Available online at www.sciencedirect.com
SCIENCE DIRECT®
Vision Research 45 (2005) 2298–2309
Vision Research
www.elsevier.com/locate/visres

The mathematics of multiple object tracking: From proportions correct to number of objects tracked

Johan Hulleman [a,b,*]

[a] Department of Psychology, University of Hull, Hull HU6 7RX, United Kingdom
[b] School of Psychology, University of Birmingham, Birmingham B15 2TT, United Kingdom

Received 26 April 2004; received in revised form 7 February 2005

Abstract

This paper examines the relation between proportions correct responses and the number of items tracked in multiple object tracking (MOT). It analyses two of the principle methods used in MOT. The *mark all* method, where the participants have to mark all the items, is shown to be equivalent to sampling without replacement. For the *probe one* method, where participants have to indicate whether a particular item belongs to the target set, formulas are derived as well.

The paper shows that it is not possible to determine the tracked number of target items (m) and distractor items (v) from the proportions correct answers when employing only one of these two methods. A combination of the *mark all* and *probe one* methods does not yield a unique relation between the proportions correct and m and v either, because of the interchangeability between tracking targets and tracking distractors.

Keywords: Multiple object tracking; Multielement visual tracking; Visual attention

Multiple object tracking (MOT) was introduced by Pylyshyn and Storm (1988). In a typical MOT-trial, participants look at a display containing eight items randomly positioned on the screen. During the definition phase, a number of items (up to four) are flashed, indicating that these are the items that the participants are supposed to track during the subsequent period of motion. During this tracking phase, which can last up to 15 s, the items move randomly across the screen. In some versions of the task, the items are prevented from occluding each other (e.g. Pylyshyn & Storm, 1988), in other versions the items are allowed to overlap (e.g. Viswanathan & Mingolla, 2002). After the tracking phase there is a test to determine to what extent the participants have managed to follow the targets. This test phase can take two different forms. In the first, the *mark all* method (used in Scholl, Pylyshyn, & Feldman, 2001; vanMarle & Scholl, 2003), the participants are invited to mark all the targets items that were flashed during the definition phase. In the second, the *probe one* method (used in Yantis, 1992), one of the items on the screen is probed, and the participants have to indicate whether the probed item was flashed at the start of the trial. When repeated many times, both methods will yield a proportion correct. In the *mark all* method, this proportion will be the number of items that were correctly marked divided by the total number of target items that should be marked. In the *probe one* method, it is the number of correct responses divided by the number of trials. The proportion correct is subsequently used to make inferences about how many target items have been tracked successfully.

The *mark all* and *probe one* methods are not the only ways of studying tracking behaviour. Recently, Tripathy

* Tel.: +44 1482 465448.
E-mail address: j.hulleman@hull.ac.uk

0042-6989/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.visres.2005.02.016

and Barrett (2004) introduced a paradigm where participants had to detect a deviation in the linear trajectory of one of the items that were moving across the display. Their task was to indicate whether the deviation was clockwise or counter-clockwise. Tripathy and Barrett (2004) found that the deviation necessary to correctly identify the direction of the deviation increased dramatically when the participants had to track more than one item.

The crucial difference between the Tripathy and Barrett method on the one hand and the *mark all* and *probe one* methods on the other is that the former involves a distinction *within* a set of target items, whereas the latter involve a distinction *between* a subset of targets and a subset of non-targets. In the *mark all* and *probe one* methods, the participants have to be able to identify all the targets, whereas in the Tripathy and Barrett method they report an attribute (change in direction) of only one of the target items.

In the Tripathy and Barrett method the participants have to track all the items, because they all might be the target, until the deviation in the trajectory of one of the target items occurs. After this event, and after they have reached their decision about the direction of the deviation, they do not need to track any of the items anymore. In the *mark all* and *probe one* on the other hand, only a subset of the items on the screen needs to be tracked, but it needs to be tracked throughout the entire trial. So, the number of task-relevant items in the Tripathy and Barrett method drops during the trial, whereas it stays constant for the *mark all* and *probe one* methods.

The formulas for the expected proportion correct derived in this paper describe expected performance for tasks that involve the distinction between targets and non-targets. They are not valid for the case where an attribute of one of the targets has to be reported. They do therefore only apply to the *mark all* and *probe one* methods. In these two methods, as will be shown later in this paper, knowing which items are non-targets will help the participants in correctly identifying the targets. This is not the case in the Tripathy and Barrett method. When the deviating item is not cued beforehand, knowing that the item(s) that were tracked did not undergo the change in direction will not help the participants in identifying the change of direction of the target item they did not track.

Under the assumption that participants only track a subset of items, what is the expected proportion correct for the *mark all* method? For the case where the number of targets n_T equals the number of distractors n_D Scholl et al. (2001) derived the following formula, where m denotes the number of targets successfully tracked, p is the proportion correct, and n is the number of targets (in total there are $2n$ items, because there are also n distractors):

$$p = \frac{1}{2}\left(\frac{m}{n} + 1\right) \quad (1)$$

rearrangement of (1) yields m, the number of targets successfully tracked:

$$m = n(2p - 1) \quad (2)$$

The line of reasoning behind (1) and (2) is the following: participants are able to follow m target items. Whenever m is smaller than the total number of target items, they will have to guess the remainder. Because half of the items on the screen are targets, the rate with which the participants guess correctly will be 0.5.

Using (2), Scholl et al. (2001), were able to convert the proportions correct that they observed in their experiments into the effective number of items tracked (p. 174). Moreover, by using (1) and solving it for $m = 1$, they determined which proportion correct corresponds with the situation where the participants have lost the ability to track multiple items, and are only able to follow a single target item. This method of estimating the proportion where participants have lost the ability to track more than one item has been used elsewhere as well (e.g. vanMarle & Scholl, 2003).

Unfortunately, (1) and (2) are only valid when the chance of picking a target item remains constant, as in sampling with replacement.[1] However, this is not typically the case when participants have to mark all the targets. Whenever they have marked an item, they do not get the opportunity to mark that same item again. Rather, they will always have to mark n different items. This is equivalent to sampling without replacement. To see how this influences the estimate of m, we will take a closer look at what happens when the participants are prompted for a response in a *mark all* trial where the number of targets n_T equals the number of distractors n_D.

Suppose that the participant has not managed to track a single target. It is as if the participant was not in the room during the definition phase, but is asked to mark the target items nonetheless. In this case, the participant does not possess any knowledge whatsoever about the test display, and will have to guess which items are the targets. Clearly, because half of the items are targets, the expected proportion correct will be 0.5.

A more circuitous way of arriving at this proportion correct is by following the formal route of combinatorics. The participant has to pick the n_T targets out of the $(n_T + n_D)$ items on the screen. The expected proportion correct depends on the probability of marking $0, 1, \ldots, n_T - 1, n_T$ target items when n_T items are picked

[1] Formula (1) actually does provide the expected proportion correct for the Tripathy and Barrett method, with m representing the number of items successfully tracked, and n the total number of items in the display.

from the total of $(n_T + n_D)$ items on the screen. The probability of picking b target items is given by

$$p = \frac{\binom{n_T}{b}\binom{n_D}{n_T - b}}{\binom{n_T + n_D}{n_T}} \tag{3}$$

because there are

$\binom{n_T}{b}$ ways of picking b target items from the total of n_T target items,

$\binom{n_D}{n_T - b}$ ways of picking the remaining items from the n_D distractor items,

$\binom{n_T + n_D}{n_T}$ ways of picking n_T items out of a total of $n_T + n_D$ items on the screen.

Sampling without replacement (as embodied in (3)) follows the hypergeometric distribution. The mean number of target items that will be marked when picking randomly is therefore given by the mean of this distribution:

$$\frac{n_T^2}{(n_T + n_D)} \tag{4}$$

Consequently, the expected proportion correct, in an experiment where participants are marking targets randomly is

$$p = \left(\frac{n_T^2}{(n_T + n_D)}\right) \Big/ n_T = \frac{n_T}{(n_T + n_D)} \tag{5}$$

For the special case where $n_T = n_D = n$ we can see that the expected proportion correct will be 0.5.

When participants have only managed to follow a single target item, they will mark this target and guess the $n_T - 1$ remaining target items. More generally, when the participants have followed m target items, they will only have to guess $n_T - m$ target items. So the expected proportion correct when they have successfully tracked m target items will be

$$p = \left(m + \frac{(n_T - m)^2}{(n_T - m + n_D)}\right) \Big/ n_T \tag{6}$$

There is one more issue to consider. It might be the case that the participants also have knowledge about the distractor items. For instance, they might know that one of the items in the test display is in fact a distractor. If this would be the case, they could bring this knowledge to bear and increase their chances of correctly marking the target items that they have to guess. When participants know that v items on the screen are distractor items, this means that they will not pick from among these. In effect, they are reducing the number of distractor items that could be picked by v. In this case, the expected proportion correctly marked target items is given by

$$p = \left(m + \frac{(n_T - m)^2}{(n_T - m + n_D - v)}\right) \Big/ n_T \tag{7}$$

In the numerator of (7), m counts the amount of information that participants have over the target items in the test display. The other part of the numerator of (7) provides an estimate for the expected number of correctly marked targets for the part of the display that the participants know nothing about. This is the part of the display where the participants are guessing. This is identical to the state of affairs described by (5), except that now, the number of target items that are still present in this part of the test display is $(n_T - m)$, rather than n_T and the total number of items to be taken into consideration is $(n_T + n_D - m - v)$, rather than $(n_T + n_D)$.

Fig. 1 shows the proportions correct for the case where $n_T = n_D = 4$. The curves for other values of n_T and n_D look similar and can easily be drawn by using (7). There are two interesting aspects about the curves. The first is that the expected proportion correct for the case where the participants only managed to track a single target, and know nothing about the rest of the test display ($m = 1$ and $v = 0$), is 0.57. Using (1), provided by Scholl et al. (2001), the computed value would be 0.625. In general, for test displays where the number of targets equals the number of distractors, (1) would yield:

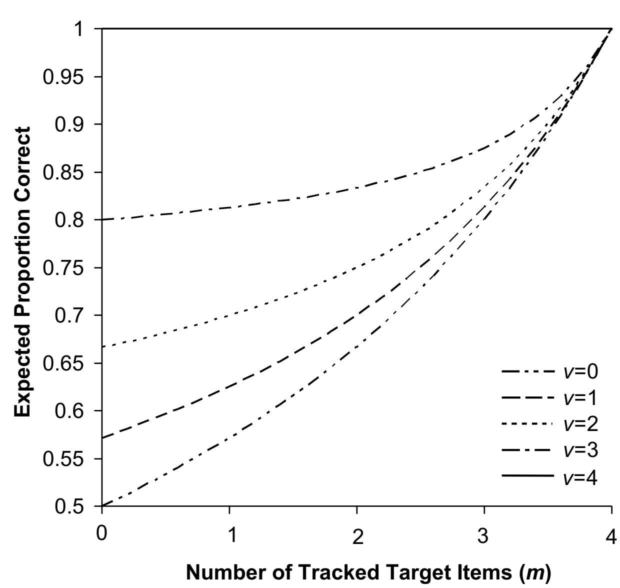

Fig. 1. Expected proportion correct for the *mark all* method, as a function of the number of target items (m) and distractor items (v) tracked. Number of targets (n_T): 4; number of distractors (n_D): 4.

$$p = \frac{n+1}{2n} \tag{1a}$$

whereas (7) would yield:

$$p = \frac{n}{(2n-1)} \tag{7a}$$

whenever $n > 1$, the value from (7a) will be smaller than the value from (1a). This is the difference between sampling with and without replacement. It means that (1a) overestimates the proportion correct where participants have only succeeded in following a single target item. This in turn leads to concluding that participants have been reduced to following only a single item (e.g. van-Marle & Scholl, 2003, who call this chance level), whereas performance actually implies that the participants have been following more than a single item. This would mean the difference between an inability to track multiple items and a diminished ability to track multiple items, with its accompanying theoretical implications. The second interesting characteristic is that a certain proportion correct in the *mark all* method can be derived with a number of different combinations of m and v. For instance, when participants get 0.67 correct in the *mark all* method (for $n_T = n_D = 4$) this could suggest that they know that 2 items in the test display are targets, and are guessing the rest ($m = 2$, $v = 0$). However it could also suggest that they know that 1 item is a target, and know as well that 1.6 of the remaining items certainly are distractors ($m = 1$, $v = 1.6$).

It is important to realize that what is tested in a test display is knowledge about the items it contains, that is, knowledge about which of the items are the targets that were flashed during the definition phase and knowledge about which of the items are the distractors. What is not tested is the way in which the participants managed to hold on to that knowledge over the tracking phase. That is something that will have to be provided by a psychological theory like FINST (Pylyshyn, 1989), or the grouping account proposed by Yantis (1992).

So, even if we were to accept that the m in (7) represents the number of target items successfully tracked, we see that a certain proportion correct is consistent with a continuum of values of m. By assuming that $v = 0$, it becomes possible to compute a unique value of m for all observed proportions correct. However, assuming that $v = 1$ would also yield a unique value of m for most proportions correct (some proportions correct—e.g. 0.5—are impossible when assuming $v = 1$).

In the same way that m could be considered the number of targets tracked, v could be considered the number of distractors tracked. Accepting these interpretations of m and v we are able to compute a minimum and a maximum of the total number of items tracked, by adding m and v. A certain proportion correct is compatible with any number of tracked items lying between these limits.

The upper limit is given by (see 1 in Appendix A)

$$(m+v)_{\max} = 4n_T\left(p - \frac{3}{4}\right) + n_D \tag{8}$$

Because the most efficient way (i.e. the smallest number of items needed to achieve a particular proportion correct) of tracking is to concentrate on the items of the kind of which there are the fewest (see A.1), the lower limit depends on whether there are more target items or more distractor items. In most MOT studies, $n_D \geqslant n_T$ so (using $v = 0$ in (7) and solving for m)

$$(m+v)_{\min} = \frac{n_T(n_T p + n_D p - n_T)}{(n_T p + n_D - n_T)} \tag{9}$$

However, as can be seen from Fig. 2 (for the case where $n_T = n_D = 4$), these limits start to diverge quite quickly, and even for reasonably modest levels of performance the difference between minimum and maximum is more than 1.

The proportions correct in the *probe one* method are underdetermined as well. This is the result of two factors. First, it is the result of the freedom that the participants have when they are guessing. The strategy that will yield the highest proportion correct (best guessing strategy), depending on the number of targets and distractors tracked is

$$p = \frac{1}{b}\frac{m}{n_T} + \left(1 - \frac{1}{b}\right)\frac{v}{n_D} + \max\left(\frac{1}{b}\left(1 - \frac{m}{n_T}\right), \left(1 - \frac{1}{b}\right)\left(1 - \frac{v}{n_D}\right)\right) \tag{10}$$

In (10), $1/b$ is the probability that a target item will be probed (i.e. the probability that 'target' is the correct

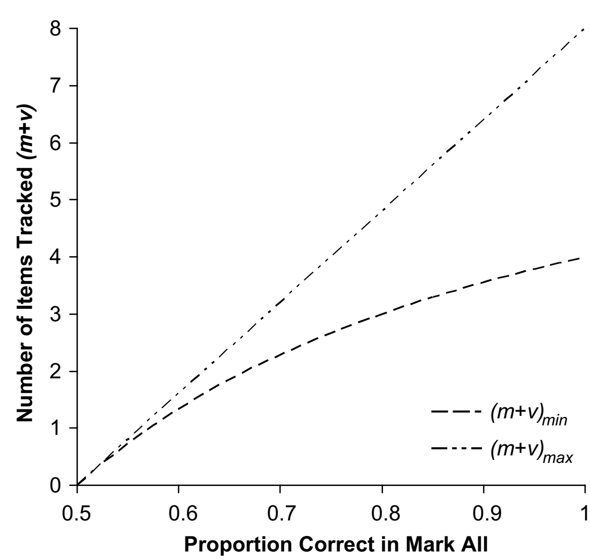

Fig. 2. The minimum and maximum values of $m + v$ as a function of the proportion correct in the *mark all* method. Number of targets: 4; number of distractors: 4.

answer, the per category probability). Together with n_T and n_D, $1/b$ determines the probability that a particular item will be probed.

In the best guessing strategy, the participants will respond 'target' for the m items they know to be targets, and they will respond 'distractor' for the v items they know to be distractors. Whenever an item is probed of which the status is unknown, the participants will base their answer on what they do know. If there are more targets than distractors amongst the known items, then there will be more distractors than targets amongst the unknown items (for $n_T \leqslant n_D$). Participants will therefore guess 'distractor' for the unknown item, because it is more likely to come from that category. Similarly, if there are more distractors amongst the known items, this makes it more likely that the probed item is a target. In this way, the participants maximize their chances of getting it right. Depending on the values of $1/b$, n_T and n_D this can involve some complicated mental arithmetic, because their combination can produce unequal weighting factors, preventing a simple decision based on the difference between m and v. Moreover, the participants will base their decision on the perceived probabilities, rather than on the real probabilities. This provides an extra source of errors, because participants might use the best guessing strategy, but use the incorrect probabilities. For instance, they might assume that the per item probability (the chance that a particular item will be probed) is identical for targets and distractors, whereas in fact the per item probabilities are different for targets and distractors, because it is actually the per category probabilities (the chance that the correct answer is either 'target' or 'distractor') that are identical.

Choosing 0.5 as the value for $1/b$ and giving n_T and n_D equal values would eliminate this source of error. With these values, using the per item probability and using the per category probability to maximize the proportion correct will yield identical results. In this case, (10) reverts to

$$p = \frac{m}{2n} + \frac{v}{2n} + \max\left(\frac{1}{2} - \frac{m}{2n}, \frac{1}{2} - \frac{v}{2n}\right) \qquad (11)$$

When $v = 0$, (11) is the sophisticated guessing strategy discussed in Yantis (1992).

A different strategy—the pure guessing strategy—that participants could employ in the *probe one* method is to answer randomly, whenever they do not know whether the probed item is a target or a distractor. For the case where n_T equals n_D and targets and distractors are probed equally often the expected proportion correct for the pure guessing strategy is given by

$$p = \frac{1}{2}\left(\frac{m}{n} + \frac{1}{2}\frac{(n-m)}{n}\right) + \frac{1}{2}\left(\frac{v}{n} + \frac{1}{2}\frac{(n-v)}{n}\right) \qquad (12)$$

This reverts to

$$p = \frac{1}{2} + \frac{m}{4n} + \frac{v}{4n} \qquad (13)$$

In Fig. 3, the curves for the best guessing (11) and the pure guessing (13) strategies are shown for $n_T = n_D = 4$, and several values of m and v. The curve for the best guessing strategy shows that the value of the expected proportion correct only depends on the maximum of m and v. Because participants who use the best guessing strategy will always guess that the probed item will be of the category most likely to be left amongst the unknown items, the expected proportion correct, when participants have tracked a category to a certain level, will be identical for the cases where they have not tracked

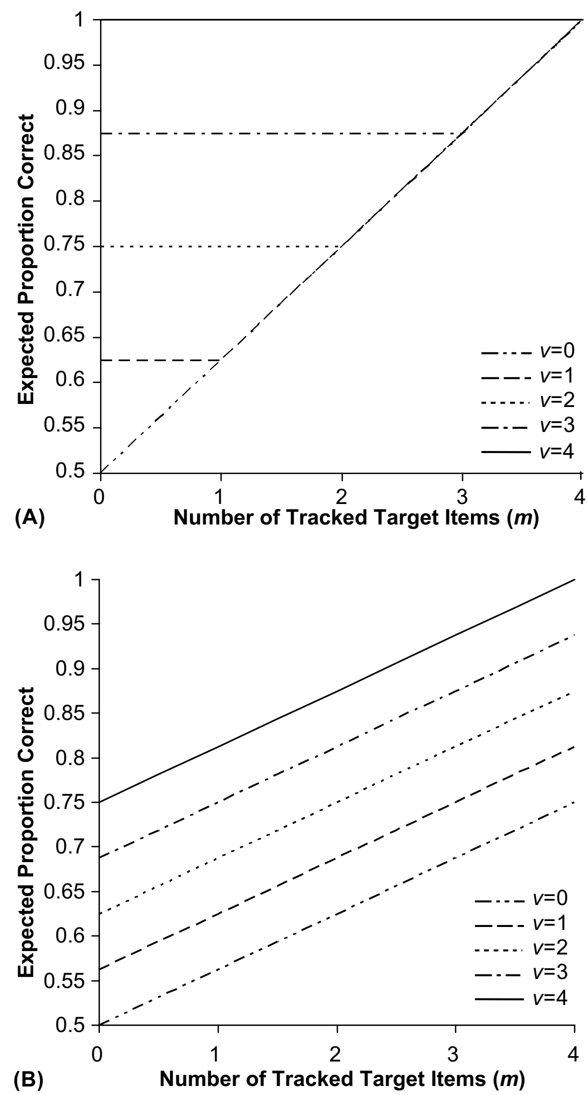

Fig. 3. (A) Expected proportion correct as a function of the number of tracked target (m) and distractor items (v) in the *probe one* method, when the participants are using the best guessing strategy. (B) Expected proportion correct as a function of the number of tracked target (m) and distractor items (v) in the *probe one* method, when the participants are using the pure guessing strategy.

the other category at all, and when they have tracked the other category up to the same level. For instance, tracking 2 target items and 2 distractor items yields the same expected proportion correct as tracking only 2 target items. From an experimental perspective this is a very unwelcome property of the best guessing strategy. As before for the *mark all* method, it is impossible to arrive at a unique combination of m and v from an observed proportion correct. Moreover, the sum of m and v varies for different combinations of m and v that are compatible with an observed proportion correct. When $n_T = n_D$, the minimum value of the sum of m and v for any proportion correct under the best guessing strategy will be the value of m that is compatible with that proportion (14), and the maximum will be twice this value ((15), see A.3):

$$(m + v)_{\min} = n(2p - 1) \quad (14)$$

$$(m + v)_{\max} = 2n(2p - 1) \quad (15)$$

For the higher proportions correct this means a large difference between the minimum and the maximum number of items tracked when participants use the best guessing strategy.

The curves for the pure guessing strategy look rather different. Instead of all terminating at a proportion correct of 1, the pure guessing strategy produces parallel lines. Again, multiple combinations of m and v are compatible with an observed proportion correct. This time however, the sum of m and v is a constant for all those combinations (see A.4):

$$m + v = 2n(2p - 1) \quad (16)$$

This means that for the pure guessing strategy, we would be able to pinpoint the total number of items that were tracked. The particular values of m and v are unknown, but we do know their sum. If the participants were to use the pure guessing strategy, then we would be able to estimate the total number of items tracked, be they targets or distractors. It is important to realize that this property of constant value for m and v only holds when n_T and n_D are equal.

In order to further pin down the way the tracked items are distributed over target and distractor items, the *mark all* method and the *probe one* method could be combined.

To ensure that the behavior of the participants during the definition and tracking phase is identical for the two methods, they should be randomly interspersed. A necessary assumption is that performance of the participants in the test phase is equal for both methods. Because the *probe one* method only requires a single response, and the *mark all* method requires n_T responses, there might be more of a memory burden in the *mark all* task. If this were to be the case, both tasks need to be equated. This probably could be accomplished by slightly increasing the mean duration of the tracking phase for the *probe one* trials. It will also be necessary to introduce some variability in the duration of the tracking phase for both the *mark all* trials and *probe one* trials. This variability would mask the increased duration of the *probe one* trials, which otherwise might be used as a cue by the participants.

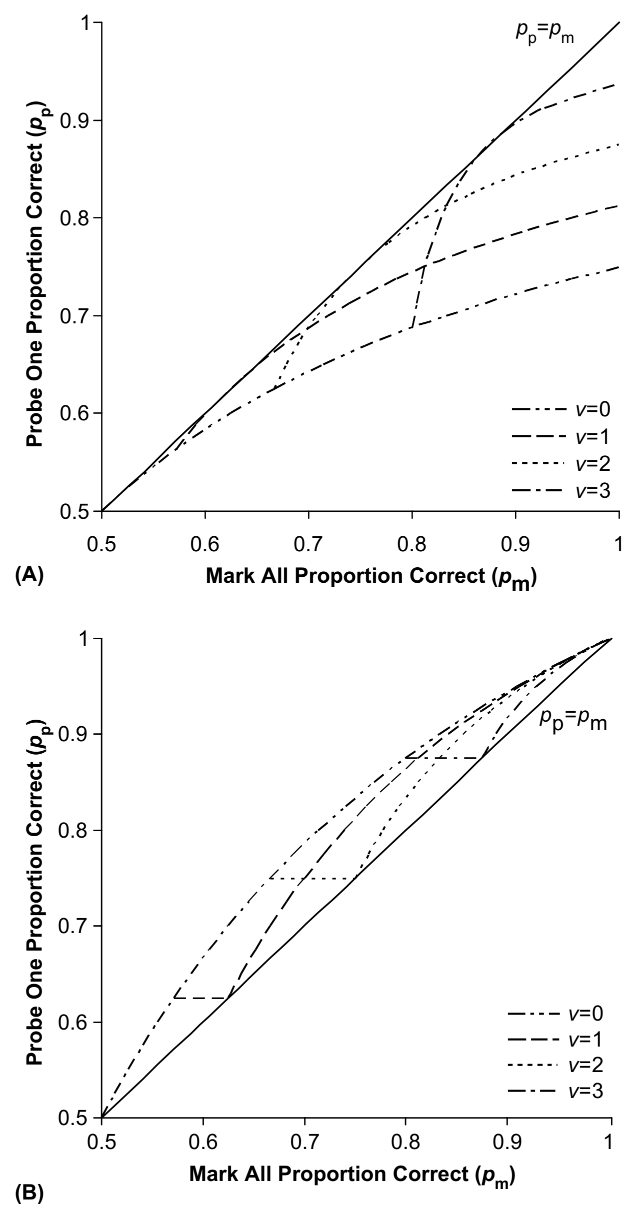

Fig. 4. (A) Curves for fixed values of the number of tracked distractor items (v) plotted for combinations of proportions correct in the *mark all* method (p_m) and in the *probe one* method with pure guessing strategy (p_p), assuming equal performance for both methods. (B) Curves for fixed values of the number of tracked distractor items (v) plotted for combinations of proportions correct in the *mark all* method (p_m) and in the *probe one* method with best guessing strategy (p_p), assuming equal performance for both methods. Note that the assumption that participants are only tracking target items would result in a combination of proportions correct that would fall somewhere on the curve $v = 0$.

Assuming equal performance, Fig. 4A and B show (for $n_T = n_D = 4$) the combinations of proportions correct for the *mark all* method and *probe one* method with pure guessing strategy and best guessing strategy, respectively. If the participants really would concentrate all their resources on tracking the target items as assumed by FINST (Pylyshyn, 1989), the observed proportions correct should fall somewhere along the curves $v = 0$. Whenever the observed proportions correct deviate significantly from this curve we would have to conclude that the participants do not only have knowledge about the target items, but also about the distractor items in the displays. It would be up to theorists to explain how the participants could be able to accumulate this kind of knowledge. If the participants are using the pure guessing strategy, their performance should fall on or below the line where the proportions correct in the *mark all* and the *probe one* methods are identical (see A.5). If the participants are using the best guessing strategy, their performance should fall on or above this line (see A.6).

The combination of the *probe one* method with a pure guessing strategy and the *mark all* method would narrow down the possible combinations of m and v to two, rather than one (see A.7). For $n_T = n_D = n$, these two possible combinations are (with p_p and p_m the proportions correct from the *probe one* and the *mark all* method, respectively):

$$m_1 = n\left(2p_p - 1 + 2\sqrt{(p_m - p_p)(1 - p_p)}\right) \tag{17}$$

$$v_1 = n\left(2p_p - 1 - 2\sqrt{(p_m - p_p)(1 - p_p)}\right) \tag{18}$$

$$m_2 = n\left(2p_p - 1 - 2\sqrt{(p_m - p_p)(1 - p_p)}\right) \tag{19}$$

$$v_2 = n\left(2p_p - 1 + 2\sqrt{(p_m - p_p)(1 - p_p)}\right) \tag{20}$$

Similarly, the combination of the *probe one* method with a best guessing strategy and the *mark all* method would also yield two possible combinations of m and v (see A.8). For $n_T = n_D = n$, these two possible combinations are (with p_p and p_m the proportions correct from the *probe one* and the *mark all* method, respectively):

$$m_1 = n(2p_p - 1) \tag{21}$$

$$v_1 = \frac{n(p_m(2p_p - 3) + 1)}{(2p_p - p_m - 1)} \tag{22}$$

$$m_2 = \frac{n(p_m(2p_p - 3) + 1)}{(2p_p - p_m - 1)} \tag{23}$$

$$v_2 = n(2p_p - 1) \tag{24}$$

The reason for this remaining uncertainty is the interchangeability of tracking target items and distractor items. Decreasing the number of tracked target items and increasing the number of tracked distractor items by the right amounts will yield a combination of proportions correct that is indistinguishable from the original number of tracked targets and distractors. However, combinations of m and v are only interchangeable with a single other combination of m and v. As can been seen from (17) to (20) and from (21) to (24), in the case where the number of targets is identical to the number of distractors, the values of m and v are interchangeable with each other (see also A.9). That is, the value of m in the first combination will be identical to the value of v in the second and vice versa. So, were it to be the case that performance in a certain experiment is compatible with $m = 2$ and $v = 0$, the performance is also compatible with $m = 0$ and $v = 2$. This means that the number of tracked items of a single category is fixed, when the number of targets equals the number of distractors. However, the label of the category (either 'targets' or 'distractors') can still be chosen freely.

Acknowledgment

This work was supported by a grant from the MRC (UK).

Appendix A

A.1. For the mark all method, the maximum value of $m + v$ is given by $4n_T(p - 3/4) + n_D$

From (7) we can find the following expression for m

$$m = \frac{n_T(n_T p + n_D p - vp - n_T)}{(n_T p + n_D - v - n_T)} \tag{A.1}$$

adding v gives an equation for the total number of items tracked ($m + v$) as a function of p and v

$$m + v = \frac{n_T(n_T p + n_D p - vp - n_T)}{(n_T p + n_D - v - n_T)} + v \tag{A.2}$$

By taking the derivative for v and equating to 0 we can find the maximum of this function

$$0 = \frac{-n_T p(n_T p + n_D - v - n_T) + n_T(n_T p + n_D p - vp - n_T)}{(n_T p + n_D - v - n_T)^2} + 1 \tag{A.3}$$

The roots of (A.3) are n_D and $2n_T(p - 1) + n_D$. Substituting these values of v into (A.3) yields:

$$(n_T + n_D) \tag{A.4}$$

$$4n_T(p - 3/4) + n_D \tag{A.5}$$

(A.5) gives the maximum of $m + v$ still compatible with a certain value of p in the *mark all* method.

A.2. The minimum value of $m + v$ in the mark all method will be observed for $v = 0$, when $n_D \geqslant n_T$

Assume two sets of values for m and v lead to the same expected percentage correct such that ($m_1 = m$, $v_1 = 0$) and ($m_2 = m - a$, $v_1 = v$). If (m_1, v_1) is the minimum for $m + v$, then $m_1 + v_1 \leqslant m_2 + v_2$ for all allowable values of a. Consequently, $m \leqslant m - a + v$ and therefore $0 \leqslant v - a$.

First of all, an expression for $v - a$ is needed. To this end, the values for m_1 and v_1 and m_2 and v_2 are substituted into (7), yielding the following equality ($a \geqslant 0$, because otherwise v would become negative, $a \leqslant m$, because otherwise m would become negative):

$$m + \frac{(n_T - m)^2}{(n_T + n_D - m)} = (m - a) + \frac{(n_T - (m - a))^2}{(n_T + n_D - (m - a) - v)} \qquad 0 \geqslant a \geqslant m \tag{A.6}$$

The relation between v and a is therefore given by

$$v = \frac{-an_D^2}{a(m - n_T - n_D) - (n_T - m)^2} \qquad 0 \geqslant a \geqslant m \tag{A.7}$$

Consequently, the function for $v - a$ is

$$v - a = \frac{-an_D^2}{a(m - n_T - n_D) - (n_T - m)^2} - a \qquad 0 \geqslant a \geqslant m \tag{A.8}$$

(A.8) has two zero crossings:

$$a = 0 \text{ and } a = \frac{(n_D^2 - (n_T - m)^2)}{(n_T + n_D - m)} \qquad 0 \geqslant a \geqslant m \tag{A.9}$$

We now need to show that (i) (A.8) is always positive between these two zero crossings and (ii) that the allowable values of a will always be between the two zero crossings. First, we will start with (i).

(A.8) has two extreme values, a local minimum:

$$a = -(n_T - m) \tag{A.10}$$

and a local maximum:

$$a = \frac{(n_T - m - n_D)(n_T - m)}{(m - n_T - n_D)} \tag{A.11}$$

because $0 \geqslant a \geqslant m$, we only have to look at (A.11). We now need to show that (A.11) is always positive. Substituting (A.11) into (A.8) yields:

$$\frac{\dfrac{-(n_T - m - n_D)(n_T - m)}{(m - n_T - n_D)} n_D^2}{(n_T - m - n_D)(n_T - m) - (n_T - m)^2} - \frac{(n_T - m - n_D)(n_T - m)}{(m - n_T - n_D)} \tag{A.12}$$

after some rearranging, (A.12) results in

$$\frac{(n_T - m - n_D)(n_D - n_T + m)}{(m - n_T - n_D)} \tag{A.13}$$

whenever $n_D \geqslant n_T$, (A.13) will be a positive value. Maximum (A.11) is always located between the two zero crossings because

$$a = \frac{n_D(n_T - m) - (n_T - m)^2}{(n_T + n_D - m)} \qquad \text{(A.11, rearranged)}$$

will be larger than or equal to 0 whenever $n_D \geqslant n_T$ and it will be smaller than or equal to

$$a = \frac{(n_D^2 - (n_T - m)^2)}{(n_T + n_D - m)} \qquad \text{(A.9, repeated)}$$

Because (A.8) has a positive maximum between its zero crossings, it is positive between its zero crossings.

We will now show that the value of a in (A.8) will always be between its two zero-crossings. By definition, $0 \geqslant a \geqslant m$, so the only thing that we need to show is that a can not become larger than the second zero crossing:

$$a = \frac{(n_D^2 - (n_T - m)^2)}{(n_T + n_D - m)} \qquad \text{(A.9, repeated)}$$

(A.9) will always be equal to or larger than m, the maximally allowable value of a, when $n_D \geqslant n_T$:

$$\frac{(n_D^2 - (n_T - m)^2)}{(n_T + n_D - m)} \geqslant m \tag{A.14}$$

rearranging yields:

$$\frac{(n_D^2 - n_T^2 + mn_T)}{(n_T + n_D - m)} \geqslant \frac{mn_D}{(n_T + n_D - m)} \tag{A.15}$$

Substituting $n_D = n_T + i$ in (A.15) gives

$$\frac{(2n_T i + i^2 + mn_T)}{(n_T + n_D - m)} \geqslant \frac{mn_T + mi}{(n_T + n_D - m)} \tag{A.16}$$

when $i = 0$ this is an equality, whenever i is larger than 0, the left hand side will be larger than the right hand side.

There is one special case that has to be taken care of: (A.8) is undefined for:

$$a = \frac{(n_T - m)^2}{(m - n_T - n_D)} \tag{A.17}$$

(A.17) will always be smaller than or equal to zero. Because $0 \geqslant a \geqslant m$, only the case where $a = 0$ is important. In (A.17) a will only be 0 when $n_T = m$. If we substitute this value for m into (A.6), the two solutions are

$$a = 0 \quad \text{and} \quad v = n_D \tag{A.18}$$

When $a = 0$, only targets are tracked. So, we only have to look at $v = n_D$. This equality shows that, when $n_T = m$, the only way to achieve the same percentage correct is to track n_D distractors. Because $n_D \geqslant n_T$, $m + v$ will always be equal or larger when n_D distractors are tracked than when only n_T targets are tracked.

A.3. For the best guessing strategy, with a value of 0.5 for 1/b, and $n_T = n_D = n$ the maximum of m + v will be twice the minimal value of m + v compatible with a certain percentage

$$p = \frac{m}{2n} + \frac{v}{2n} + \max\left(\frac{1}{2} - \frac{m}{2n}, \frac{1}{2} - \frac{v}{2n}\right) \qquad (11, \text{repeated})$$

if $m \geqslant v$

$$p = \frac{1}{2}\left(\frac{m}{n} + 1\right) \qquad (A.19)$$

Thus, p does not depend on v, due to the application of the max-rule. The value of m, compatible with a value p will be

$$m = n(2p - 1) \qquad (A.20)$$

Because the minimum of $m + v$ will occur for $v = 0$, (A.20) also gives the minimum for $m + v$:

$$(m + v)_{\min} = n(2p - 1) \qquad (A.21)$$

The maximum of the sum of $m + v$ still consistent with a certain value of p will be at the point where $m = m$ and $v = m$. Applying the max-rule in this case will still yield (A.20). Because (A.20) gives the value of m, and because $v = m$, the expression for $(m + v)_{\max}$ becomes:

$$(m + v)_{\max} = n(2p - 1) + n(2p - 1) = 2n(2p - 1) \qquad (A.22)$$

if $m \leqslant v$

$$p = \frac{1}{2}\left(\frac{v}{n} + 1\right) \qquad (A.23)$$

Here, p does not depend on m. Following the same logic as above, the expression for $(m + v)_{\min}$ becomes:

$$(m + v)_{\min} = n(2p - 1) \qquad (A.24)$$

The maximum of the sum of $m + v$ still consistent with a certain value of p will be at the point where $m = v$ and $v = v$. Applying the max-rule in this case will still yield (A.23). Because (A.23) gives the value of v, and because $m = v$, the expression for $(m + v)_{\max}$ becomes:

$$(m + v)_{\max} = n(2p - 1) + n(2p - 1) = 2n(2p - 1) \qquad (A.25)$$

So, the maximum of the sum of $m + v$ is two times the minimum in the best guessing strategy, when $n_T = n_D = n$.

A.4. For the pure guessing strategy, with $n_T = n_D = n$ and 1/b = 0.5 m + v will be constant for an observed proportion correct

The observed proportion correct:

$$p = \frac{1}{2} + \frac{m}{4n} + \frac{v}{4n} \qquad (13, \text{repeated})$$

rearranging yields:

$$m = 4n\left(p - \frac{1}{2} - \frac{v}{4n}\right) \qquad (A.26)$$

the expression for $m + v$ will therefore be

$$m + v = 4n\left(p - \frac{1}{2} - \frac{v}{4n}\right) + v \qquad (A.27)$$

this simplifies to

$$m + v = 2n(2p - 1) \qquad (16, \text{repeated})$$

Clearly, $m + v$ only depends on the value of p, making the sum of m and v independent of the values of m and v when the participants are applying the pure guessing strategy.

This property does not hold for the case where n_T and n_D are unequal. Here (16) becomes:

$$m + v = 2n_T(2p - 1) - v\left(\frac{n_T}{n_D} - 1\right). \qquad (A.28)$$

A.5. The proportion correct of mark all (p_m) will always be larger or equal to the proportion correct of probe one (p_p), when the participants use the pure guessing strategy during probe trials and $n_T = n_D = n$

Assuming that m and v are constant across mark all and probe one trials, from (7) and (13), the difference between p_p and p_m will be

$$p_p - p_m = \frac{1}{2} + \frac{m}{4n} + \frac{v}{4n} - \left(m + \frac{(n - m)^2}{(2n - m - v)}\right) \Big/ n \qquad (A.29)$$

rearranging yields:

$$p_p - p_m = \frac{-(m - v)^2}{4n(2n - m - v)} \qquad (A.30)$$

The expression in (A.30) will always be negative, except when m equals v and $p_p - p_m$ will be 0. (A.30) is not defined for $m = v = n$. However this is the case where the expected proportions correct are 1 for both the *mark all* trials and the *probe one* trials. Hence the difference will be 0 here as well.

A.6. The proportion correct of mark all (p_m) will always be smaller than or equal to the proportion correct of probe one (p_p), when the participants use the best guessing strategy during probe trials and 1/b = 0.5 and $n_T = n_D = n$

There are two cases to consider: $v \geqslant m$ and $v < m$.
For $v \geqslant m$:

Assuming that m and v are constant across mark all and probe one trials, from (7) and (A.19), the difference between p_p and p_m will be

$$p_p - p_m = \frac{v}{2n} + \frac{1}{2} - \left(m + \frac{(n-m)^2}{(2n-m-v)}\right) \Big/ n \qquad (A.31)$$

Rearranging yields:

$$p_p - p_m = \frac{(n-v)(v-m)}{2n(2n-m-v)} \qquad (A.32)$$

This result will always be positive or zero, because n will always be larger than or equal to v, v is larger than or equal to m, and $2n$ will always be larger than or equal to $(m + v)$. For the special case where $2n$ equals $(m + v)$, (A.32) is undefined. However, this is the case where the expected proportion for both p_p and p_m is 1.

For $v < m$, from (7) and (A.23):

$$p_p - p_m = \frac{m}{2n} + \frac{1}{2} - \left(m + \frac{(n-m)^2}{(2n-m-v)}\right) \Big/ n \qquad (A.33)$$

Rearranging yields:

$$p_p - p_m = \frac{(n-m)}{n}\left(\frac{1}{2} - \frac{1}{1 + \frac{(n-v)}{(n-m)}}\right) \qquad (A.34)$$

Because $v < m$ and $n \geqslant m$, this result will always be positive. For the case where $m = n$, (A.34) is not defined. However, this is the case where the expected proportion correct is 1 for both p_p and p_m.

A.7. The combination of the percentage correct from a mark all experiment (p_m) and a probe one experiment (p_p) with pure guessing yields two estimates for m and v, when $n_D = n_T = n$, under the assumption that m and v are constant across the two types of experiment

The expected proportion correct in a *probe one* experiment is given by

$$p_p = \frac{1}{2} + \frac{m}{4n} + \frac{v}{4n} \qquad (13, \text{repeated})$$

The expected proportion correct in a *mark all* experiment is given by

$$p_m = \left(m + \frac{(n-m)^2}{(2n-m-v)}\right) \Big/ n \qquad (A.35)$$

((A.35) is (7) repeated, with $n_D = n_T = n$).

Substitution for v from (13) in (A.32) yields:

$$p_m = \left(m + \frac{(n-m)^2}{(2n-m-(4np_p-2n-m))}\right) \Big/ n \qquad (A.36)$$

Solving (A.36) for m and substituting this value in (13) results in the following two combinations of m and v:

$$m_1 = n\left(2p_p - 1 + 2\sqrt{(p_m - p_p)(1-p_p)}\right) \qquad (17, \text{repeated})$$

$$v_1 = n\left(2p_p - 1 - 2\sqrt{(p_m - p_p)(1-p_p)}\right) \qquad (18, \text{repeated})$$

$$m_2 = n\left(2p_p - 1 - 2\sqrt{(p_m - p_p)(1-p_p)}\right) \qquad (19, \text{repeated})$$

$$v_2 = n\left(2p_p - 1 + 2\sqrt{(p_m - p_p)(1-p_p)}\right) \qquad (20, \text{repeated})$$

As expected, the value of m computed with (17) is identical to the value of v computed with (20), as are the values of v and m computed with (18) and (19), respectively.

There is only a single solution when p_m and p_p are identical, and there are no solutions when p_p is larger than p_m. This is consistent with the previously observed property that the value of p_m is always equal to or larger than p_p, when the participants are using a pure guessing strategy.

It is also possible to give more general versions of (17) and (19) because it is not necessary to use equal values for n_T and n_D when the guesses are not made by the participants, because the possibility of a misunderstanding of the per item probabilities are irrelevant. They are however, quite unwieldy:

$$m_1 = \frac{-(n_D(4p_p - 3 + p_m) + (n_T(1-p_m))) - \sqrt{(n_D(4p_p - 3 + p_m) + n_T(1-p_m))^2 + 4n_D(n_Tp_m + 3n_Dp_m - 4n_Dp_mp_p - n_T)}}{\frac{-2n_D}{n_T}}$$

$$m_2 = \frac{-(n_D(4p_p - 3 + p_m) + (n_T(1-p_m))) + \sqrt{(n_D(4p_p - 3 + p_m) + n_T(1-p_m))^2 + 4n_D(n_Tp_m + 3n_Dp_m - 4n_Dp_mp_p - n_T)}}{\frac{-2n_D}{n_T}}$$

The corresponding values v_1 and v_2 can be found by entering the values of m_1 and m_2 into:

$$v = 4n_D\left(p_p - \frac{1}{2} - \frac{m}{4n_T}\right) \qquad (A.37)$$

A.8. The combination of the percentage correct from a mark all experiment (p_m) and a probe one experiment (p_p) with best guessing yields two estimates for m and v, when 1/b = 0.5 and $n_D = n_T = n$, under the assumption that m and v are constant across the two types of experiment

The expected proportion correct in a *mark all* experiment is given by

$$p_{\mathrm{m}} = \left(m + \frac{(n-m)^2}{(2n-m-v)}\right) \Big/ n \qquad \text{(A.35, repeated)}$$

For the expected proportion correct in the *probe one* experiment with best guessing strategy, there are two cases that need to be considered: $m \geqslant v$ and $m < v$.

For $m \geqslant v$ the expected proportion correct is

$$p_{\mathrm{p}} = \frac{1}{2}\left(\frac{m}{n} + 1\right) \qquad \text{(A.19, repeated)}$$

(A.19) yields the following expression for m:

$$m = n(2p_{\mathrm{p}} - 1) \qquad \text{(21, repeated)}$$

Substituting (21) into (A.35) and solving for v yields:

$$v = \frac{n(p_{\mathrm{m}}(2p_{\mathrm{p}} - 3) + 1)}{(2p_{\mathrm{p}} - p_{\mathrm{m}} - 1)} \qquad \text{(22, repeated)}$$

(22) is not defined for $p_{\mathrm{m}} = 2p_{\mathrm{p}} - 1$. This is the case where $m = n$:

Substitution of (A.35) and (A.19) into $p_{\mathrm{m}} = 2p_{\mathrm{p}} - 1$ yields

$$\left(m + \frac{(n-m)^2}{(2n-m-v)}\right) \Big/ n = \frac{m}{n} \qquad \text{(A.38)}$$

rearranging terms we get

$$(n-m)^2 = 0 \qquad \text{(A.39)}$$

The only solution for (A.39) is $m = n$. This is the case where we would expect $p_{\mathrm{m}} = p_{\mathrm{p}} = 1$. When $m = n$, v can adopt any value between 0 and n.

When $m < v$:

$$p_{\mathrm{p}} = \frac{1}{2}\left(\frac{v}{n} + 1\right) \qquad \text{(A.23, repeated)}$$

(A.23) yields the following expression for v:

$$v = n(2p_{\mathrm{p}} - 1) \qquad \text{(24, repeated)}$$

Substituting (24) into (A.35) and solving for m yields:

$$m = \frac{n(p_{\mathrm{m}}(2p_{\mathrm{p}} - 3) + 1)}{(2p_{\mathrm{p}} - p_{\mathrm{m}} - 1)} \qquad \text{(23, repeated)}$$

(23) is not defined for $p_{\mathrm{m}} = 2p_{\mathrm{p}} - 1$. This is the case where $v = n$:

Substitution of (A.35) and (A.23) into $p_{\mathrm{m}} = 2p_{\mathrm{p}} - 1$

$$\left(m + \frac{(n-m)^2}{(2n-m-v)}\right) \Big/ n = \frac{v}{n} \qquad \text{(A.40)}$$

rearranging terms we get

$$(v-n)^2 = 0 \qquad \text{(A.41)}$$

The only solution for (A.41) is $v = n$. This is the case where we would expect $p_{\mathrm{m}} = p_{\mathrm{p}} = 1$. When $v = n$, m can adopt any value between 0 and n.

Because we do not know whether m is larger or smaller than v for a given participant, each combination of p_{p} and p_{m} has two solutions, given by (21) and (22) and by (23) and (24).

If we allow different values for n_{T} and n_{D} (but still $1/b = 0$), the formulas become more unwieldy again:

$m \geqslant v$:

$$m_1 = \left(p_{\mathrm{p}} - \frac{1}{2}\right)(n_{\mathrm{D}} + n_{\mathrm{T}})$$

$$v_1 = \frac{n_{\mathrm{D}}^2\left(p_{\mathrm{p}} - \frac{1}{2}\right) + n_{\mathrm{T}}^2\left(\frac{3}{2} - \frac{3}{2}p_{\mathrm{m}} - p_{\mathrm{p}} + p_{\mathrm{p}}p_{\mathrm{m}}\right) + n_{\mathrm{T}}n_{\mathrm{D}}\left(p_{\mathrm{p}}p_{\mathrm{m}} - \frac{3}{2}p_{\mathrm{m}}\right)}{n_{\mathrm{D}}\left(p_{\mathrm{p}} - \frac{1}{2}\right) + n_{\mathrm{T}}\left(p_{\mathrm{p}} - p_{\mathrm{m}} - \frac{1}{2}\right)}$$

$m \leqslant v$:

$$m_2 = \frac{n_{\mathrm{T}}^2\left(p_{\mathrm{p}}p_{\mathrm{m}} - \frac{3}{2}p_{\mathrm{m}} + 1\right) + n_{\mathrm{T}}n_{\mathrm{D}}\left(p_{\mathrm{p}}p_{\mathrm{m}} - \frac{3}{2}p_{\mathrm{m}}\right)}{n_{\mathrm{D}}\left(p_{\mathrm{p}} - \frac{3}{2}\right) + n_{\mathrm{T}}\left(p_{\mathrm{p}} - p_{\mathrm{m}} + \frac{1}{2}\right)}$$

$$v_2 = \left(p_{\mathrm{p}} - \frac{1}{2}\right)(n_{\mathrm{D}} + n_{\mathrm{T}})$$

A.9. In mark all trials, with $n_T = n_D = n$, the number of target items tracked m and the number of distractor items tracked v are interchangeable

The expected proportion correct as a function of m and v:

$$p = \left(m + \frac{(n_{\mathrm{T}} - m)^2}{(n_{\mathrm{T}} - m + n_{\mathrm{D}} - v)}\right) \Big/ n_{\mathrm{T}} \qquad \text{(7, repeated)}$$

Rearranging (7) yields m as a function of p and v:

$$m = \frac{n_{\mathrm{T}}(n_{\mathrm{T}}p + n_{\mathrm{D}}p - vp - n_{\mathrm{T}})}{(n_{\mathrm{T}}p + n_{\mathrm{D}} - v - n_{\mathrm{T}})} \qquad \text{(A.42)}$$

Another rearrangement of (7) gives v as a function of p and m:

$$v = (n_{\mathrm{T}} - m + n_{\mathrm{D}}) - \frac{(n_{\mathrm{T}} - m)^2}{(n_{\mathrm{T}}p - m)} \qquad \text{(A.43)}$$

If we assume that v takes the arbitrary value a and substitute this in (A.42):

$$m = \frac{n_{\mathrm{T}}(n_{\mathrm{T}}p + n_{\mathrm{D}}p - ap - n_{\mathrm{T}})}{(n_{\mathrm{T}}p + n_{\mathrm{D}} - a - n_{\mathrm{T}})} \qquad \text{(A.44)}$$

in the special case where $n_{\mathrm{T}} = n_{\mathrm{D}}$ we get:

$$m = \frac{n(2np - ap - n)}{(np - a)} \qquad \text{(A.45)}$$

If we assume that m takes the same arbitrary value a and substitute this in (A.43):

$$v = (n_{\mathrm{T}} - a + n_{\mathrm{D}}) - \frac{(n_{\mathrm{T}} - a)^2}{(n_{\mathrm{T}}p - a)} \quad \text{(A.46)}$$

after some rearranging we get

$$v = \frac{n_{\mathrm{T}}(n_{\mathrm{T}}p - ap + n_{\mathrm{D}}p - n_{\mathrm{T}}) + a(n_{\mathrm{T}} - n_{\mathrm{D}})}{(n_{\mathrm{T}}p - a)} \quad \text{(A.47)}$$

in the special case where $n_{\mathrm{T}} = n_{\mathrm{D}}$ this yields:

$$v = \frac{n(2np - ap - n)}{(np - a)} \quad \text{(A.48)}$$

So, interchanging the values of m and v will result in exactly the same expected proportions correct, if, and only if, $n_{\mathrm{T}} = n_{\mathrm{D}}$.

References

Pylyshyn, Z. W. (1989). The role of location indexes in spatial perception: A sketch of the FINST spatial index model. *Cognition, 32*, 65–97.

Pylyshyn, Z. W., & Storm, R. W. (1988). Tracking multiple independent targets: Evidence for a parallel tracking mechanism. *Spatial Vision, 3*, 179–197.

Scholl, B. J., Pylyshyn, Z. W., & Feldman, J. (2001). What is a visual object? Evidence from target merging in multiple object tracking. *Cognition, 80*, 159–177.

Tripathy, S. P., & Barrett, B. T. (2004). Severe loss of positional information when detecting deviations in multiple trajectories. *Journal of Vision, 4*, 1020–1043http://journalofvision.org/4/12/4/.

vanMarle, K., & Scholl, B. J. (2003). Attentive tracking of objects versus substances. *Psychological Science, 14*, 498–504.

Viswanathan, L., & Mingolla, E. (2002). Dynamics of attention in depth: Evidence from multi-element tracking. *Perception, 31*, 1415–1437.

Yantis, S. (1992). Multielement Visual Tracking: Attention and perceptual organization. *Cognitive Psychology, 24*, 295–340.