12 Numerosity Measures

Maybe the most straightforward method to quantify eye-movement events is simply to count them, either in absolute numbers, in proportion to the total number of events, or as a rate over time. Indeed, event-counting measures appeared very early in eye-movement research. For instance, Dodge and Cline (1901, p. 154) analysed reading data from three participants and noted "a little less than five reading pauses per line, making an average of a little less than four movements to the right for every movement to the left."

In Part II of the book, over 25 events were defined. Many of them are counted in research, some of them very often (like number of fixations), and some very seldom (like number of look-aheads). Not only are events counted, however, but also the number of participants and trials that fulfil some given criterion. For example, it may be of interest to count the number of participants that looked at a specific area of interest. This chapter deals with the *number, proportion* and *rate* of such countable entities in eye-movement research. It differs from all other measure chapters, which are only about single events or representations.

Countable entities	How researchers count them	Page
Saccades	*Number, proportion and rate*	403
Glissades	*Proportion*	405
Microsaccades	*Rate*	406
Square-wave jerks	*Rate*	407
Smooth pursuits	*Rate*	408
Blinks	*Rate*	410
Fixations	*Number, proportion and rate*	412
Dwells	*Number, proportion and rate*	417
Participants, areas of interest and trials	*Number and proportion*	419
Transitions	*Number, proportion and rate*	422
Regressions, backtracks, look-backs, and look-aheads	*Number and rate*	425

What are number, proportion, and rate?

In Figures 12.1 and 12.2, we exemplify the three types of counting which are done with events and other entities in eye-tracking experiments. The *number of* (*N*) is a unit that simply represents how many occurrences of something there are, for example 12 saccades as in Figure 12.1.

Proportion is a fraction between a part and a whole, having values between 0 and 1, but is typically reported as %. In Figure 12.2(b), the proportion of glissadic saccades, seen as small notches directly after the large saccadic peaks, is 42%, and in Figure 12.2(c), we calculate the proportion of saccades with peak velocities above 75°/s to be 50%. Since two quantities with the same unit are divided, proportion is dimensionless.

Rate is the ratio between a number and the extension of the temporal range over which that number was counted. The example in Figure 12.2(d) is saccadic rate, which counts 12

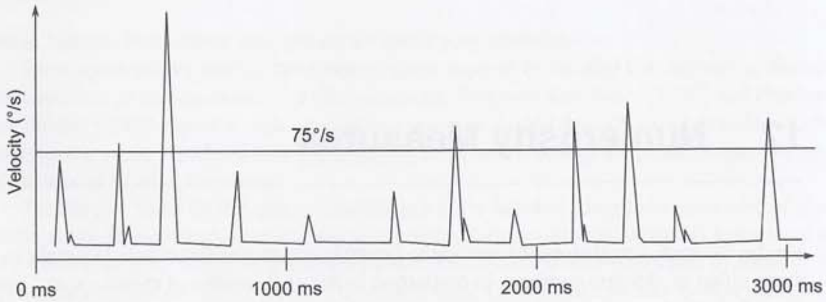


Fig. 12.1 Velocity-over-time diagram for a 3 second trial of synthetic data.

(a) 12 saccades. An example of number.

(b) Five out of twelve ≈ 0.42 = 42% glissadic saccades. This is a proportion calculation.

(c) 6/12 = 0.5 = 50% of the saccades have velocities above 75°/s. This is another proportion calculation.

(d) 4 saccades per second, or 4 s^{-1}. This is a rate calculation.

Fig. 12.2 Examples of number, proportion, and rate calculations for the data in Figure 12.1.

saccades over 3 seconds, equalling an average of 4 per second (or 4 s^{-1}).

All three types of counting are also used with the term 'frequency': in the terminology of the statistical analysis software SPSS, for instance, number is called 'absolute frequency', while 'relative frequency' covers both rate, proportion and various normalizations. These terms sift through to eye-tracking literature as well: Gilchrist and Harvey (2006) write 'frequency of saccades' when referring to number of saccades, and Foulsham and Kingstone (2010) use 'relative frequency' to denote proportion values, just to take two examples.

The term 'number of saccades' is about five times as common on Google Scholar as 'frequency of saccades' (January 2011). In this book, we reserve 'frequency' for the rate of regular repetitions per unit time, such as sampling frequency, and use the unit Hz for it. The term frequency will not be used with eye-movement data for other types of counting.

From an event-detection perspective, 'counting measures' are the least demanding. To count an event, it only needs to be detected, and no additional information about the event itself is needed: we only need to know that there is a saccade, but not its direction, velocity, duration, or skewness. Thus, event counting requires detection, but silently circumvents calculating values for event properties, and remains unaffected by the various noise this can introduce (Chapter 5).

However, as all events of a type are treated as equal when counting them (i.e. a 2° saccade equals a 25° saccade, and a 50 ms fixation is indistinguishable from a 1500 ms fixation) another type of noise, in the form of information reduction, is introduced. Although it may seem as though the events behind a count value are all equal, they very seldom are.

Grouping and limiting the counting range

There is often little point in calculating the total number of saccades or any other event in the whole recorded data. Instead, most researchers limit the counting process in a number of ways to extract count values that correspond to conditions in their experimental design. Data range restrictions can be imposed on the input to other measures also, but for counting measures this range is an integral part which defines the scope in which counting will be carried out.

A *spatial* division of data means, for instance, counting the number of fixations in each area of interest, as is common in scene perception research (p. 412).

A division of data into *temporal* portions can involve counting the event only in a limited temporal range: select to count in a single time period, for instance, the number of blinks during the three seconds before the participant clicks to end the trial, or the number of microsaccades during the first 250 milliseconds after fixation onset.

Some events can form temporal counting restrictions for other events. For instance, the number of microsaccades per fixation (Otero-Millan *et al.*, 2008) involves counting how many microsaccades (events) there are in each fixation (temporal counting range).

Limiting the counting process to *conditions in the experimental design* means, for example, counting the number of first skips of the word 'of' in an easy text condition of a reading experiment (perhaps including only the data collected from a group of 12-year old readers). This example illustrates the counting processes for one condition of the experimental design to familiarize the reader with the principle, but usually however the required events will be counted across *all* conditions of the experiment, and counted in relation to each of the independent variables you manipulate.

Multiple counting ranges: histograms

If we select to count in multiple time periods, each of these portions of time can be divided into a bin, and the count in each bin can be used to make a histogram, such as the number of microsaccades for each 250 ms bin. Counting data from multiple areas of interest can also be made into a histogram. This is possible with other measures as well, but the selection of both single and multiple counting ranges is particularly emphasized in the case of counting measures.

The implicit presuppositions to event counting

Some events can only exist in the presence of other events or when certain conditions are fulfilled, and this affects how they are counted. For instance, while fixations can be counted in their own right, glissades always co-exist with saccades. Therefore, glissades are counted as a proportion per saccade, and never in absolute numbers or rates per second. Microsaccades, when occurring during long fixations, can be counted per second, however. Table 12.1 summarizes the presuppositions that events have. Dwells, transitions, skips, and returns are always counted with reference to one or more specific areas of interest.

Measures and other ways to count

Table 12.2 summarizes what we have said above about how to construct a multitude of counting measures; not all of the many possibilities to count events have actually been utilized in research however. As always, the vast freedom we have to build measures does not mean we *should* build and calculate values for all of them. In fact, the selection of precise counting restrictions is an essential part of designing the experiment where you plan to use the counting

Table 12.1 Groups of events in eye-movement data: glissades can only co-exist with saccades, for instance, and dwells can exist only when areas of interest have been set up. These dependencies affect how the events are counted.

Events	Presupposition—required data parsing	Chapter
Saccades, fixations, smooth pursuits, blinks, nystagmuses, and artefacts	Event detection	5
Drifts, microsaccades, inter-microsaccadic intervals, square-wave jerks, glissades, catch-up-, back-up-, and leading saccades	Fixation, saccade or smooth pursuit	5
Area of interest hits, dwells, transitions, total skips, first skips, entries, and returns	Area of interest division of stimulus	6
Backtracks, regressions, look-backs, and look-aheads	Scanpath	8

Table 12.2 The five steps of forming counting measure.

1	Select restriction/range in space, time, and data
2	Select what event to count
3	Select absolute number, proportion, or rate
4	Count
5	Optionally form averages or normalize

measures, and as such they are part of your hypotheses.

Counting measures are very prevalent in research, for instance "number of saccades" gives roughly the same number of hits on Google Scholar as "saccade [saccadic] amplitude" and "dwell time". However, the effects studied with some counting measures vary much more than for other measures. In this chapter, we present a selection of existing counting measures sorted by the events they count, address operational definitions and methodological issues, and in most cases give examples of common effects that the measure has been used to study.

Notable measures that involve counting but have nevertheless been left out of this chapter are those based on transition matrices and proportion over time representations. The transition matrix counts transitions to investigate overall eye-*movement* behaviour, and measures using it are found in Chapter 10. Proportion over time representations count participants, but they are used to measure latency, and are therefore in Chapter 13.

Counting participants and trials

Participants are counted using restriction criteria such as having had to look at a specified area of interest, which allows us to count the number of participants who met this requirement. Using thresholding, unlimited forms of criteria can be applied to participant counting: for instance, the number of participants who had a peak saccadic velocity larger than 600°/s in more than 30% of the saccades. Apart from its use as measures, participant counts are reported as numbers of data quality, for instance if 23% of participants had to be removed from data analysis because of inadequate precision (p. 181).

Data quality is often the reason for counting (excluded) trials, but trials can also be counted using restriction criteria such as the number of trials with fewer than 20 saccades, or the number of trials with four or more dwells. Researchers also count number of correct trials, where 'correct' refers to the participant succeeding in performing a task as per the experimental design.

Averaging and normalizing count data

Count data lend themselves excellently to averaging and normalizing. Proportion and rate measures in this chapter are forms of normalization, with respect to the whole or the duration. Other measures rely on normalization against other properties on the counting context. For example, the measure 'fixation density' [number/cm^2] counts the number of fixations and divides it by the area within which the fixations were counted. Fixation density pertains more to the location of fixations in space, hence we include it as measure of position (see p. 390).

Counting and statistics

When analysing count data, there are at least two issues to be wary about from a statistical standpoint. The first is that count data are not inherently normally distributed (Gaussian), but are best described by a Poisson or a binomial distribution. The solution is either to use a statistical model that can handle Poisson-distributed data, which is the proper solution, or, approximate the distribution with a Gaussian distribution. The latter is done when the expected count (typically denoted as λ in a Poisson distribution) is sufficiently large, typically when $\lambda > 10$.

The second issue concerns proportions. Proportions are, as previously described, the ratio between a part to the whole, and as such constitute a number between 0 and 1. They typically suffer from a restricted range, i.e. a tail of the distribution is truncated and the variance becomes restricted the further we go to either extreme point. This issue is mainly a problem if the scores occur outside of 0.3–0.7. This issue and its possible solutions were discussed earlier (pp. 87–90).

12.1 Saccades: number, proportion, and rate

Target question	*How many saccades?*
Input representation	*Event-processed data*
Output	*The amount of saccades in number, proportion, or rate (s^{-1})*

The chosen event detection algorithm operationalizes, detects, calculates, and outputs the saccades in a trial. They are then counted in absolute number, in proportion to something, or as a rate. Saccades may be distinguished based on different saccadic properties (see Table 12.3).

For still images, the number of saccades should be equal or ±1 to the number of fixations. An increase in the number of saccades over a fixed trial duration equals a decrease in the average fixation duration. When there is motion and smooth pursuit, catch-up saccades may increase the number of saccades to be larger than that of fixations.

With reference to a well-known neurological patient, Land (2006) points out that people may often make more saccades than needed for the tasks at hand. A.I. is a participant with a condition called oculomotor apraxia in which saccades cannot be initiated; A.I instead makes saccadic-like head movements. In spite of this disorder A.I is, nevertheless, able to execute tasks with normal speed and competence, making only one third as many head saccades as controls make eye saccades. Thus, the implication is that we make more saccades than are strictly necessary.

Table 12.3 Common saccadic properties, and example criteria for proportion of saccade calculations from published research.

Saccadic property	Examples where proportion of saccades is commonly calculated
Direction towards locations	Saccades to each distractor type; saccades that ultimately end up in the target location; saccades that were captured by the onset; saccades that were directed at same-coloured targets or in the direction of the mouth in a face image
Direction, angular	Saccades in the correct direction; saccades where the direction changes relative to the previous saccade
Amplitude	Saccades with amplitudes smaller than X°; saccades that landed outside a specified 2° ring, for example
Latency	Saccades at a zero latency; saccades triggered after a certain time delay
Onset	Saccades launched during set timebands, of 500 ms for instance.

12.1.1 Number of saccades

The *absolute number of saccades* is calculated by using one of two operational definitions. The first counts saccades along *a distance with a constant length*, for instance the same line of text, or with sinusoidal smooth pursuit stimuli where the lengths are identical. The second definition instead uses *a fixed trial duration*, and counts the number of saccades from trial start to trial end. Both these operational definitions assure that count data are comparable between participants, trials, and conditions. It goes without saying that different stimuli evoke different numbers of saccades.

There are several sub-events to saccades that can be counted, for instance the number of large regressive saccades, or the number and extent of backtracking regressions, just to mention a few. The studies reviewed by O'Driscoll and Callahan (2008) use cyclic stimuli, and tend to show that participants with schizophrenia have significantly more saccades during smooth pursuit, both catch-up and leading saccades. Studies with higher proportions of medicated patients show smaller effect sizes for leading saccades and larger effect sizes for catch-up saccades.

12.1.2 Proportion of saccades

The proportion of saccades is counted by dividing the counted number of saccades in a subgroup by the number in the whole group. The subgroup and main group selected varies, and can depend on conditions in the experimental design and particular paradigm used. Nevertheless, when studies report the proportion of saccades, very often the subgroup is defined by saccadic properties that we recognize from other chapters.

Again, the specific subevents of saccades also form proportions, such as the proportion of saccades that were revisits, or the percentage of forward saccades from word N which skip word $N+1$. These are just two examples of many which can be derived.

12.1.3 Saccadic rate

Saccadic rate is measured as the number of saccades per second (s^{-1}), and is also known as 'number of saccadic movements per unit time' (Ohtani, 1971), 'eye transition frequency' (Wierwille & Connor, 1983), and 'eye-movement rate' (Bergstrom & Hiscock, 1988). For still picture stimuli, the saccadic rate should be identical to the fixation rate, and thus also correlate closely to fixation duration. For stimuli that elicit only smooth pursuit, the saccadic rate would be a measure of the prevalence of catch-up saccades. For mixed stimuli, like standard video films or the real world (i.e. which contain both static and dynamic items), the saccadic rate is not the same as fixation rate, and it remains unclear exactly what it signifies.

In the following examples from research, we select only studies that explicitly report saccadic rate, but a comparison to studies reporting fixation duration is advisable.

Mental workload, arousal and fatigue Saccadic rate decreases when task difficulty or mental workload increases (Nakayama, Takahashi, & Shimizu, 2002). This interpretation is utilized in several applied studies, for instance Pan *et al.* (2004), who found a difference in saccadic rate between two types of web pages. Also, arousal increases (Morris & Miller, 1996) and higher fatigue levels reduce saccadic rate (Van Orden *et al.*, 2000).

Visual imagery A task that requires visual imagery involving up/down or right/left movements increases saccadic rate significantly compared to visual imagery of inside/out (Demarais & Cohen, 1998), because the inside/out dimensions do not map as easily to stimulus space as up/down and right/left.

Disorders During smooth pursuit saccadic rate is higher for many psychiatric and neurological disorders, such as psychosis (Van Tricht *et al.*, 2010), and schizophrenia in general O'Driscoll and Callahan (2008).

The neural sites responsible for saccade programming have also been examined by taking saccadic rate measurements. For instance, Paus, Marrett, Worsley, and Evans (1995) found that with increasing saccadic rate, cerebral blood flow also increased in the frontal eye fields, the superior colliculus, and the cerebellar vermis.

12.2 Glissadic proportion

Target question	*How many glissades per saccade?*
Input representation	*Event-processed data*
Output	*The proportion of glissades*

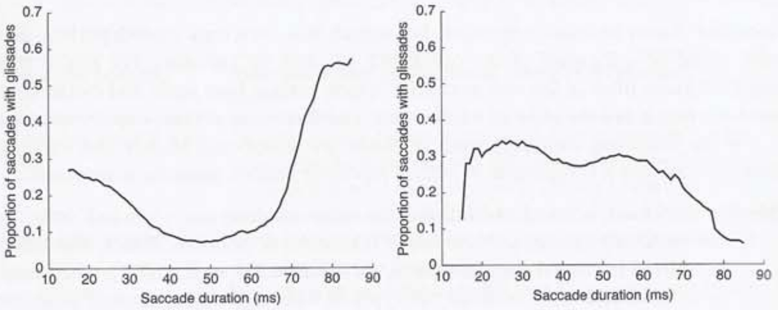
Glissades are always counted in proportion to the saccades which precede them. Although operational definitions measuring glissades differ considerably in the literature, and some authors and algorithms have even considered glissades to be measurement errors (p. 183), glissades are nevertheless real; moreover, they are also quite common (between 20 and 50% of all saccades end in a glissade (Nyström & Holmqvist, 2010; Wang, 1998)).

Studies have interpreted glissadic proportion in relation to:

Vigilance Wang (1998) interprets an increase in the percentage of glissadic saccades within anticipatory and return saccades (over time-on-task) as a possible index of vigilance decrement, and a similar interpretation is given by Ciuffreda and Tannen (1995).

Neurological diseases In the 1970s, glissades were used as an aid in diagnosing multiple sclerosis and vascular lesions (Bahill, Hsu, & Stark, 1978), and such neurological interpretations are reiterated by Ciuffreda and Tannen (1995, p. 39).

Saccade direction and amplitude Kapoula *et al.* (1986) found glissades to be overrepresented in the abducting eye, they also observed an over-representation of glissades in populations of saccades with short amplitude. In their much larger data set, with text and pictorial stimuli, Nyström and Holmqvist (2010) found no support for such a relation, neither in reading nor in scene perception.



(a) Reading (b) Scene perception

Fig. 12.3 Proportion of glissades for saccades made during reading and during scene perception, adapted from Nyström and Holmqvist (2010). Reprinted with kind permission from Springer Science + Business Media: *Behaviour Research Methods*, An adaptive algorithm for fixation, saccade, and glissade detection, *42*(1), 2010, Marcus Nyström and Kenneth Holmqvist.

Saccadic amplitude in reading Interestingly however, in reading data, the long (return sweep) saccades have a much higher percentage of glissades than equally long saccades during scene perception (see Figure 12.3). Nyström and Holmqvist (2010) speculate that this may be evidence that the oculomotor system adapts to the low velocities of the short saccades necessary when reading a line of text, but this adaptation does not spread to the programming of higher velocity return sweeps in reading. Conversely, in scene perception, where saccadic velocities are not bimodally distributed and much more random, the effect is rather the reverse.

12.3 Microsaccadic rate

Target question	*How often do microsaccades occur?*
Input representation	*Event-processed data*
Output	*Rate* (s^{-1})

Microsaccadic rate, or 'frequency' as referred to by Martinez-Conde and Macknik (2007), is defined as the number of microsaccades per second, and has the unit s^{-1} (Engbert & Kliegl, 2003). When calculating the value, only divide the number of microsaccades by the period spent in fixations. Dividing by the trial duration means including saccadic duration and underestimates the rate values. Note that detection algorithms for microsaccades may also detect square-wave jerks as microsaccades, and that—in some studies—a distinction may have to be made between the two types of events.

Although microsaccades occur during fixations only, they are virtually never counted in relation to fixations (i.e. microsaccades per fixation), the exception being Otero-Millan *et al.* (2008), who counted the number of microsaccades per fixation in order to plot the prevalence of microsaccades in connection with fixation duration.

Reported microsaccade rate values tend to range from 0.2–3 s^{-1}. Microsaccadic rate is one of the major measures in the research field which investigates the role of microsaccades in visual perception. Studies with this measure have focused on:

Rate over time Microsaccades have systematically been shown to have a *minimum* rate approximately 150 ms after cue onset, known as 'microsaccade inhibition'. The *maximum* rate of microsaccades occurs approximately 350 ms after presentation of the cue, known as the 'rebound effect' (see Engbert & Kliegl, 2003, and Rolfs *et al.*, 2005). In a task where participants scanned pictures, Otero-Millan *et al.* (2008) show that microsaccadic rate increases rapidly after around 300 ms; this is not the case when a blank screen is presented, however.

Visual illusions When a peripherally attended ring fades during maintained fixation due to neural adaptation (Troxler fading), this corresponds to a decrease in microsaccadic rate from around 3 to over $5\ \text{s}^{-1}$ (Martinez-Conde, Macknik, Troncoso, & Dyar, 2006). The microsaccadic rate decrease precedes the subjective experience of perceptual fading, and rate increases again just before the peripheral ring is once again visible. This highlights a functional role of microsaccades during fixation, in that they enable a stable perception. Another finding is that subjective experience of illusory motion in what is known as the Enigma painting coincides with an increase of microsaccadic rate (Troncoso, Macknik, Otero-Millan, & Martinez-Conde, 2008).

Intermediate trajectories Engbert and Mergenthaler (2006) propose that the individual microsaccadic rate can be predicted by what is known as the fractal dimension (Theiler, 1990) of the intermediary trajectories, known as inter-microsaccadic intervals (IMSIs), between microsaccades.

Manual response preparation Microsaccadic rates appear to be affected by (manual) response preparation processes as part of target discrimination (Betta & Turatto, 2006).

Much of microsaccade research is neurologically motivated, with the goal of revealing the function(s) of microsaccades in vision and brain areas which subserve visual perception.

12.4 Square-wave jerk rate

Target question	*How often do square-wave jerks occur?*
Input representation	*Event-processed data*
Output	*The number of square-wave jerks per minute* (min^{-1})

Square-wave jerk rate (SWJ-r), also known as 'saccadic intrusion rate' and 'multiple saccadic wave intrusion rate' (MSWI), is defined as the number of square-wave jerk pairs per minute, resulting in the unit min^{-1} (see p. 183 for a reminder of the square-wave jerk event).

Most studies report that the range for square-wave jerk rate is around 1–40 min^{-1}, with an average of around 10 min^{-1}, in healthy adult participants (Abadi & Gowen, 2004). Many studies report that all participants tested exhibit SWJs, although rates vary between individuals. Sixty per cent of the participants in Abadi and Gowen exhibited SWJ rates up to 20 min^{-1}, even though clinical textbooks such as Wong (2008) consider rates above 15 min^{-1} as pathological. As pointed out by Salman, Sharpe, Lillakas, and Steinbach (2008): "The prevalence and frequency of square-wave jerks in normal adults vary between studies, largely because of variations in the recording sensitivity, alertness, and arbitrarily selected saccade-amplitude threshold or duration for definition or detection of square-wave jerks, and also because of the varied periods of fixation sampled" (p. 18).

The three less known varieties, the biphasic square wave, the intrusion single saccadic pulse, and the double saccadic pulse, have frequencies of 0.5–5 min^{-1} (Abadi & Gowen, 2004). Since the threshold for detection is variable across studies, frequencies also vary.

The following factors have been shown to increase square-wave jerk rate:

Disorders Increased square-wave jerk rates are seen in a wide variety of disorders. This includes progressive supranuclear palsy (Troost & Daroff, 1977), strabismus (Ciuffreda, Kenyon, & Stark, 1979), multiple sclerosis (Doslak, Dell'Osso, & Daroff, 1983), Parkinsons (Rascol *et al.*, 1991), Friedrich's ataxia (Fahey *et al.*, 2008), and many more. According to clinician Wong (2008), square-wave jerks can also result from basal ganglia and cerebral cortical diseases. When square-wave jerks occur continuously, they are symptomatic of Parkinson's disease, alcoholic cerebral degeneration, and some forms of palsy.

Age Although all participant have some square-wave jerks, they are particularly common with elderly people. For instance, Herishanu and Sharpe (1981) found that the mean rate was 4.7 min^{-1} for young people (mean age 32), compared to a mean rate value of 27.0 min^{-1} for elderly people (mean age 71).

Increased catch-up saccade amplitude There is possibly a relationship between the amplitudes of catch-up saccades during smooth pursuit and the rate of square-wave jerks, according to Friedman, Jesberger, Abel, and Meltzer (1992), who speculate that programming both types of eye movements simultaneously taxes the saccade system heavily.

Tobacco and other drugs During the first 5 minutes after smoking one cigarette, the participants of Sibony, Evinger, and Manning (1988) exhibited an increased rate of square-wave jerks on both vertical and horizontal smooth pursuit. The effect is found both in schizophrenic patients and healthy controls (Thaker, Ellsberry, Moran, & Lahti, 1991). Drugs such as L-tryptophan (Baloh, Dietz, & Spooner, 1982), and intravenous opioids (Rottach, Wohlgemuth, Dzaja, Eggert, & Straube, 2002), have also been found to increase square-wave jerk rate.

Lower visual task demands Shaffer, Krisky, and Sweeney (2003) found significantly lower square-wave jerk rates when observers fixated remembered target locations compared to visual targets which were present. Smooth pursuit tracking of faster-moving and less predictable targets also reduced square-wave jerk rate.

A high square-wave jerk rate may jeopardize accuracy in data, as a square-wave jerk that occurs when a participant looks at a calibration point will cause miscalibration at that location (see p. 130).

12.5 Smooth pursuit rate

Target question	*How often do smooth pursuits occur?*
Input representation	*Event-processed data*
Output	*The number of smooth pursuits or catch-up saccades per second* (s^{-1})

There are very few studies which count smooth pursuits, simply because of the lack of algorithms which detect the smooth pursuit event. Indeed, this gap in the event detection domain means that smooth pursuit has a very uncertain status as an event per se. Often, instead of smooth pursuit, the complimentary measure number of catch-up saccades is used. With commonly used sinusoidal velocity stimuli, the number of catch-up saccades is approximately equal to the number of smooth pursuits minus one, as exemplified in Figure 12.4 and Table 12.4. Either way, rate (or 'frequency') measures are preferred to simply counting the number of occurrences. With stimuli like motion pictures or the real world, it is very likely that smooth

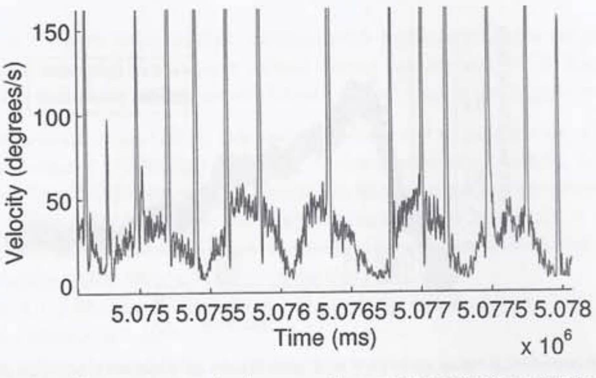


Fig. 12.4 Data collected from a person viewing a pendulum movement using a head-mounted 500 Hz system. A number of catch-up saccades interrupt the low-speed smooth pursuit. Values for relevant measures can be found in Table 12.4.

Table 12.4 Measurement values resulting from the data in Figure 12.4

Measure	Value
Number of catch-up saccades	14
Number of glissadic catch-up saccades	1
Number of smooth pursuits	15
Duration of this trial	3.6 seconds
Smooth pursuit rate	4.17 s^{-1}
Rate of catch-up saccades	3.89 s^{-1}

pursuits will not correspond to any other occulomotor event in terms of number (hence, using catch-up saccades to indirectly count smooth pursuits is unlikely to be a tenable alternative). This highlights the need for an algorithm which captures smooth pursuits in raw data samples for a variety of stimulus types.

The reader should also be aware that the rate of saccades within smooth pursuit (catch-up, back-up etc.) increases when the velocity of the stimulus target increases. Typical stimulus point velocities range from 10–40°/s.

Kremenitzer *et al.* (1979) measured the number of smooth pursuit segments per second on infants tracking smoothly moving targets, and found values of 0.07–0.38 s^{-1}. Smooth pursuits were counted by manual inspection. In this study, the number of smooth pursuits per second is a measure of the ability to keep tracking smoothly and continuously.

Measuring the rate of catch-up and back-up saccades during continuous smooth pursuit, Radant and Hommer (1992) found values of 0.62 s^{-1} for catch-up saccades, 0.16 s^{-1} for back-up saccades, and 1.04 s^{-1} for all forward-going saccades (these data come from their control group of "normal" participants). This rate measure is also frequently used in the study of schizophrenia, depression, and various other clinical disorders. Moreover, it is used to investigate the effects of medication and substance abuse, which tends to increase the rate.

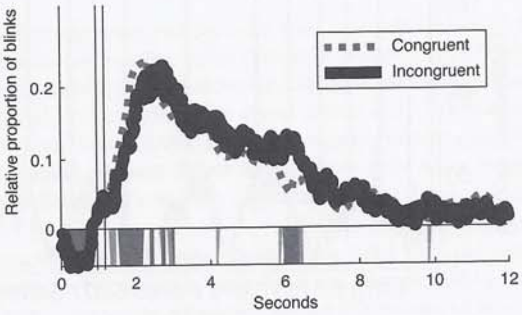


Fig. 12.5 A continuous measure of blink rate, also known as 'Relative proportion of blinks'. Vertical bars signify the onset of Stroop test trials, which could either be congruent or incongruent. The bar below the x-axis shows significance calculations between the congruent and incongruent trial types. Reprinted from Siegle *et al.* (2008) with kind permission from John Wiley and Sons: *Psychophysiology*, Blink before you think: Blinks occur prior to and following cognitive load indexed by pupillary responses, *45*(5), 2008, Greg J. Siegle, Naho Ichikawa, and Stuart Steinhauer, pp. 679–687.

12.6 Blink rate

Target question	*How many blinks?*
Input representation	*Event-processed data*
Output	*The number of blinks per minute* (min^{-1})

Blink rate ('frequency') is defined as the number of blinks per second or minute. Some researchers in ophthalmology prefer to use the inverse of blink rate, which may be called the 'inter-blink interval', which is defined as the duration between blinks in seconds (Montes-Mico, Alio, & Charman, 2005). Blink rate is a measure known to laypersons, who may ask "do people with lenses blink less?" and "do people blink more in front of a computer?".

Because blink rate only requires detecting the blink event, not measuring its amplitude, duration, and other properties, blink rate may appear the easiest blink measure to record with a video-oculographic eye-tracker. However, coverage of the pupil by reflection in glasses, and motion of the participant when using a remote eye-tracker causes a loss of pupil detection which many event detection algorithms may misinterpret as blinks. Such artefacts may be very common, as examplified by the histogram on page 325.

There appears to be a considerable variation in blink rate both between and within participants (Doughty & Naase, 2006), as well as over time for individual participants. For instance, Zaman and Doughty (1997) showed that participants exhibit large fluctuations in blink rate over time during a five minute recording. Clearly this can have important consequences for your results because the calculation of blink rate depends on *when* the data was recorded with respect to the length of the experiment. Moreover, the distribution of data may often be non-Gaussian; thus, alternative statistical methods may be necessary.

Blinks do not easily provide a continuous measure comparable to pupil dilation or proportion over time graphs, but Siegle, Ichikawa, and Steinhauer (2008) describe a method to calculate blink rate over time by comparing a short moving window of samples marked as belonging to a blink (or not) against a prerecorded baseline sample of individual blink rate. From this one can then produce graphs such as Figure 12.5.

The first studies of blink rate were conducted in the 1890s. Most authors, including those mentioned in the excellent overview of blink rate research from Stern *et al.* (1994), conclude that blink rate is a robust measure that increase with time-on-task and fatigue. Blink rate exhibits average values of 3–7 blinks per minute during reading and 15–30 during most non-reading tasks. The following effects on blink rate can be found in the literature:

Dry eyes Montes-Mico *et al.* (2005) conclude that the need to moisten the eyeball is not a major determinant of blinking. However, Montes-Mico *et al.* (2005), amongst others, report that " the typical inter-blink interval in normal patients is approximately 4 to 5 seconds and in those with dry eyes is approximately 1 to 2 seconds in normal conditions" (p. 1618). It is intuitive that a function of blinking is to keep the eyeball moist, and the results of the latter authors support this.

Air pollutants Air particulates, such as those resulting from cigarette smoking, increase blink rate (Stern *et al.*, 1994).

Contact lenses Blink rates initially increase when wearing lenses, but may return to base values following long-term wear (Collins, Seeto, Campbell, & Ross, 1989).

Monitors In normal healthy participants, blink rate drops on average by 20% during use of monitors (Patel, Henderson, Bradley, Galloway, & Hunter, 1991).

Time on task The vast majority of earlier studies (before 1990) were conducted using prolonged reading tasks, which could be up to six hours in length. Blink rate generally increases with time on task (i.e. length of time of reading), but intermediate pauses with comprehension questions eradicate this effect.

Time of day Blink rate appears to be higher during late afternoon than during morning (Stern *et al.*, 1994).

Mental workload Several studies associate the rate of blinks with mental workload (Wolkoff *et al.*, 2005). In a study where car drivers had simultaneous auditory tasks, Tsai, Viirre, Strychacz, Chase, and Jung (2007) found an increase in blink rate compared to the driving-only task. Van Orden *et al.* (2000) found that blink rate increased as a function of tracking error in a human factors tracking task. Brookings *et al.* (1996) showed that blink rate positively correlates with mental workload in air traffic controllers. However, Veltman and Gaillard (1996) conclude that blinking is independent of mental workload. This conclusion is based on physiological observations from a flight simulator study, where participants experienced in aviation had a concurrent auditory memory task, as well as the primary task of controlling and landing the simulated aircraft. Veltman and Gaillard (1996) found that during landing, when pilots had to process more visual information, the number and duration of blinks decreased, irrespective of the demands of the auditory task. The authors suggest this decrease reflects a strategy to maximize the time available for visual processing when visual demands are highest. It is possible the differences in findings for blink rate and mental/visual workload reflect the different types of task. One can easily envisage, for example, a situation where blinks are infrequent owing to maintained concentration when visual (and cognitive) demands are high. However when the requirements of the task exceed available cognitive resources, concentration may falter, and blink rate may increase due to mental workload.

Saccade amplitude Blinks are more likely to co-occur with large-amplitude than with small-amplitude saccades. Evinger *et al.* (1994) observed that activation of the lid-closing muscle occurred with 97% of saccadic gaze shifts larger than 33°.

Age Young children blink significantly less frequently than adults. When younger than two months, infants have almost no blinks at all (Stern *et al.*, 1994).

Psychoticism Participants with higher scores on the psychoticism scale showed higher blink rates (Colzato, Slagter, Van Den Wildenberg, & Hommel, 2009).

Individual differences Using a cluster analysis, Doughty and Naase (2006) group their participants into two groups: normal blinkers, with around 10 blinks per second, and frequent blinkers, with more than 20 blinks per second.

12.7 Fixations: number, proportion, and rate

Target question	*How many fixations?*
Input representation	*Event-processed data*
Output	*The amount of fixations in number, proportion or rate* (s^{-1})

Fixations are detected and their properties calculated by an event-detection algorithm. Fixations are then counted in the selected space, time, or portion of data. In practice, the number of fixations often refers implicitly to an entire trial (such as reading one page, or looking at an image), and in some cases to the number of fixations in an area of interest (AOI) during a trial. When comparing two numbers of fixations, it is important that space, time etc. are of equal size—if not, consider using fixation rate.

With still images, and if the event-detection algorithm is good enough, the number of fixations should equal the number of saccades ±1. With animated stimuli, smooth pursuit may occur, and the number of fixations may be lower than the number of saccades. Also, slow smooth pursuit may be taken to be fixations and artificially increase their number.

There is also an inverse relationship between number of fixations and fixation durations when the trial duration is constant: the more fixations, the lower the durations, and vice versa, as exemplified in Figure 12.6. There are also several other relationships between number of fixations and other measures, illustrated in Figure 12.6.

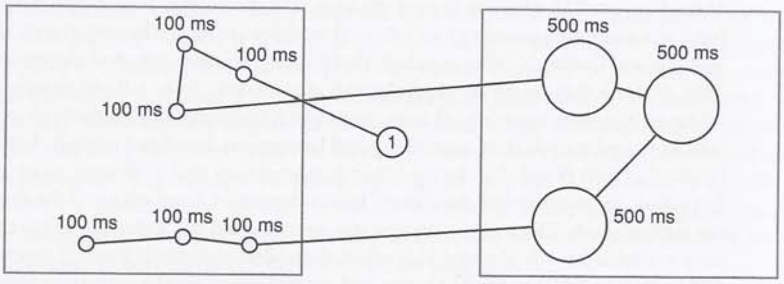


Fig. 12.6 Two areas of interest, left and right, and one scanpath starting at the central fixation (1). Trial duration is 2450 ms, of which saccades make up 300 ms and the initial central fixation is split so that 50 ms falls into this trial. Small fixation rings represent 100 ms fixations, and large rings 500 ms fixations. Measure values can be found in Table 12.5.

12.7.1 Number of fixations

When inside an area of interest, the number of fixations has also been called 'fixation density' (Henderson *et al.*, 1999). When counted over the whole of the stimulus, another term has been 'fixation frequency'—not to be confused with the measure fixation rate (p. 416) which sometimes goes under the same name. Number of fixations can also be counted per area unit, as an operational definition of reading depth (pp. 390–391). Number of fixations from onset until task completion or object detection is a latency measure (pp. 437–438).

Table 12.5 Values for a variety of duration and counting measures, derived from the synthetic data in Figure 12.6. The 'target' in post-target fixations is the selected area of interest. Stars indicate measures included within this chapter.

	Left AOI	Right AOI
Total dwell time (incl. saccades)	772 ms	1582 ms
Proportion of total dwell time	31.5%	64.5%
Number of fixations*	6	3
Proportion of fixations*	67%	33%
Number of dwells*	2	1
Number of fixations per dwell*	3	3
Number of post-target fixations*	6	3
Number of returns*	1	0
Fixation rate*	$2.4\ s^{-1}$	$1.22\ s^{-1}$
Dwell rate*	$0.83\ s^{-1}$	$0.41\ s^{-1}$

The number of fixations in an area of interest is a very general measure. It is correlated to total dwell time (p. 389), but is not the same, since number of fixations ignores fixation durations. The difference between dwell time and number of fixations will be particularly apparent if a number of short fixations in one area of interest (such as a text) is compared to a smaller number of long fixations in another area of interest (such as a face image). This is illustrated in Figure 12.6. Number of fixations is one of the most used metrics in usability research, according to the review in Jacob and Karn (2003), and it often appears in scene perception research, and in different forms in reading research.

Number of fixations has been used as an indication of the following:

Semantic importance Buswell (1935) and Yarbus (1967) present many scanpath visualizations which give the impression that the semantic importance of a region within a picture affects the number of fixations on that region. Many usability researchers (e.g. Poole *et al*., 2004, Jacob & Karn, 2003, Fitts *et al*., 1950) argue that the general importance or noticeability of an object increases the number of fixations in the area of interest allocated to that object. In later scene perception research, Henderson *et al*. (1999) and Loftus and Mackworth (1978), found that significantly more fixations landed on semantically informative areas.

Search efficiency and difficulty The number of fixations overall is thought to be negatively correlated with search efficiency (see, for example, Goldberg & Kotval, 1999, who identify a number of measures for assessing usability). Rötting (2001) posits that a low number of fixations could mean either that the task goal has been reached, that the participant is experienced, or that the search task is (too) simple. A high number of fixations would then be indicative of difficulty in interpreting the fixated information or layout. This interpretation is supported by Ehmke and Wilson (2007). In a survey by Jacob and Karn (2003) covering early usability studies and the measures they used, the number of fixations measure was found to be the most common, used in 11 out of 24 studies.

Experience A large number of studies employing various tasks report that experts have fewer fixations in their domain of experience (and sometimes this also transfers to other areas of visual skill). Here follows a few pertinent examples. More accurate circuit chip inspectors required fewer fixations to complete an inspection cycle (Schoonahd, Gould, & Miller, 1973). More experienced tin can inspectors have fewer fixations because they omit fixations on irrelevant parts of the tin can (Megaw & Richardson, 1979). In chess, it has been argued that experts have a larger visual span, therefore make fewer fixations than their less experienced counterparts (Reingold *et al.*, 2001). Fewer fixations is also true of reading skill, where proficient readers make fewer fixations than beginners, as well as other tasks where experts have faster processing of information conveyed in the form of high-resolution detail. However, there are some exceptions to this general rule: Kasarskis, Stehwien, Hickox, Aretz, and Wickens (2001) report that expert pilots have significantly more fixations than novices on all cockpit instruments, and Williams, Davids, Burwitz, and Williams (1994) found that experienced soccer players exhibit more fixations of shorter duration. This pattern of results is not contradictory however if one considers that an expert's fixations are more efficient. In Kasarskis *et al.* (2001), although more fixations were observed for experienced pilots, their dwell times were shorter than novices, suggesting that with greater experience information uptake is enhanced, and search optimally directed to relevant areas. This interpretation also holds for soccer players, and is supported in the literature on expertise differences in eye-movements more generally. Clearly however, the interaction between task and experience is important for the number of fixations measure, also explaining differences in results: beginning tennis players show more fixations towards the head of their opponent than do experts (Singer, Cauraugh, Chen, Steinberg, & Frehlich, 1996); and in golf putting, Vickers (1992) found that a low handicap group directed more fixations to the ball, whereas a high handicap group directed more fixations to the club.

Memory build-up For every new fixation upon an object in a room, memory for the position of that object will improve (Tatler, Gilchrist, & Land, 2005). This has the implication that fewer fixations will be required to locate objects once they have been encoded and a memory representation accumulated through multiple fixations.

Word properties in reading In reading, the number of fixations is used as a measure of morphological complexity, word frequency, and familiarity (Clifton *et al.*, 2007). For example, an area of interest containing a long composite word, or an unfamiliar word, receives more fixations.

Dysfunctions It has been frequently reported that dyslexic readers as a group tend to make more fixations than control readers. Participants with peripheral visual field defects make significantly more fixations than controls (Coeckelbergh, Cornelissen, Brouwer, & Kooijman, 2002). Macular pathology, the loss of foveal vision, also increases the number of fixations, at least during reading (Rohrschneider *et al.*, 1996). During oral reading, stutterers evidence more fixations and regressions per line than do non-stutterers (Brutten & Janssen, 1979). Phobias can result in fewer fixations on anxiety-inducing stimuli (such as spiders). This may reflect an avoidance coping strategy, particular to some phobic participants but not others (Pflugshaupt *et al.*, 2007). Hori, Fukuzako, Sugimoto, and Takigawa (2002) found that schizophrenic patients made a smaller number of fixations (as well as a smaller number of fixated areas) when looking at Rorschach pictures—these are some of the most well known ink-blot pictures used in assessment of personality and schizophrenia.

Age Older participants make significantly more fixations when scanning driving scenes (Ho, Scialfa, Caird, & Graw, 2001), although when their participants made passing manoeuvres in a simulator, Lavallière, Tremblay, Cantin, Simoneau, and Teasdale (2006) found the reverse effect: elderly drivers made fewer fixations.

Spectacles with progressive lenses When reading with progressive lenses, participants exhibit impaired reading in a number of measures, including an increase in the number of fixations (Han *et al.*, 2003). These multi-focal lenses have a gradient for people who are both near-sighted (Myopia) and far-sighted (Hyperopia). The gradient begins at the top of the lens, correcting for viewing objects in the distance, and progresses downwards to correct for close viewing, commonly reading. As a result the wearer is forced to find the best gaze direction through the variable glass, and if failing to do so, may have blurred vision where they read, and with more fixations and regressions as a result.

Sex Both sexes direct more fixations at female rather than male targets (Locher, Unger, Sociedade, & Wahl, 1993), which could explain the domination of female faces on the front pages of gossip magazines, which need to be seen to be sold.

Calculating the measure variety *number of fixations per dwell in each area of interest* (Figure 12.6), Gajewski and Henderson (2005), using a copy task where source and copy are the two areas of interest, argue that a small number of objects fixated per dwell in combination with frequent transitions from one area of interest to another indicates a strategy to minimize the use of visual working memory to one object at a time.

The *number of post-target fixations* for an area of interest is the number of fixations on other areas of interest, from the first entry of the area of interest until the end of the trial. Goldberg and Kotval (1999) who define the measure, speculate that it "can indicate the target's meaningfulness to a user. High values of non-target checking, following initial target capture indicate target representations with poor meaningfulness or visibility" (p. 643). After it was stated, this speculation has been repeated, but the measure itself has hardly been used. In fact, post-target duration (from leaving the area of interest the last time until the end of trial), or alternatively the number of post-target dwells, would probably be a better measure of how much a participant attends (or rather ignores) an area of interest.

12.7.2 Proportion of fixations

Proportion of fixations have been operationalized in one out of two ways.

The *first group* of operarationalizations compare number of fixations between areas of interest and between experimental groups. For instance, Adolphs *et al.* (2005) calculate the number of fixations on eyes in a face picture, divided by the total number of fixations made in the face. Behrmann, Watt, Black, and Barton (1997) found a significantly higher proportion of fixations on the ipsilesional right side of space for patients with unilateral neglect.

Other operational definitions commonly found in literature include a higher proportion of fixations 'on objects of the specified color' (of many colours), 'on salient regions' (rather than unsalient), 'on relevant pieces' (for the chess task), 'to the critical region' (in a nuclear control room), 'on the face' (in stimuli with people) etc. The areas of interest in focus can be given meaning by an uncontrolled action from the participant, as in the operational definition 'the proportion of fixations on the chosen alternative'. When comparing between areas of interest, the measure has also been called 'the ratio of on-target versus all-target fixations' by Goldberg and Kotval (1999).

Comparisons between participants involve examples such as finding the proportion of fixations made by the clinical group (53%) versus control participants (47%), over fixed trial durations. Commonly, studies use a combined experimental design where both participants and areas of interest are taken into account in the calculation of fixation proportion. Among many results, we find that "The larger visual span of experts in this [chess] task results in fewer fixations per trial, and a greater proportion of fixations between individual pieces, rather than on pieces" (Reingold *et al.*, 2001, p. 54).

The *second type* of operational definition utilizes the fact that fixations have properties that allow us to divide them into groups. The proportion of fixations with durations greater than 150 ms, for instance, or the proportion of fixations whose durations fall between 50 ms and 70 ms, and upward in 20 ms bins. Different fixation durations are argued to convey different cognitive processes (p. 377), which was used as a counting criterion by Schleicher, Galley, Briest, and Galley (2008), who argue that: "the proportion of fixations with a duration between 150 and 900 ms is associated with cognitive processing, while the remaining percentage of fixations are interpreted as express and overlong staring" (p. 15).

Dividing up fixations by their function, Land and Hayhoe (2001) reported that the majority of fixations (more than 50%) in studies on sandwich making, had task-relevant functions. There are many other examples where the *function* of fixations is used as a criterion.

The binocular disparity between fixations of each eye can be thresholded by the distance between characters in reading (and in principle other spatial cutoffs), and so Liversedge, White, *et al.* (2006) report that 47% of fixations have a disparity greater than one character space.

Other subgroups for dividing fixations into categories so that they can be counted and proportionalized include first fixations, such as the first ones in an area of interest, and ambient versus focal fixations. In reading literature, it is common to attribute saccadic properties to fixations, such as 'the proportion of fixations that were regressive', which should be read as 'the proportions of fixations following a saccade going in the opposite direction to the text'.

12.7.3 Fixation rate

Fixation rate ('fixation frequency', 'gazing time') is the number of fixations divided by a period such as the duration of a trial in seconds, giving the unit per second s^{-1}. The number of fixations in a period of time is roughly proportional to the inverse of the average fixation duration, but differs in two ways: first, fixation rate indirectly includes saccade and blink durations, smooth pursuit, and various other noise. Second, in recordings where there are a few very long and otherwise mostly short fixations, the fixation rate may be much lower than would be expected from the average fixation duration of the same data, as is the case in the study of a painter's eye-movements by Miall and Tchalenko (2001). This is because a small number of long fixations biases the mean fixation duration upwards, thus affecting the approximately inverse relationship between fixation duration and fixation rate.

Fixation rate values can also be divided into areas of interest. For instance, Colvin, Dodhia, and Dismukes (2005) compare pilots scanning in the cockpit, and find large differences in fixation rate within the different areas of interest, with some instruments very regularly hit by fixations, and others much more seldom.

Fixation rate is mostly used by reading, human factors, and usability researchers. Below are a few examples where fixation rate has been explicitly used as a measure in its own right.

Time on task Fixation rate is reported by McGregor and Stern (1996) and Morris and Miller (1996) to decline with time on task.

Task difficulty Fixation rate was found to be negatively correlated to task difficulty by Nakayama *et al.* (2002).

Performance quality McGregor and Stern (1996) found that fixation rate was correlated with performance only in the flight manoeuvre portion of their experiment, during which participants needed to focus attention on a greater number of cockpit displays than in the straight-and-level flight segments. Van Orden *et al.* (2001) found fixation rate to be variably correlated to tracking performance across participants, and showed that fixation rate is predictive of target density in a surveillance task, and that it could therefore be used as a measure of mental workload.

In reading research, fixation rate is a central measure used in the operational definition of reading speed (p. 330).

12.8 Dwells: number, proportion, and rate

Target question	*How many dwells?*
Input representation	*Area of interest-processed data*
Output	*The number, proportion, or rate* (s^{-1}) *of dwells*

The number of dwells is one measure in a whole family, some of which are illustrated in Figure 12.6 on page 412. Dwells are of course a different kind of events from fixations, so bear in mind that these measures refer to dwells.

12.8.1 Number of dwells (entries) in an area of interest

The number of dwells ('gazes', 'glances', 'visits', or 'transitions into and out of the area of interest') is counted over a limited time period which is usually a trial, but can also be a period ended with a participant action like selection or recognition. Precise counting criteria vary somewhat: most definitions allow for a new dwell (entry into an area of interest) to start directly after the previous dwell ended (gaze exited the area of interest). However, in their standard for car-driving research, ISO/TS 15007-1 (2002) require each new dwell in an area of interest to be separated by at least one dwell to a different area of interest. The two definitions only differ when there is whitespace in the stimulus, which is often the case in car-driving studies.

Figure 12.6 shows two areas of interest with whitespace in between. The left area of interest contains two dwells and the right area of interest one dwell. Table 12.5 shows measure values for these data and AOIs. As each dwell requires exactly one entry, the *number of dwells* and the *number of entries* are identical. Both of these are one more than the *number of returns* into an area of interest, that is

$$N_{dwells} = N_{entries} = N_{returns} + 1 \tag{12.1}$$

Number of dwells is a common measure in human factors and driving studies, which are often published in less accessible reports, PhD theses and conference papers. The measure has been given the following interpretations:

Semantic informativeness Although entering an area many times is a different behaviour compared to making many fixations in it, both are sensitive to semantic informativeness (Loftus & Mackworth, 1978; Henderson *et al.*, 1999).

Difficulty of instruments Presenting car drivers with the task of operating a simple and a difficult MP3 player, Chisholm, Caird, and Lockhart (2008) found many more dwells on the difficult player. The correlation to task completion time was high, see Figure 12.7. The safety limit used by several car manufacturers, based on the study by Zwahlen *et al.* (1988), is to design in-car instruments so that maximum three dwells are required until task completion. Above that level, the danger of drifting between lanes when driving increases rapidly.

Practice When submitted to a training programme in airport security screening, participants improved from originally on average 1.24 dwells on the target (a knife etc.) before recognizing it, to 1.08 dwells after four training sessions (McCarley, Kramer, Wickens, Vidoni, & Boot, 2004).

Animation Car drivers make a significantly lower number of dwells towards passive roadside advertisement signs (0.64 dwells per participant and sign) compared to animated signs (1.31 dwells per participant and sign) (Beijer, Smiley, & Eizenman, 2004).

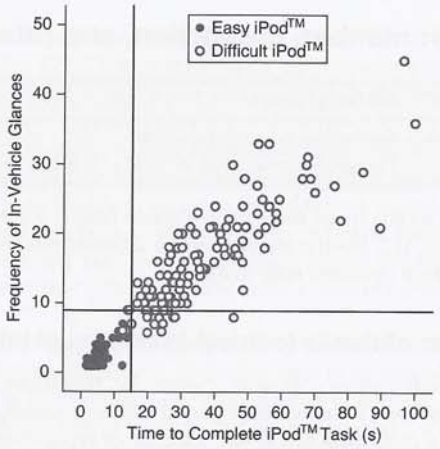
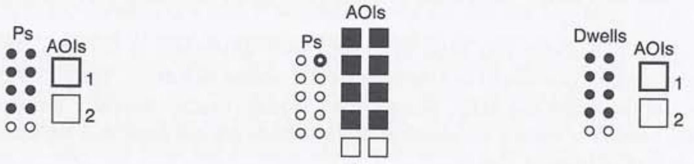


Fig. 12.7 Number (here called 'Frequency') of in-car dwells ('glances') towards an easy to manipulate MP3 player (iPod™) and difficult one during car driving. Reprinted from *Accident Analysis & Prevention*, S.L. Chisholm, and J.K. Caird, J. Lockhart, 704–713, Copyright (2008) with permission from Elsevier.



(a) 80% of the participants looked at AOI 1. Page 419.

(b) 80% of the AOIs were looked at, number of participants may vary. Page 421.

(c) 80% of the dwells were made to AOI 1. One or many participants performed the dwells. Page 418.

Fig. 12.8 Proportion calculations with participants (Ps), dwells, and AOIs. Note the variation not only in what is counted, but also whether participants, AOIs, and trial extent is specified or all-including.

Simulator versus real situation The number of dwells to in-car instruments is larger in field driving than in simulator driving, Wang *et al.* (2010) find in a validation study.

12.8.2 Proportion of dwells to an area of interest

The proportion of dwells tells us, for a single participant, or as an average over all participants, what percentage of the dwells made were directed at a specific area of interest; for example in Figure 12.8(c).

When the proportion of dwells to an area of interest is zero for a participant, the area of interest is considered totally skipped (see pages 192 and 419).

This measure is used infrequently in research, typically appearing in real-world studies with head-mounted eye-trackers, where dwells are calculated rather than fixations. Kuhn, Tatler, and Cole (2009) use this measure to show that when a magician looks away from the point in space where the trick takes place, participants follow the gaze of the magician and fail to see the trick.

12.8.3 Dwell rate

Dwell rate is the number of entries into a specific area of interest per minute. It is also known as 'gaze rate' (Fitts *et al.*, 1950). In particular when social and interactional scientists observe and estimate gaze of interlocutors without the use of eye-trackers (Vigil, 2009; Shrout & Fiske, 1981), 'gaze rate' is the most common term.

In the context of usability and human factors research, Jacob and Karn (2003) point out that the proportion of dwell time at a particular area of interest may indeed reflect the importance of the underlying area, but also that researchers using this metric should be careful to note that proportion of time confounds the rate of looking at an area of interest with the duration of those dwells. They quote Fitts *et al.* (1950) in that dwell rate and dwell time should be treated as separate metrics, with dwell time reflecting difficulty of information extraction, and dwell rate reflecting the importance of that area to the task.

Note that dwell rate is not equal to transition rate. While dwell rate is calculated *per area of interest*, transitions rate is often an overall measure *for two or more areas of interest*, and sometimes a measure counting *pairwise transitions between two areas of interest*.

The dwell rate measure is remarkably little used in eye-tracking research. These are two examples:

Drugs When driving, alcohol but not marijuana leads to decreased dwell rate to critical areas (Moskowitz *et al.*, 1976).

Preference task: Selecting a photo among many based on subjective preferences ('Do I like it?') requires fewer dwells per minute on average than selecting the photo on the basis of subjective recency ('Is it recently taken?') (Glaholt & Reingold, 2009).

12.9 Participant, area of interest, and trial proportion

Target question	*What proportion of participants, areas of interest, or trials fulfil a certain criterion?*
Input representation	*Data from multiple participants and trials*
Output	*The proportion of participants, areas of interest, or trials that fulfil a specified criterion*

Participants, areas of interest, and trials are not events, and should not have to be counted using the same principles as counting eye-movement data. Researchers do know how many participants and trials they have recorded, how many areas of interest they have specified etc, as it is defined in their experimental design (Chapter 3). Thus, these are not dependent variables that we wish to find out the number of.

Nevertheless, many researchers count both participants, areas of interest, and trials as part of their data analysis. This is because *it can be revealing to quantify the number of times a particular criterion is fulfilled* with respect to the experimental design in question and the eye-movement data it produces.

12.9.1 Participant looking and skipping proportions

As researchers know how many participants they have recorded, the absolute number is not of much interest. Instead we calculate the proportion of participants that fulfil a criterion which usually relates to a specific area of interest (Figure 12.8(a)). We might find, for instance, that 25% of the participants looked at the brand logo in an advertisement, or 94.5% skipped the high frequency word 'of'. More complex counting criteria are the proportion of participants that leave a target word with a regression back to prior text, and the probability that a participant will make more than two fixations on a specified word (see Pollatsek & Hyönä, 2005).

The *proportion of participants that look at an area of interest* is a very different measure from number of dwells and dwell time. A 100% proportion of participants for an area of interest only means that every participant must have looked at the area of interest once, but says nothing about how long, since a very short dwell time is enough to be included within the proportion calculation. Other areas of interest may have received many dwells or long dwell times from only some of the participants, which make them high in number of dwells and dwell time, but low in proportion of participants.

Percentage of participants fixating on areas of interest is interpreted by Albert (2002) and Jacob and Karn (2003) as a sign of attention-grabbing properties of an interface element. This interpretation was picked up very early in advertisement studies such as Treistman and Gregg (1979), who quantify the proportion of participants noting each primary ad element. Poole and Ball (2005) add that if a low proportion of participants is fixating an area that is important to the task, it may need to be highlighted or moved.

Participant *skipping proportion* is, for a selected area of interest, the proportion of all participants who never looked at the area of interest (total skipping proportion), or the proportion of readers who skipped the word area of interest when first encountering it (first skipping proportion).

In reading research, skipping proportion is often called 'skipping rate'. Skipping rate is one of eight selected reading measures in the review by Inhoff and Radach (1998), and many other reading researchers point out its importance. The reported skipping proportion for words in sentences constructed for experiments ranges from 0–80%. The important theoretical questions revolve around whether words are skipped because they were identified parafoveally during the previous fixation, and the extent to which lexical identification of the skipped word can be achieved from that previous fixation position. Four factors have been investigated:

High versus low frequency Several studies have used identical sentence pairs with a high-frequency word in one and a low-frequency in the other. For instance, Rayner and Fischer (1996) compared sentences like "He invested his money to build a store and was soon bankrupt" with sentences like "He invested his money to build a wharf and was soon bankrupt". Schilling, Rayner, and Chumbley (1998) report that a low-frequency word (1–10 occurrences per million—opm) has a 10% probability of being first skipped, while a high-frequency word (more than 10,000 opm) will be skipped 67% of the time.

Word length The shorter a word is, the more often it will be skipped (Inhoff & Radach, 1998).

Contextual predictability Given the sentence "The man decided to shave his ____", participants are much more likely to fill in "beard" (83%) than "chest" (8%) (Rayner & Well, 1996). When reading such sentences with the contextually predictable word filled in, that word is also more likely to be skipped than a contextually unpredictable word in the same sentence.

Writing system In Chinese, skipping proportion of a single character is reported to be between 40–80% (Yang & McConkie, 1999). Readers of Chinese even skip two or three characters, although the skipping proportion is then lower.

12.9.2 Proportion of areas of interest looked at

The area of interest proportion measure tells us by reviewing all the areas of interest shown to a single participant, how many of the areas of interest were looked at (Figure 12.8(b)), or for that matter how many were skipped. This is a measure used in traffic research, where drivers are shown videos of more or less hazardous situations, and researchers then count the percentage of pedestrians, cyclists, grandmothers with baby carriages, and other potentially dangerous things outside the vehicle (even a suddenly appearing moose has been used as a hazard in such studies!) which were fixated. The fixated proportion can then be related to the position, type, or movement direction of the target, or to the level of blood alcohol of the driver.

For instance, Underwood, Chapman, Berger, and Crundall (2003) found that 83.1% of central hazardous events were looked at, but only 66.8% of central non-hazardous events. Further comparisons reveal that central events were looked at more often (74.9%) than 'incidental events' (49.8%), and that dynamic events (69.2%) were looked at more often than static hazards (55.6%). Measuring how much drivers later recall objects in these scenes, (Underwood, Chapman, Berger, & Crundall, 2003) found that "fixated objects were recalled on around 50% of occasions, but details about non-fixated objects were recalled approximately 20% of the time." (p. 301)

Researchers using this proportion measure often warn against interpreting non-fixated as non-processed. In addition to driving research, the same observation has been made in psycholinguistic research, where for instance Griffin and Spieler (2006) point out that people often speak about objects in a scene although the objects were not fixated, and are able to include detail such as agency (actor versus receiver), relative humanness (human versus other primate), and animacy (animate versus inanimate).

12.9.3 Proportion of trials

Trials are counted like participants: each trial is identified using a criterion from the eye-movement data within it, for example, how many saccades are made during the trial, or how many transitions are made, etc.

Reingold *et al.* (2001) counted the proportion of trials where participants could evaluate the positions on a chess board without making any eye movements, reporting 15.9% of trials for experts, 2.6% for intermediate players, and 1.6% for novices. Using data from a visual search task, Gilchrist and Harvey (2000) first counted the number of saccades in each trial; then for each such number of saccades, the number of trials with that number were counted. The resulting histogram in Figure 12.9 shows that there were more trials with many saccades in which the target was absent compared to when the target was present.

In their problem solving task, Peebles and Cheng (2002) divide trials on the basis of the number of transitions between question and graph areas of interest. 37.1% of the trials involved only one transition from the question to the graph, 48% involved two such transitions, 11.9% involved three transitions and 2.8% involved four or more transitions. The one-, two-, three-, and four-transition(s) trials did not differ in how long participants took to read the elements in the question, but the strategies used did differ: "in the 1 transition trials, participants took approximately 2.28 s to scan the graph before entering an answer, whereas in the other trials, participants looked at the graph for a shorter time before looking again at the question for approximately 300 ms. In the two- and three-transition trials, participants then looked at the graph again for over a second before entering an answer or looking for a third time at the question, respectively. Participants in the three-transition trials scanned the graph for a third time before entering an answer". (p. 82)

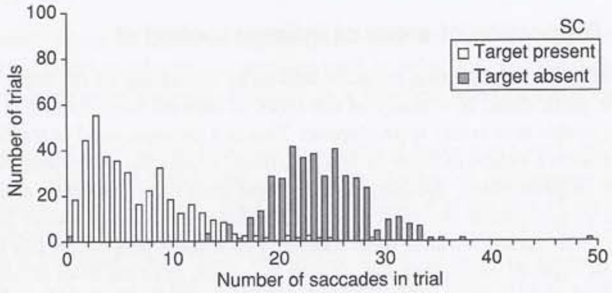


Fig. 12.9 Number of trials as a function of the number of saccades made in the trial. Data from a single participant (SC) who performed visual search trials with the target either present or absent. Reprinted from *Current Biology, 10*(19), Iain D Gilchrist and Monika Harvey, Refixation frequency and memory mechanisms in visual search, 1209–1212, Copyright (2000), with permission from Elsevier.

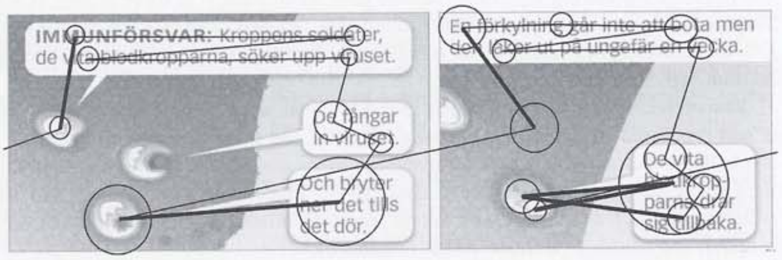


Fig. 12.10 A participant reading a piece of information graphics describing an attack of viruses inside the nose, and the response of the immune system. Bold saccades are *integrative* in the sense of bringing the reader between content-identical visual and textual information. Recorded 2005 with a 50 Hz head-mounted system with magnetic headtracking. Events are counted in Table 12.6. Data are described in Holsanova *et al.* (2008).

12.10 Transition number, proportion, and rate

Target question	*How many transitions?*
Input representation	*Area of interest-processed data*
Output	*The number of transitions, their proportion or their rate* (s^{-1})

Transitions are movements between areas of interest, and are counted for pairs of areas of interest, for instance the number of transitions from the mouth to the eyes in a face stimulus. Transition matrix measures are used to study overall transition behaviour and longer sequences of transitions (p. 339). The number of transitions measure concerns counting and comparing transitions in small numbers of pairs.

12.10.1 Number of transitions

The number of transitions between two areas has been used as a measure to evaluate the following:

Design Holsanova *et al.* (2008) compared reading of identical information graphics material in a radial versus a serial design, finding that the serial design resulted in a larger num-

Table 12.6 Number of integrative and non-integrative saccades shown in Figure 12.10.

	Left box	Right box
Number of saccades	10	12
Number of integrative saccades	2	4

ber of transitions between content-identical text and graphics areas of interest. Stolk and Brok (1999) report that the number of transitions between text and picture is related to the complexity of their stimulus material. Hannus and Hyönä (1999) often found only one transition from text to visualization at the end of school textbook reading, and that high-ability children were better at utilizing the relationship between text and visualization.

Expertise Expert pilots make a larger number of transitions between the runway and the airspeed indicator compared to novices, which indicates that expert pilots are following their training and making more of the runway, Kasarskis *et al.* (2001) note. Kramer and McCarley (2003) suggest that instruments with a large number of transitions between them should be located in close proximity. Studying computer menu search, Card (1982) argues that *conceptual chunking* of menu items can be seen in eye-movement data as sequences of short-amplitude transitions to spatially neighbouring menu items, interrupted by longer transitions.

Importance of an area When an area is so important that it must be more or less continuously monitored, transitions rates drop. Thus Vatikiotis-Bateson *et al.* (1998) found that the number of transitions between eye and mouth areas of the face is decreased in the presence of noise during a speech task, while dwell time on the mouth is increased.

12.10.2 Number of returns to an area of interest

Returns ('refixations', 'rechecks') are a specific type of transition into an area of interest. To count as a return it is required that there has been at least one previous dwell in the area of interest. In Figure 12.10, the only returns are made to the text at the bottom right, and to the white blood cell that it points to. Note that the number of returns to an area of interest always equals the number of dwells in the area of interest minus one (Figure 12.6). There is at least one proportion variety of this measure: Mannan *et al.* (2005) defined 're-fixation proportion' for each participant as the total number of re-fixations divided by the total number of targets fixated. A number of effects on this measure have been observed in the literature:

Semantically informative areas Returns are interesting for the same reason as number of dwells: they show that the area is semantically informative. The results that semantically informative and interesting areas are often looked at (many dwells) largely results from many returns in those areas.

Memory Amnesia patients make fewer returns to manipulated areas in a scene, compared to controls, which Ryan, Althoff, Whitlow, and Cohen (2000) take to indicate a deficit in relational (declarative) memory processing.

Need to confirm In eye tracking on radiological images, the number of returns to a lesion is taken as a measure of confirmatory scanning before the final judgement as to whether it is a malign structure or not (Mello-Thoms *et al.*, 2005).

Age In a change detection task with simple visual patterns, older participants made more returns to already visited areas. Veiel, Storandt, and Abrams (2006) speculate that more

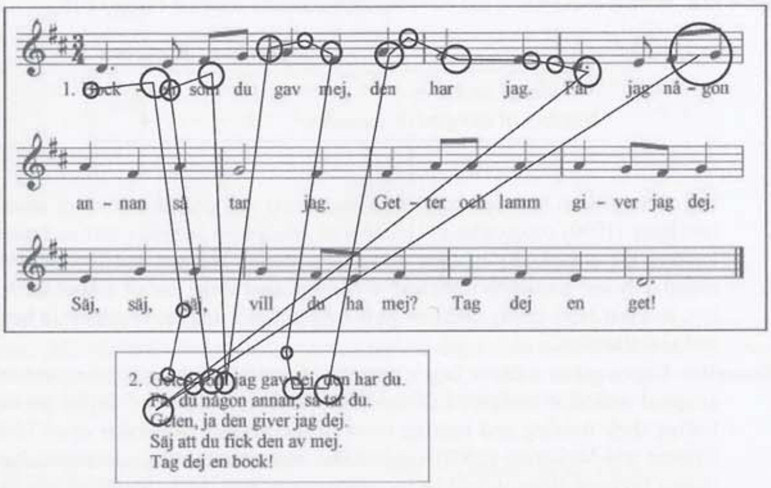


Fig. 12.11 Transition rate in prima vista singing; participants had not seen scores or lyrics before, and had to start singing directly. The division of lyrics at the bottom and music score at the top forces the singer to move up and down in regular intervals, giving a transition rate of 1.48 s^{-1} on average. Recorded 2001 with the an older remote system.

returns indicate a need to refresh visual working memory more often. Gilchrist, North, and Hood (2001) find that the number of returns to an area of interest is much lower if the cost for returning is increased, which increases memory load.

12.10.3 Transition rate

Transition rate is defined as the number of area of interest transitions—movements from one area of interest to another—per second. Transition rate can be calculated *for single pairs of areas of interest over all participants*, in which case transition rate equals dwell rate for the areas of interest. As there are often more than two areas of interest in a stimulus, there can be different transition rates for different pairs of areas of interest. A transition rate matrix has rate values in the cells rather than count values, and can be easily constructed by dividing all transition matrix values by the trial duration. Alternatively, transition rate can be calculated *for participant groups or conditions over all areas of interest.*

Nobody has purported a general interpretation of a high transition rate, but the measure has been used in two broad types of studies:

Working memory capacity Berséus (2002) use transition rate, in combination with dwell time, to study memory buffers of singers who have music and text information separate. Figure 12.11 shows typical data. Results showed that these choir singers had to make 1.48 transitions per second to perform smoothly. Thus, transition rate could be a measure of the demand on (visual) working memory buffers in repetitive tasks. Similarly, Miall and Tchalenko (2001) recorded the eye (and hand) movements of a portrait artist making a pencil sketch of a model. The regular alternation between sitter and drawing gave a transition rate of about 0.5–0.7 s^{-1}.

Integration between modalities There is no consensus that a higher transition rate between two modalities implies better integration between the two information sources. Yet the transition rate measure has often shown effects in studies with multimodal stimuli. For instance, Schmidt-Weigand, Kohnert, and Glowalla (2010) computed the transition rate between text and visualization, after initial instructions varying between fast, medium, and slow pace. Learners in medium and slow instruction pace exhibited more transitions per second between text and visualization compared to the fast instruction pace. Bartels and Marshall (2006) used the measure in a study of the effect of colour codings in displays used by air traffic controllers.

12.11 Number and rate of regressions, backtracks, look-backs, and look-aheads

Target question	*How many regressions?*
Input representation	*Event- or area of interest-processed data*
Output	*The number of regressions, backtracks, look-backs, or look-aheads*

There are hundreds of journal papers that have investigated reading processes by counting the number of regressions. There are two recurring themes: first, dyslexia leads to an increase in regressions. Second, 'garden-path' (leading) sentences and other sentences which require search for an antecedent are common stimuli used with the regression measure.

Compared to regressions, backtracks, look-backs and look-aheads are much more seldom counted in published research.

12.11.1 Number of regressions in and between areas of interest

As we have noted on pages 263–264, the regressions *inside words* are thought to reflect *lexical activation processes* (understanding the word), while regressions *between words* reflect *sentence integration processes* (understanding how several words relate). See Chapters 4 and 5 in Underwood (1998) for further discussion of this difference.

When studying reading of texts longer than one sentence, what is known as the sentence wrap-up effect may cause more regressions near sentence borders.

Spelling errors Zola (1984) found that introducing spelling errors into text resulted in an increase in the number of regressions (in addition to an increase in the fixation duration at the error). When Yang and McConkie (2001) showed even more deteriorated texts, where letters are replaced with various non-text patterns, the number of regressions greatly increased.

Reader proficiency It has been argued that better readers have a spatial coding of the text that allows them to resolve anaphoric expression by executing a single long saccade to the antecedent (Murray & Kennedy, 1988). Readers without such a spatial coding instead must search for the antecedent using backtracking.

Visual disorders Several visual disorders increase the number of regressions, for instance macular pathology (Trauzettel-Klosinski, Teschner, Tornow, & Zrenner, 1994; Rohrschneider *et al.*, 1996). Training may help in some cases, as Reinhard *et al.* (2005) show by giving participants with homonymous visual field defects visual restitution training.

Other disorders Many studies have found that readers with dyslexia have more regressions (Elterman, Abel, Daroff, Dell'Osso, & Bornstein, 1980; Rayner, 1985). Also

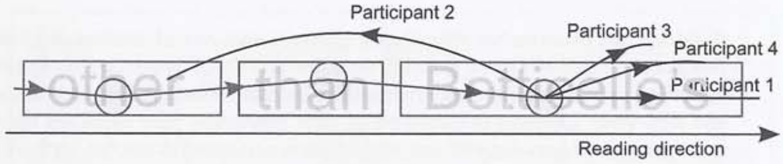


Fig. 12.12 The definition of regression-out is exemplified by these four participants reading. After the fixation of 'Botticello's', there is a 25% regression-out proportion, provided by participant 2.

Alzheimer patients make more regressions compared to controls (Lueck *et al.*, 2000), and the number of regressions in patients with Broca's and Wernicke's aphasia has been found to be approximately two or threefold the mean value in a normal group (Klingelhöfer & Conrad, 1984).

Age Throughout the primary school years reading skills improve, and the number of regressions continue to decrease up to late university years. However, Rayner, Reichle, Stroud, Williams, and Pollatsek (2006) claim to have found that higher age again increases the number of regressions.

12.11.2 Number of regressions out of and into an area of interest

The reading measure *regression-out* is the proportion of exits from a [word] area of interest that go backward in the reading direction, relative to all exits. This measure is particularly used in first-pass reading (Clifton *et al.*, 2007). Figure 12.12 shows how three participants leave the word area of interest "Botticello's" in the reading direction, while one participant makes a regression-out. The proportion of regression-out is therefore 25%. The regressions-out measure is indicative of both early and late processing.

The *regression-in* measure is for an area of interest the proportion of participants that have regressed to that area of interest, or the proportion of trials where regressions happened. Pollatsek and Hyönä (2005) use the terms 'probability of refixating target word', and 'probability of refixating target word at least twice'.

Both these reading measures are used in branches of reading research. For instance, Pollatsek and Hyönä (2005) compared sentences with fully opaque compounds such as "humbug" (where the meaning of neither constituent is related to the meaning of the compound) to sentences with fully transparent ("milkbottle") and partially opaque compounds such as "strawberry" (where the meaning of one of the constituents is related to the meaning of the compound). The number of regressions from opaque words was almost double the number of regressions from transparent words.

12.11.3 Regression rate

Regression rate is counted as the number of regressions per second, per 100 words, per line, sentence, or paragraph. The measure is used to compare reading processes between conditions such as blood alcohol levels, paragraph width, and type of glasses. For instance, Han *et al.* (2003) found that using progressive spectacles when reading results in a higher regression rate than single-vision spectacles. Beymer *et al.* (2005) found regression rate values of 0.54 s^{-1} for text with paragraph width 9 inches, while it was only 0.39 s^{-1} for paragraph width 4.5 inches.

12.11.4 Number of backtracks

Backtracks are notoriously ambiguous events, as explained on pages 262–264. Count values in the literature rely on varying operational definitions, and cannot be easily compared.

Nevertheless, the number of backtracks was found by Goldberg and Kotval (1999) to be one of the best predictors of usability. However the studies that use this measure (and different operational definitions of backtracks) to investigate webpage use, appear not to have been able to replicate this finding.

12.11.5 Number of look-aheads

Look-ahead fixations tend to cluster around 3 s ahead of the action, which is taken as support for the view that they are anticipatory and preparatory for the upcoming action.

Mennie *et al.* (2007) compute number of look-aheads in a grasping task. Look-aheads prove to be much less common than guiding fixations, and the look-ahead fixation is also shorter in duration. Finding no decrease in the number of look-aheads over the one hour sessions, Mennie *et al.* conclude that look-aheads are not influenced by familiarity with the scene layout.