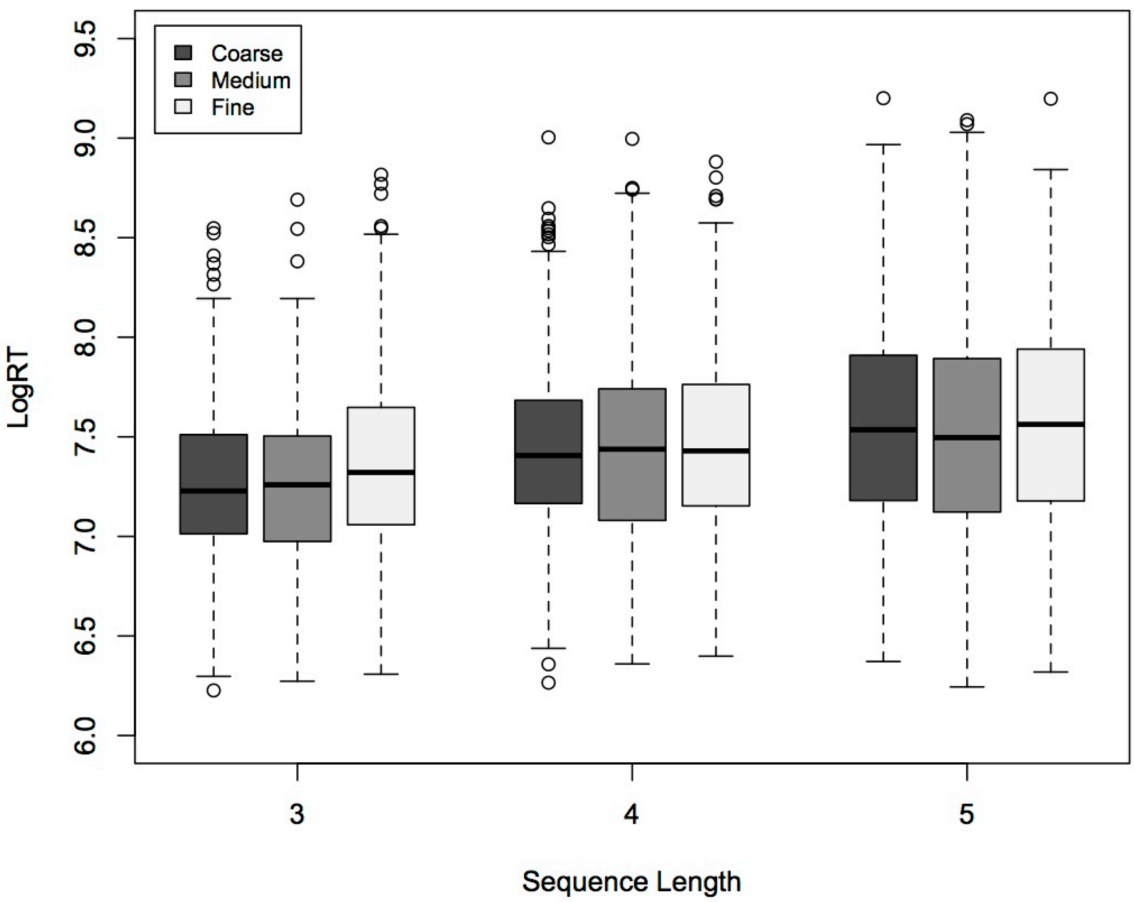

Figure 3.

2.1.6 Discussion

The overall question addressed in the present research is whether round numerical values are less effortful to process than their non-round counterparts, as is assumed in diverse theoretical accounts of round number usage and interpretation.

In an experimental task designed to simulate one aspect of processing, namely remembering strings of numerical values (here, clock times), the prediction of a processing advantage is borne out: ‘round’ clock times such as 6:25 and 6:30 are recalled more accurately and quickly than ‘non-round’ times such as 6:21. This effect was observed to disappear at the longer sequence lengths that were presented later in the experiment, suggesting that participants may have developed some sort of strategies over the course of completing the task. However, the fact that the effect is observed most clearly in the first block of trials (sequence length 3) speaks to its robustness. In short, round clock times are in fact easier for recipients.

Our experiment included times occurring on scales of three distinct granularity levels: coarse (15-minute granularity), medium (5-minute) and fine (1-minute). Of these, times at the coarse and medium levels are both round in the sense of van der Henst et al. (2002) (divisible by 5); those at the fine level can be characterized as non-round. That we found a difference in accuracy and reaction times between coarse/medium and fine, but not between coarse and medium, suggests that on this task at least, there is no effect of domain-specific scale granularity beyond that of roundness.

Beyond considerations of scale granularity, the present experiment can only shed limited light on the source of the advantage in short-term memorability for round clock times. One possibility in particular that is not ruled out is that the roundness effect found on this task was due simply to the length of the stimuli items. As discussed in the introduction, round numerical expressions are typically shorter (in morphemes or syllables) than their non-round counterparts, and shorter expressions are known to be easier in short-term memory tasks. When (generalized) linear mixed models are fitted to our experimental results in which the granularity of stimuli items is replaced by their length in syllables, the goodness of fit (as measured by the Akaike Information Criterion, or AIC) is no different from that of the above-described models; the syllable-based models are furthermore not improved by the inclusion of granularity as an additional predictor.

These last two points, however, lead to an important observation about the nature of this experimental task, namely that it did not force respondents to process the stimuli as clock times. Perhaps 6:20, for example, was not stored as a time, but as the numeral 620, or even the string of syllables *six-twen-ty*.[4] In this case, granularity levels specific to the domain of clock times would not be expected to play a role. We address this in Experiment 2, with a task that is likely to be less sensitive to the simple length of stimuli items, and which forces participants to process stimuli as clock times.

2.2 Experiment 2

Our second experiment used a novel "clock time arithmetic" paradigm, intended to simulate the everyday task of reasoning about temporal differences. For example, if my train leaves at 8:20 and it takes 45 minutes to get to the station, what time do I need to leave home? If the train trip itself takes 53 minutes, what time will I arrive at my destination? Crucially, because addition or subtraction of clock times must be carried out modulo 60, this task forces participants to process stimuli as times rather than numbers or simple syllable strings.

2.2.1 Participants

[4] More accurately, the syllable string in this case would be *sechs-uhr-zwei-und-zwan-zig*, as the task was administered in German to German-speaking subjects.

Participants were 22 native German speakers (17 female, mean age 24.4), students at Humboldt University Berlin and/or recruited by word of mouth in Berlin. Participants were paid 4 euros for participation.

2.2.2 Materials

Test items were clock time addition/subtraction problems, each consisting of a start time, an operation (plus or minus) and an increment time in minutes. For example:

A)	3:45 plus 30 ----- 4:15	B)	6:48 minus 26 ----- 6:12

Each problem was followed by a possible answer; participants' task was to decide whether this answer was correct (as in A) or incorrect (as in B).

A total of 720 test items were created, in which the following factors were varied:

i) Start time granularity level:

- Coarse (15-minute): minute-place :00, :15, :30, :45
- Medium (5-minute): minute-place :05, :20, :35, :50
- Fine (1-minute): minute-place :03, :18, :33, :48

ii) Increment granularity level:

- Coarse (15-minute): 30, 45
- Medium (5-minute): 25, 40
- Fine (1-minute): 26, 41

iii) Operation: plus or minus

Half of test items featured correct answers, half featured incorrect answers; incorrect answers differed by +/– 5 minutes or +/– 10 minutes from the correct answer, with the consequence that participants could not answer correctly based on parity alone. The design of the stimuli was such that for some items the correct answer featured the same hour place as the start time (as in B), but for others the answer ‘spilled over’ to the next or previous hour vs. the start time (as in A); this is crucial because to answer these items correctly, participants had to add/subtract modulo 60, and thus were forced to treat the numerical values as times. The hour place for start times ranged from 2 to 8; due to spillover, the hour place for the end times ranged from 1 to 9. Note also that in some cases of medium start granularity and medium increment granularity, the correct end time was at the coarse granularity level (e.g. 4:50 plus 40 equals 5:30); this has the effect of blurring the distinction between coarse and medium granularity levels, and as such has a conservative effect.

2.2.3 Procedure

The task was administered in German on a PC running E-Prime. Participants read the instructions on the screen, and completed three practice trials. They then completed 3 blocks of 48 trials each, drawn randomly from the list of 720 test items. On each trial, participants saw a fixation cross for 2 seconds, followed by the start time for 1 second, and then the operation and increment time for 1.5 seconds, and then the end time. Participants pressed one of two keys to indicate whether the displayed end time was correct or incorrect. Response and reaction time were recorded.

2.2.4 Predictions

The predictions for this experiment are parallel to those for Experiment 1: if round clock times are easier to manipulate (in the relevant respect) than non-round times, we predict greater accuracy and shorter reaction times for items with coarse and medium starting granularity versus those with fine starting granularity, and likewise for items with coarse/medium increment granularity versus those with fine increment granularity (recall that coarse and medium values are both ‘round’ in the sense of being divisible by 5, while fine are non-round). If ease of manipulation is affected by the granularity of the scales on which time expressions occur, such that those interpretable relative to coarser-grained scales are easier than those occurring only on finer-grained scales, we predict not just these differences but also greater accuracy and shorter reaction times for items with coarse start/increment granularities than those with medium start/increment granularities. If instead, numerical roundness alone is responsible for degree of processing effort required, no such differences between coarse and medium are predicted.

2.2.5 Results

Before analysis, one trial with a reaction time <200 msec was removed as an outlier.

Results for percentage correct are shown in Table 3. A generalized linear mixed effects model was fitted to the results, with response (correct or incorrect) as dependent variable, start granularity, increment granularity and spillover (yes/no) as fixed effects, and random intercepts for subject and item. Both start granularity and increment granularity were Helmert contrast-coded as described in Experiment 1. The effect of spillover was significant ($z = -6.911$, $p < 0.001$), with subjects making more errors on trials in which the correct answer 'spilled over' into the previous or next hour. More importantly, for start granularity, the first contrast was significant ($z = -2.225$, $p < 0.05$), indicating subjects were less accurate for items with medium versus coarse starting granularity level. The second contrast was also significant ($z = -5.198$, $p < 0.001$), indicating an additional decline in accuracy for fine starting granularity in comparison to coarse/medium. For increment granularity, the first contrast was likewise significant ($z = -1.987$, $p < 0.05$), indicating lower accuracy at medium vs. coarse increment granularity. The second contrast was also significant ($z = -4.015$, $p < 0.001$), signifying that accuracy was still lower for fine increment granularity relative to coarse/medium increment granularity. This model outperformed simpler models and was not significantly improved by adding additional predictor variables (including interaction terms).

Start Granularity	Increment Granularity			
	Coarse	**Medium**	**Fine**	**Total**
Coarse	97.1	93.0	89.2	93.1
Medium	91.3	92.8	85.1	89.7
Fine	86.9	83.6	82.8	84.5
Total	91.7	89.9	85.7	89.1

Table 3.

Results for reaction time on correctly answered trials are shown in Table 4, and log-transformed reaction times are displayed graphically in Figure 4. A linear mixed effects model was fitted with log-transformed reaction time as dependent variable, start granularity, increment granularity, spillover, direction (forwards/backwards) and correct answer (correct/incorrect) as fixed effects, and random intercepts for subject and item. The effect of spillover was significant ($t = 15.403$, $p < 0.001$), such that responses were slower on spillover trials. There was also a significant effect of correct answer ($t = -5.437$, $p < 0.001$) and direction ($t = -2.310$, $p < 0.05$), with reaction times shorter on correct trials, and when the direction was forwards (i.e. when the operation was addition rather than subtraction). More crucially, for start granularity, both contrasts were significant (<u>contrast 1</u>: $t = 2.829$, $p < 0.01$; <u>contrast 2</u>: $t = 8.973$, $p < 0.001$), indicating that reaction times were longer for medium vs. coarse starting granularity, and still longer for fine vs. coarse/medium. Both contrasts for increment granularity were likewise significant (<u>contrast 1</u>: $t = 3.724$, $p < 0.001$; <u>contrast 2</u>: $t = 8.704$, $p < 0.001$), meaning that reaction times were longer for medium versus coarse, and also longer for fine vs. coarse/medium. Finally, there was a significant interaction between the second contrasts for start granularity and increment granularity ($t = -2.802$, $p < 0.01$), such that the effect of fine vs. coarse/medium increment granularity was less pronounced at fine versus coarse/medium start granularity. As above, this model was superior to simpler ones including fewer predictors (in particular, to one without the interaction term).

Start Granularity	Increment Granularity: Coarse	Medium	Fine	Total
Coarse	808.7 (648.8)	944.8 (626.4)	1221.8 (1053.2)	986.3 (812.9)
Medium	940.1 (672.9)	969.9 (707.3)	1259.3 (1061.2)	1051.9 (838.1)
Fine	1252.8 (1190.1)	1293.2 (934.5)	1434.7 (1133.7)	1325.2 (1097.3)
Total	993.6 (882.2)	1059.2 (773.7)	1301.9 (1084.7)	1114.5 (929.9)

Table 4.

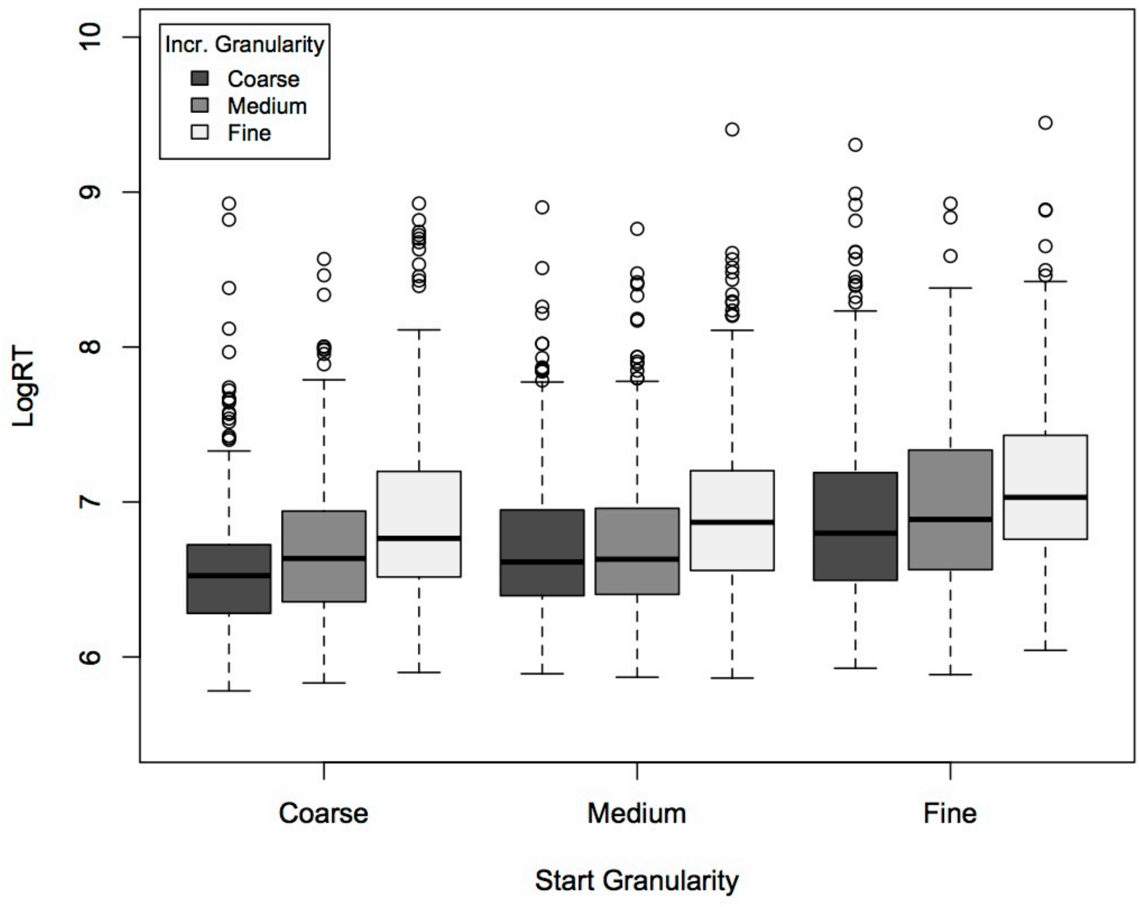

Figure 4.

2.2.6 Discussion

The task in Experiment 2 was designed to simulate a common everyday sort of temporal reasoning, namely that involved in determining for instance what is 30 minutes after 3:45, or 26 minutes before 6:48. On this task, as in Experiment 1, we found an effect for the roundness of the temporal expressions involved: subjects were more accurate and responded more quickly when the starting time was round than when it was non-round, and also when the temporal increment was round versus non-round. Thus while this experiment assessed a very different aspect of 'processing' than the previous one, the overall finding is the same: rounder clock times are easier.

Importantly, in the present experiment – in contrast to Experiment 1 – we found a difference not just between round (i.e. coarse/medium) and non-round (fine) items, but also within the round category between coarse and medium items; this was the case for both starting and increment granularities. We attribute the difference between these findings and those from Experiment 1 to the fact that the present task required participants to encode and manipulate stimuli items as times in order to compute the necessary modulo 60 calculations.

In simple terms, our findings are the following: values that correspond to the measurement of time in units of 15 minutes require less effort to process than those that correspond to measurement in increments of 5 minutes, and these in turn require less effort than those corresponding to measurement at the 1-minute level. Thus on this task, the processing advantage accrues not simply to values that are round in the sense of their divisibility properties, but further to those that correspond to points on a coarser-grained scale of time measurement. This interpretation tends to be reinforced by our finding of an interaction between start granularity and increment granularity, such that the cost of fine increment granularity was most pronounced at coarse/medium start granularity, and had less effect when the start granularity itself was fine; this suggests that when participants were forced by the value of the starting time to adopt a fine-grained time representation, there was less incremental cost incurred in calculating fine-grained increments. Overall, the results of this experiment thus provide evidence for the role of scale granularity in the processing of numerical expressions.

It is plausible that the advantage we found for time expressions at the 15- and 5-minute levels has its source in part in participants' familiarity with the analogue clock face, which typically makes visible precisely these levels of granularity (see Fig. 1). Visualizing the clock face might have also have helped participants calculate differences in increments of quarter or half hours, i.e. at the coarse granularity level. These findings do not let us differentiate between an explanation in terms of the structure of measurement scales in an abstract sense and one based on the visual representation of time on the clock face. But in either case the more general point remains the same: it is not simply the form of a time expression or its mathematical properties that are that is relevant to ease of processing (in the aspect considered here), but also its role in our conventional system of time measurement.

3. Conclusions and topics for further investigation

When it comes communicating numerical information, speakers typically have a choice between multiple potentially acceptable expressions: *three twenty seven* or *half past three*; *fifty nine percent* or *six in ten*; and so forth. The starting point for this paper was the observed tendency for speakers to choose rounded expressions, and the hypothesis that such rounding reflects a strategy aimed at reducing processing load on the part of the hearer, thereby maximizing the relevance of the overall utterance. In two experiments designed to simulate different sorts of everyday numerical tasks, we found the hypothesized processing advantage for roundness to be substantiated: round clock times are recalled and manipulated more accurately and quickly than their non-round counterparts. These findings are supportive of a theory of communication according to which speakers' choices of numerical expression are shaped in part by processing-related concerns. One possible such theoretical approach is the relevance-theoretic account outlined in Section 1; but the same insight might also be incorporated into another framework, such as the Rational Speech Acts model of Kao et al. (2014) or the Constraint Based system of Cummins (2015).

Our research focused on time expressions. A natural question is whether these findings would generalize to other sorts of numerical information, given that the tendency towards rounding is observed in other domains beyond the temporal one. For example, is *one hundred* easier to recall and reason with than *one hundred and three*, or *one third* easier than *34 percent*? We hypothesize that we would obtain similar results for expressions such as these, and thus that a unified theory of rounding across domains will be possible. But this of course needs to be experimentally verified.

The present research does not fully address the question of the <u>source</u> of the lower processing cost for round numbers, but we can offer some initial observations. Our findings are compatible with the possibility that the processing advantage is due at least in part to low-level factors such as expression length, round expressions being on average shorter in than non-round ones. But certain of our findings – particularly those from Experiment 2 – suggest that something more than this is going on. Specifically, on at least one sort of task, the expressions that are advantaged are not simply those that are shorter or rounder in a mathematical sense, but further those that correspond to a coarser-grained level of measurement. This suggests that the ease of processing is correlated not only to the form of a numerical expression but also to the representation of measurement that it encodes.

This latter point gives rise to a further question. Our experiments showed an advantage for temporal expressions <u>associated with</u> the approximate or coarse-grained representation of time, but not for coarse-grained representations themselves. In fact, both tasks we employed (short term memory and clock time arithmetic) required respondents to recall or manipulate the stimulus items exactly. We are thus led to ask whether approximate or coarse-grained representations of measurement are themselves less costly to process than precise ones. We leave this as a topic for future investigation.

References

Bates D., M. Maechler, B. Bolker and S. Walker. 2014. lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1-7, <URL: http://CRAN.R-project.org/package=lme4>.

Balota, D. A. and J. I. Chumbley. 1984. Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. *Journal of Experimental Psychology: Human Perception and Performance* 10(3): 340-357.

Baddeley, A.D., N. Thompson and M. Buchanan. 1975. Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior* 14(6): 575-589.

Blutner, R. 2000. Some aspects of optimality in natural language interpretation. *Journal of Semantics* 17: 115-162.

Cummins, C. 2015. Constraints on Numerical Expressions. Oxford: Oxford University Press.

Dehaene, S. and J. Mehler. 1992. Cross-linguistic regularities in the frequency of number words. *Cognition* 43: 1-29.

Dubois, B.L. 1987. Something on the order of around forty to forty-four: Imprecise numerical expressions in biomedical slide talks. *Language in Society* 16: 527-541.

Gibbs, R.W. and G. A. Bryant. 2008. Striving for optimal relevance when answering questions. *Cognition* 106: 345-369.

Hayashi, H. 2005. Truthfulness and relevance in Japanese. In B. G. Bara, L. Barsalou and M. Bucciarelli (eds.), *Proceedings of Cogsci 2005: XXVII Annual Conference of the Cognitive Science Society*, 2490. Mahwah, NJ: Lawrence Erlbaum.

Jansen, C. J. and M. M. Pollmann. 2001. On round numbers: pragmatic aspects of numerical expressions. *Journal of Quantitative Linguistics* 8: 187-201.

Kao, J. T., J. Y. Wu, L. Bergen and N. D. Goodman. 2014. Nonliteral understanding of number words. *Proceedings of the National Academy of Sciences* 111(33): 12002-12007.

Krifka, M. 2002. Be brief and vague! And how bidirectional optimality theory allows for Verbosity and Precision. In D. Restle and D. Zaefferer (eds.*), Sounds and Systems. Studies in Structure and Change: A Festschrift for Theo Vennemann*, 439-458. Berlin: Mouton de Gruyter.

Krifka, M. 2007. Approximate interpretation of number words: A case for strategic communication. In G. Bouma, I.Krämer and J. Zwarts (eds.), *Cognitive Foundations of Interpretation*, 111-126. Amsterdam: Koninklijke Nederlandse Akademie van Wetenschapen.

Krifka, M. 2009. Approximate interpretations of number words: A case for strategic communication. In E. Hinrichs and J. Nerbonne (eds.), *Theory and Evidence in Semantics*, 109-132. Stanford: CSLI Publications.

Kuznetsova, A., P. B. Brockhoff and R. H. B. Christensen. 2015. lmerTest: Tests in Linear Mixed Effects Models. R package version 2.0-29. http://CRAN.R-project.org/package=lmerTest.

Mason, J. D., A. F. Healy and W. R. Marmie. 1996. The effects of rounding on memory for numbers in addition problems. *Canadian Journal of Experimental Psychology* 50(3): 320-323.

Oldfield, R.C. and A. Wingfield. 1965. Response latencies in naming objects. *Quarterly Journal of Experimental Psychology* 17(4): 273-281.

R Core Team. 2015. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.

Sperber, D. and D. Wilson. 1986. *Relevance: Communication and Cognition*. Oxford: Blackwell.

Sperber, D. and D. Wilson. 1995. *Relevance: Communication and Cognition* (2nd Edition). Oxford: Blackwell.

Sperber, D. and D. Wilson. 2008. A deflationary account of metaphors. In R. W. Gibbs Jr. (ed.), *The Cambridge Handbook of Metaphor and Thought*, 84-105. Cambridge: Cambridge University Press.

Sternberg, S. 1966. High-speed scanning in human memory. *Science* 153: 652-654.

Van der Henst, J. B., L. Carles and D. Sperber. 2002. Truthfulness and relevance in telling the time. *Mind & Language* 17: 457–466.

Van der Henst, J. B. and D. Sperber. 2012. Testing the cognitive and communicative principles of relevance. In D. Wilson and D. Sperber (eds.), *Meaning and Relevance*, 279-306. Cambridge: Cambridge University Press.

Williams, S. and R. Power. 2009. Precision and mathematical form in first and subsequent mentions of numerical facts and their relation to document structure. *Proceedings of 12th European Workshop on Natural Language Generation*, 118-121. Stroudsburg, PA: The Association for Computational Linguistics.

Wilson, D. and D. Sperber. 2004. Relevance theory. In L.R. Horn and G. Ward (eds.), *The Handbook of Pragmatics*, 607-632. Oxford: Blackwell.

Zhang, G. 2005. Fuzziness and relevance theory. *Foreign Language and Literature Studies (Quarterly)* 22(2): 73-84.