

Rejoinder: **Missing the Target: A Reply to Koehler & Macchi (2009)**

BEN R. NEWELL*, CHRIS J. MITCHELL and BRETT K. HAYES
School of Psychology, University of New South Wales, Sydney, Australia

ABSTRACT

Koehler and Macchi (2009) criticize the experiments presented in Newell, Mitchell, and Hayes (2008) as being “virtually irrelevant” to exemplar cuing theory. This reply addresses that interpretation and argues that the experiments dealt with issues at the heart of the theory and provided evidence highly relevant to understand how people think about low-probability events. The role of the ‘target’ in probabilistic statements is examined, highlighting the need for further theoretical and empirical clarification of the concept. The remaining specific criticisms raised in the commentary are discussed as well. Copyright © 2009 John Wiley & Sons, Ltd.

KEY WORDS exemplar cuing; probability; frequency format; imaginability

Newell, Mitchell, and Hayes (2008) (NMH) presented three experiments aimed at comparing two theories of the way people think about low-probability events: the statistical format account and exemplar cuing theory (EC). The statistical format account predicts that information presented as frequencies (e.g., 50 out of 5000) will be easier to imagine than equivalent information presented as probabilities (e.g., 1%) and this will have a subsequent effect on judgments (i.e., people will be more willing to participate in potentially positive events—(e.g., lotteries)—and less willing to participate in potentially negative events—(e.g., taking vaccinations). EC theory predicts that format per se will not impact judgments but that certain combinations of factors lead to exemplar generation and that these exemplars affect the weight people attach to the possibility that a low-probability event will occur. NMH’s results supported the general notion that imaginability plays a role in thinking about low-probability events, but questioned one of the mechanisms of EC theory.

NMH’S EXPERIMENTS

Koehler and Macchi (2009) (KM) present a summary of NMH’s experiments which fails to acknowledge that the overall conclusion is consistent with the main ideas behind exemplar cuing: namely that the imaginability of outcomes plays a key role in thinking about low-probability events. They also did not acknowledge the observed support for one mechanism¹ of EC theory: that integer numerators (e.g., 1 out of 1000) are more imaginable than fractional numerators

* Correspondence to: Ben R. Newell, School of Psychology, University of New South Wales, Sydney 2052, Australia.
E-mail: ben.newell@unsw.edu.au

¹KM question the use of the words “mechanism” and “multiplicative” and the use of quotation marks. The quotation marks were not used to indicate direct quotes, merely to suggest a degree of uncertainty about the appropriateness of the term multiplicative mechanism. The term provided a shorthand way of describing the claim that reference class size and incidence rates are multiplied. KM use the word “mechanism” in their own conclusion so appear to be comfortable with the term even if it did not appear in their original article.

(e.g., 0.1 out of 100) (although this finding is also consistent with the statistical format account). In addition NMH reported evidence pertaining directly to the imaginability of outcomes (imaginability and affect ratings) which showed some direct support for the notion that imaginability and willingness to participate are more strongly and consistently related in frequency than probability formats. Rather than being virtually irrelevant, the experiments dealt directly with the key issues and mechanisms of concern to EC theory.

MISSING THE TARGET

The substantive point of the critique is that NMH misunderstood an important qualifier to the claim that exemplars are cued when the product of the reference class size and the incidence rate exceeds one. The qualifier is that such multiplication *only* occurs in a multiple target situation. We argue below that missing this qualification does not render the results irrelevant, and we welcome this opportunity to gain some clarification on the nature of the target in probabilistic statements. Koehler and Macchi (2004, p. 541) described the difference between multiple and single targets thus:

The target of a probabilistic statement identifies a problem-relevant sample space or reference class. Single targets (e.g., “Phoebe’s lottery ticket”) offer the smallest reference class ($n = 1$) by focusing attention narrowly on one event. This focus discourages exemplar production because there are no other events to consider. Consequently in the lottery example, people who receive single targets are unlikely to think about other lottery tickets that could win. In contrast, multiple targets (e.g., “tickets sold in the lottery”) promote exemplar production because they offer a large reference class ($n > 1$) within which to consider other winning tickets.

They provided the following two statements to illustrate the difference between (a) single and (b) multiple targets (p. 540):

- a) There is a 0.1% chance that a given ticket will win.
- b) One in every 1000 tickets sold out of the 500,000 tickets that are sold each day will win.

Although Koehler and Macchi (2004) used the example of a lottery to generate predictions about EC, the multiplicative mechanism was only *tested* in a single experiment using arguably less transparent and more complex statistics pertaining to DNA evidence. These statistics were embedded in a two-page summary of a court case. NMH sought to expand the tests of the theory to other low-probability situations (including lotteries) and make the test more transparent by presenting the statistical information succinctly in short vignettes. The lottery statement in the probability version of NMH’s exemplars cued condition reads as follows:

“You are offered the opportunity, today, to buy a ticket in the new Sydney lottery. Five thousand tickets are sold each day and it is estimated that you have a 1% chance of buying a winning ticket.”

NMH argued that because the product of the reference class size and the incidence rate ($5000 \times 1\%$) exceeded one (50), such a statement would, according to EC theory, cue exemplars. KM argue that such a statement does not cue exemplars because attention is drawn to the single ticket that one is about to purchase. They claim this is the case despite the fact that the statement includes information about the 5000 tickets sold each day (unlike Koehler and Macchi’s statement (a) above). The key difference appears to be that the other tickets mentioned in our vignette are not specifically identified as possible winning tickets (compare to (b) above). However, how can KM be confident that attention is only drawn to the single ticket in the statements NMH used? Why is attention not drawn to the reference class size (5000) or the product of incidence rate and reference class size (50)? Moreover, why do people not focus on the ‘one ticket’ when given a statement like (b) above?

The onus is on KM to explain how a participant identifies the target of a given statement, and how one can be confident that the target is as it is intended by the experimenter. This is both a conceptual and empirical question that remains to be resolved. Koehler and Macchi noted in their original paper that discussion of the role of ‘target’ is almost completely absent from the literature (“near invisibility,” Koehler & Macchi, 2004, p. 541); the discussion in their original paper and in their commentary on NMH’s paper does little to improve this situation. Clarification of how one identifies the target of a

statement is paramount for any progress to be made in understanding the applicability and generalizability of EC theory. Moreover, it is not clear why the target is the crucial factor upon which the whole theory appears to rest, and not the format or the reference class size, or the product of the reference class size and the incidence rate—the manipulation of which were uncontroversial in NMH’s experiments. Indeed, the first line of the above quoted passage from Koehler and Macchi (2004) suggests that it is difficult to disentangle the target from the reference class.

The substantive point of KM’s critique is based on speculation about the way in which attention is directed to aspects of the statements NMH used. While this speculation may be justified (indeed NMH acknowledged the potential ambiguity of the statements in the original paper), it does not render the experiments “virtually irrelevant” to EC theory. Rather they contribute to refining, empirically, how participants interpret statistical information about low-probability events. Future experimentation is clearly required to further our understanding of the specific role played by the target of a statement.

OTHER CRITICISMS

The Problem of probability level

KM suggest that the events used in NMH’s scenarios were not of a sufficiently low probability and therefore not appropriate for testing EC theory. KM admit that there is no objective cut-off for what is presumed to be low or very low and so this criticism is also speculative. Without a cut-off or sufficiently well-supported theoretical reason for implementing one, this criticism cannot be evaluated satisfactorily.

Misstatements: fractional vs. integer numerators

NMH stated “jurors were relatively less impressed by the evidence when a fractional numerator was used” (NMH, 2008, p. 319); we acknowledge that this statement was poorly worded. The intending meaning was that jurors provided with fractional numerators were relatively less likely to interpret the statistical evidence as suggesting other possible matches, and were therefore more likely to convict. But we agree that this meaning is not conveyed clearly. However, on page 332 NMH state: “This numerator effect is consistent with Koehler and Macchi’s (2004, Experiment 2) demonstration that DNA match statistics with integer numerators were less convincing of guilt than those with fractional numerators.” This statement clearly communicates the finding, and moreover one of the general messages of NMH was support for the numerator mechanism, making it unlikely that a reader would come away with an incorrect understanding of Koehler and Macchi’s (2004) original findings.

Anchoring on the size of the town

NMH argued that the key result in Koehler and Macchi (2004) that supported EC theory was a finding that the statistical DNA evidence was less convincing when police believed that the murderer was thought to reside in a large town than in a small one. NMH reasoned that because incidence rates were kept constant, participants could have anchored on the size of the town rather than multiplying the reference class size and incidence rate. Thus they might have reasoned that if there were only 500 people (rather than 5,000,000) in the town there is a lower chance that police have “got the wrong man.” KM argue that this is not an accurate representation of the predictions of EC theory nor of their data; below we present some evidence that supports NMH’s re-interpretation.

KM argue that they obtained data consistent with the more complex predictions of EC theory—in other words, interactions between size of the town and the target. Specifically, Koehler and Macchi predicted lower values on their dependent measures only for conditions with multiple targets and large reference class sizes. While this pattern was found, it is instructive to look also at the main effect of the size of the town (the reference class size) on Koehler and Macchi’s (2004) dependent measures. This analysis was not reported in the original article, but if one looks at each dependent measure there appears to be evidence for an independent contribution of the reference class size.

The third and sixth rows of Table 1 (highlighted in bold) show that, on average, for each dependent measure in turn, participants were more impressed (evidence strength), more likely to say that the defendant was the source of the DNA evidence (P(source)), more likely to say the defendant committed the crime (P(guilt)), and, crucially, more likely to reach a verdict of guilty for cases in which police believed the murder resided in a small town (small reference class size) than for those in which the murderer was thought to live in a large city (large reference class size).

Table 1. Mean judgments as a function of reference class size and target from Koehler and Macchi (2004, Experiment 1) collapsed across form (probability or frequency format)

		Evidence strength 1-7	P (source)	P (guilt)	Verdict of guilty (%)
Large reference class size	Single target	4.2	.59	.58	20.0
	Multiple target	3.6	.34	.35	9.0
	Average	3.9	.46	.46	14.5
Small reference class size	Single target	4.0	.53	.51	26.0
	Multiple target	3.9	.52	.50	22.0
	Average	3.95	.525	.505	24.0

Note: reference class size refers to the size of the town or city where police believed the murderer resided (Large = 5,000,000 inhabitants; Small = 500 inhabitants). Single/Multiple target refers to whether the evidence statement included information about an individual suspect (single) or “people in the city” (multiple). P (source) = probability that the defendant was the source of the DNA evidence; P (guilt) = probability that the defendant committed the crime. Evidence strength judgments were made on a Likert-type scale (1 = not at all strong, 7 = extremely strong).

For the first three dependent measures shown in Table 1 these differences are qualified by crossover interactions with target. For single targets the values for the large reference class size are greater than those for the small class size (compare the numbers in row 1 with those in row 4); while for multiple targets the reverse is true (compare rows 2 and 5). However, for the verdict of guilty, which is arguably the most important as it pertains to the participant’s final decision, the interaction is *not* due to a crossover. The direction of the target difference is the *same* for both sizes of reference class but bigger for multiple targets (22 vs. 9% for small and large reference class sizes respectively) than for single targets (26 and 20%). The substantive point is that regardless of target type a higher percentage of respondents gave a guilty verdict when the murderer was believed to be from a small town than from a large city. The average difference of almost 10% in guilty verdicts for small and large towns appears to provide evidence for an independent contribution of the reference class size².

Misstatements: probability vs. frequency

KM point out that NMH provided a technically correct account of Koehler and Macchi’s (2004) results in stating that Koehler and Macchi did not find a significant effect of form (frequency or probability) on judgment, but that NMH should have said that the effect was marginally significant. Koehler and Macchi (2004) stated “we predict that variations in form will not influence decision makers’ probability judgment” (p. 541). To the extent that a null effect can support a theory, Koehler and Macchi’s data did support EC theory and thus NMH described the lack of effect in a way that was consistent with Koehler and Macchi’s interpretation. The importance of form is in need of further empirical clarification given that KM go on to suggest that form *may* influence exemplar production (contrary to the thrust of their original article); doing so appears to lead to further conflation between the predictions of EC theory and the statistical format account (e.g., Slovic, Monahan, & MacGregor, 2000) and narrows the domains in which EC theory makes unique and testable predictions³.

CONCLUSION

KM state that NMH’s findings are “virtually irrelevant to exemplar cuing theory because the theory is misstated and the experiments provided. . .mistake non-exemplar conditions for exemplar conditions”. NMH’s experiments provided clear tests of EC theory as it pertains to predictions about statistical format, the numerator mechanism and, more general claims

²We contacted Jay Koehler to enquire about the significance of the main effect of reference class size on verdicts of guilty. He responded that the simple effects analysis indicated that reference class size was significant for multiple targets, $\chi^2(1, N = 218) = 6.47, p = .01$ (22% vs. 9%) but not for single targets, $\chi^2(1, N = 210) = 0.98, p = .32$ (26% vs. 20%). Thus the 24% (small town) vs. 14.5% (large town) difference collapsed across target type was irrelevant. He did not indicate whether the main effect was significant.

³In an earlier paper Koehler (2001) reported data showing that DNA statistics presented in probability formats were more persuasive of guilt than those presented in frequency formats and took the result as evidence for EC theory. Thus there is some confusion over whether/when EC theory predicts format effects and how it explains such effects.

about the imaginability of low-probability outcomes. Whether the exemplar-cuing scenarios really failed to cue exemplars remains open to debate and further empirical investigation.

REFERENCES

- Koehler, J. J. (2001). When are people persuaded by DNA match statistics? *Law and Human Behavior, 25*, 493–513.
- Koehler, J. J., & Macchi, L. (2004). Thinking about low-probability events: An exemplar cuing theory. *Psychological Science, 15*, 540–546.
- Koehler, J. J., & Macchi, L. (2009). Comments on getting scarred and winning lotteries: Effects of exemplar cuing and statistical format on imagining low-probability events by Newell, Mitchell, and Hayes (2008). *Journal of Behavioral Decision Making, 22*, 523–527.
- Newell, B. R., Mitchell, C. J., & Hayes, B. K. (2008). Getting scarred and winning lotteries: Effects of exemplar cuing and statistical format on imagining low-probability events. *Journal of Behavioral Decision Making, 21*, 317–335.
- Slovic, P., Monahan, J., & MacGregor, D. G. (2000). Violence risk assessment and communication: The effects of using actual cases, providing instruction and employing probability versus frequency formats. *Law and Human Behavior, 24*, 271–296.