

# A rose by any other name: A social-cognitive perspective on poets and poetry

Maya Bar-Hillel\*    Alon Maharshak\*    Avital Moshinsky†    Ruth Nofech‡

## Abstract

Evidence, anecdotal and scientific, suggests that people treat (or are affected by) products of prestigious sources differently than those of less prestigious, or of anonymous, sources. The “products” which are the focus of the present study are poems, and the “sources” are the poets. We explore the manner in which the poet’s name affects the experience of reading a poem. Study 1 establishes the effect we wish to address: a poet’s *reputation* enhances the evaluation of a poem. Study 2 asks whether it is only the reported evaluation of the poem that is enhanced by the poet’s name (as was the case for The Emperor’s New Clothes) or the enhancement is genuine and unaware. Finding for the latter, Study 3 explores whether the poet’s name changes the reader’s experience of it, so that in a sense one is reading a “different” poem. We conclude that it is not so much that the attributed poem really differs from the unattributed poem, as that it is just ineffably better. The name of a highly regarded poet seems to prime quality, and the poem becomes somehow better. This is a more subtle bias than the deliberate one rejected in Study 2, but it is a bias nonetheless. Ethical implications of this kind of effect are discussed.

Keywords: categorization, expectations, experience, focusing illusion, label effects, priming, poetry, reputation bias.

## 1 Introduction

Hearing a star soprano, or attending an exhibition by a famous painter, are expected to be exceptional experiences. And so they should be—their reputation was acquired precisely by their ability to provide such exceptional experiences. Reputation seems capable of enhancing an experience even in retrospect, as when we only discover the next day that we had just heard a diva or visited the season’s hottest exhibition. But hindsight can affect the remembered experience, it cannot affect the past experience itself (e.g., Kahneman, 2005). Can the *actual* experience of a poem (rather than the *expected* or *remembered* experience) be affected by expectations in real time? And if so—how does it happen?

When the experience of a product is changed by its label, the change does not occur in the product. It would be undetected by an audio recording or a photograph. This is the intuition underlying Juliet’s famous: “That which we call a rose by any other name would smell as sweet.” However, experience is not determined by bottom-up processes alone. Distal stimuli are only experienced through

the proximal stimuli to which they give rise, and total experience takes place “in the eyes of the beholder”—or even, ultimately, in the mind of the beholder. Hence, for cognitive psychologists, it is obvious that expectations can alter experience (see, e.g., Wilson, Lisle, Kraft & Wetzel, 1994).

Social psychologists and cynics, however, will be quick to point out that one cannot rely on people’s reports of their experiences to decide the matter, because reports are not always sincere or unbiased. A glowing evaluation can be a form of cognitive “snobbery” (“it’s supposed to be excellent, so I will say it is excellent”), or of social desirability.

This caveat notwithstanding, in studies of consumer behavior, expectations are typically manipulated through brand name, price, or other marketing actions, and evaluations are solicited via expressed judgments or revealed preferences (see surveys in Shiv, Carmon & Ariely, 2005; Lee, Frederick & Ariely, 2006).

Allison & Uhl (1964) were among the first to study these variables. They found that people could not identify their favorite brand of beer in blind tasting. Later results were more startling, such as that blind tasters cannot distinguish dog food from pâté (Bohannon, Goldstein & Herschkowitsch, 2009) or that experienced violinists cannot distinguish Stradivari violins from new violins (Fritz, Curtin, Poitevineau, Morrel-Samuels and Tao, 2012). On the other hand, judges purport to distinguish among identical stimuli, when these are labeled differ-

---

Over the years, this paper was presented in many colloquia, and was read by many people, some as referees. We wish to thank the many friends and colleagues who commented and helped, but cannot list them all. We thank in particular: Jon Baron, David Budescu, Shane Frederick, Amia Lieblich, Yaacov Ritov and Roi Zultan

\*Center for the study of rationality, The Hebrew University, Jerusalem 91904, Israel.

†The National Institute for Testing and Evaluation, Jerusalem.

‡Adam-Milo Institute, Jerusalem.

ently (e.g., labeling isovaleric acid as “cheddar cheese” versus “body odor”; de Araujo, Rolls, Velazco, Margot & Cayeux, 2005) or even just framed differently (e.g., labeling beef as “75% fat free” versus “25% fat”; Levin & Gaeth, 1988).

When comparing informed evaluations to blind evaluations, or ratings for differently labeled but identical products, some researchers automatically assume that a real change occurred in the experience (e.g., Makens, 1964, p. 261: “a well-known brand positively affected *the taste* [italics ours] which Ss experienced for samples of turkey meat”), while others assume that it will not (e.g., Goldstein, Almenberg, Dreber, Emerson, Herschkowitsch, & Katz, 2008, p. 1: “non-expert wine consumers should not anticipate greater enjoyment of the intrinsic qualities of a wine simply because it is expensive”). In fact, however, we should acknowledge that it is certainly possible, psychologically speaking, that the actual experience *is genuinely different* for blind and for informed consumers, as it is possible that the actual experience *is just the same* for blind and for informed consumers.

The question is clearly an empirical one, and, moreover, its answer could well differ from context to context, or from individual to individual. Yet few studies have tackled this problem. Lee et al., (2006), when reviewing the literature, stated that “. . . it remains unclear whether [manipulating the participant’s] knowledge also changes the experience itself . . . , just as it remains unclear in most taste-test studies whether brand identity is just another input to . . . overall evaluation . . . or whether it modifies the actual gustatory experience” (p. 1055). Their own study is an exception, and will be described in Study 2 below. The present paper is another exception.

Few studies on expectation effects used cultural products. Yet cultural products are of particular interest, both because of our intrinsic interest in them, and because the question of how expectations affect cultural experiences, and how to distinguish between sincere effects and cynical or hypocritical ones, is particularly vexing with regard to ineffable or ambiguous experiences, such as artistic ones. Whereas wines, energy drinks and pain-killers affect the consumer’s physiology, lending credence to the term “marketing placebos” (Shiv et al., 2005), cultural products such as paintings or music are consumed primarily for their effect on the mind. It is harder to test whether a mental experience is altered than whether a physiological one is. This conundrum has itself been the focus of various cultural products (e.g., Yasmina Reza’s play *Art*). The present paper will focus on such an ineffable product—poetry.

Study 1 sets the stage by establishing the effect we will later study in depth. It consists of 2 experiments. Experiment 1 shows that readers of poetry are influenced by the poet’s name. Experiment 2 incidentally adds that

without the cue to quality imparted by the name of a reputable poet, readers cannot reliably distinguish good poetry from bad. Taken together, Study 1 shows that poems’ ratings are sensitive to the poet’s reputation, but not to the poem’s quality. That raises the sad possibility that the effect may be wholly due to pretension or to social desirability, as many outside critics of modern and contemporary art suspect.

Study 2 sets out to explore this question. It tests one particular model that we call The Emperor’s New Clothes effect (ENC, for short), honoring Anderson’s famous parable. According to this model, the reading of the poem is the same with or without the poet’s name, giving rise to the same aesthetic experience; the enhanced rating is solely due to a deliberate and conscious adding of points when the poem is attributed to a famous poet, motivated perhaps by a desire to appear discriminating and cultivated.

Study 3 tests an alternative model, which posits that the inclusion of the poet’s name alters the very experience of the poem, so that once the poem’s author is known, the poem is no longer “the same”. In other words, the poem—unchanged on the written page—is somehow changed in the reader’s mind. This we study by looking at judgments of many specific poem attributes.

We regard the main, and novel, contribution of this paper not in showing what happens, even in this previously unstudied context of poetry appreciation, but rather in attempting to understand the mental process whereby it happens. In particular, we offer experimental paradigms that allow one to infer whether the enhanced evaluation of a poem (or any object), when labeled in an expectation-raising manner, is driven by deliberate social considerations (a System 2 product), or happens out of awareness (a System 1 product). Is it an unfortunate social bias, or an inevitable cognitive bias? The answers have ethical as well as scientific ramifications, inasmuch as they pertain to the merits and drawbacks of “blind judgment”.

## 2 Study 1—Poem or poet?

Recently, a professional wine critic published a book called *The Wine Trials* (Goldstein, 2008). Although not a scientific book, it is based on an intriguing experiment (Goldstein et al., 2008), the abstract of which states: “Individuals who are unaware of the price do not derive more enjoyment from more expensive wine. . . . on average [they] enjoy more expensive wines slightly *less* [italics mine].” (p. 1). Not so when the wine’s price is known. Plassman, O’Doherty, Shiv & Rangel (2008) asked their participants to taste five wines. Unbeknownst to them, the same wine was presented once with its true price and once with a more expensive, or less expensive, price tag. The

participants' expressed preferences, bolstered by fMRI evidence from their brain scans, indicated that they enjoyed a wine more when they *thought* it was expensive, rather than when it really *was* expensive.

Even more recently, Fritz et al. (2012) asked 21 experienced violinists to compare [3] violins by Stradivari and Guarneri del Gesù with [3] high-quality new instruments" (p. 760) under double-blind conditions. The total market value of the former was about 100 times that of the latter. "Player's judgments about a Stradivari's sound may be biased by the violin's extraordinary monetary value . . . , but no studies designed to preclude such factors have yet been published" (p. 760). Not unlike the wine studies, the authors reported: "We found that (i) the most-preferred violin was new; (ii) the least-preferred was by Stradivari; (iii) there was scant correlation between an instrument's age and monetary value and its perceived quality; and (iv) most players seemed unable to tell whether their most-preferred instrument was new or old" (p. 760).

Neither of these studies attempted to find out what mental process, exactly, caused the difference between blind and informed judgments. To borrow an expression from Goldstein (2008), the effect can be attributed to "the taste [or sound] of money" (p. 12). We have given the taste of wine and the sound of violins special attention in this brief review, because, like poetry, the experiences they give rise to are perhaps more complex and subtle than ordinary consumer products.

In the present research, poems replace wine or violins, and poet's reputation replaces price.<sup>1</sup> Poetic analogues of "expensive" come naturally. We selected four Israeli poets (Yehuda Amichai, 1924–2000; Nathan Zach, 1930– ; Leah Goldberg, 1911–1970; Dalia Rabikovitsh, 1936–2005) from the literary canon—they are critically acclaimed, received prestigious prizes and awards, are included in Israel's high-school curriculum, and are well represented in major poetry anthologies. We chose 2 poems for each poet from collections regarded as central to their output—though not their best-known poems, to reduce the chance that our participants will recognize the poems. All poems were short, ranging between 12 and 18 lines, and up to 100 words.

Analogues of "cheaper" were harder to come by. Arguably, any published poem is "good" in some minimal sense (e.g., it passed the threshold for publication), as is any poem by a poet of high repute. We wanted to avoid debating the quality of our "bad poems", and yet give them a fighting chance (as the high-quality new violins, or the store-carried wines, have). We opted for generat-

ing the "bad poems" ourselves, while constraining them to resemble the "good poems" superficially.

Each of the authentic poems was mimicked by one that we generated ourselves.<sup>2</sup> For example, for the genuine poem that was a sonnet, we wrote a counterpart that was also a sonnet; the genuine poem whose rhyming pattern was A B C A B D D E F D D F, had a similarly rhyming counterpart; etc. The imitation poems also aimed for a similar number of words and similar vocabulary richness.<sup>3</sup> For our "unesteemed poets", we made up four bogus poems, using common Hebrew names with little cultural connotations.<sup>4</sup>

Study 1 consists of two experiments. In the first, participants rated poems, with or without poets' names. In the second, they had to distinguish between real poems and faked ones.

## 2.1 Experiment 1

### 2.1.1 Method

*Design.* Table 1 shows the 8 between-subject conditions. Authentic poems were paired either with the name of the famous poet who wrote them, or with a bogus name of the same gender. Fake poems were paired either with the name of the poet whose poem they mimicked, or with a bogus name of the same gender. Participants read and rated 4 poems each -- either those written by the two male poets (authentic poems or fake poems, but not both) or those written by the two female poets (likewise). Their 4 poems were all either attributed to the famous poets, or to bogus poets.

*Participants:* Respondents were 281 students, mostly undergraduates, mean age 25, 59% female, all fluent in Hebrew (this after discarding the data of 8 participants who didn't recognize the names of one or more of the four famous poets; 17 who recognized one or more of the eight authentic poems; 8 who "recognized" the bogus poems; and 2 who "recognized" a fake poem).

*Procedure:* Participants were approached either individually or at the end of class, and asked to answer a short questionnaire (which took up to 15 minutes). They were promised participation in a lottery for five prizes of 400 NIS each (then about \$100). Participants were randomized into the 8 conditions, and asked to "rate the quality of the poem" on a scale from 0 to 100. The questionnaires also elicited some personal data, such as the

<sup>2</sup>Our poems were in Hebrew, but to give their flavor, the Appendix contains a real poem by Emily Dickinson, and an imposter poem, generated in the same quick and rough manner that we used for Experiment 1.

<sup>3</sup>The Wine Trials also compared a Chardonnay to a Chardonnay, a Merlot to a Merlot, etc.

<sup>4</sup>The names were Rivka Sela, Hanna Caspi, Benjamin Shakhar and Shalom Dagan. Hitherto they will be referred to only by their initials, as in Table 1.

<sup>1</sup>Our own original idea was also conceived with regard to wine, but we switched to poetry because it is so much simpler and cheaper to run, requiring only paper and pencil, and minutes of the respondents' time. Chronologically, the poetry studies all preceded the wine and violins studies.

Table 1: Design and results of Experiment 1.

Poem	Poet	Mean	SD	Poem	Poet	Mean	SD
<u>Authentic poetry, Famous poetess, N = 30</u>				<u>Fake poetry, Famous poetess, N = 37</u>			
Olive trees	LG	82	12	Wheat fields	LG	75	13
Road to Granada	LG	75	15	Road to Siberia	LG	74	15
In praise of peace	DR	77	17	True dream	DR	78	13
The blue lizard	DR	75	16	The girl sleeping in the garden	DR	75	15
<u>Authentic poetry, Famous poet, N = 32</u>				<u>Fake poetry, Famous poet, N = 42</u>			
Sitting on the curb <sup>a</sup>	NZ	76	18	Museum visit	NZ	73	17
Sometimes when it's late	NZ	75	19	Sometimes when watching TV	NZ	72	19
Verbs sonnet	YA	73	22	Numbers sonnet	YA	76	14
Now, when the water surges	YA	77	18	Yesterday, when the earth quaked	YA	81	20
<u>Authentic poetry, Bogus poetess, N = 38</u>				<u>Fake poetry, Bogus poetess, N = 31</u>			
Olive trees	RS	74	13	Wheat fields	RS	72	17
Road to Granada	RS	67	17	Road to Siberia	RS	68	21
In praise of peace	HC	74	18	True dream	HC	72	18
The blue lizard	HC	76	15	The girl sleeping in the garden	HC	74	14
<u>Authentic poetry, Bogus poet, N = 34</u>				<u>Fake poetry, Bogus poet, N = 37</u>			
Sitting on the curb	BS	70	19	Museum visit	BS	62	27
Sometimes when it's late	BS	65	24	Sometimes when watching TV	BS	64	19
Verbs sonnet	SD	66	26	Numbers sonnet	SD	70	19
Now, when the water surges	SD	71	15	Yesterday, when the earth quaked	SD	73	19

<sup>a</sup> One data point was missing in this cell, which is therefore based on just 31 observations.

respondent's educational background in literature. After all data had been collected, participants were debriefed, and informed about the experiment and its results.

### 2.1.2 Results and discussion

Table 1 presents mean ratings and standard deviations of the individual poems in each condition. A 3-way ANOVA was performed, with the following factors: Authentic vs. fake poem; Famous vs. bogus poet; Male vs. female poet. Individual poets and poems were treated as repeated measures.

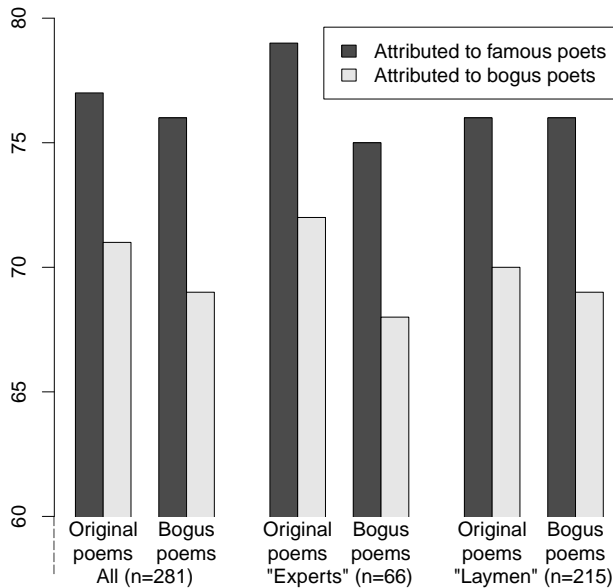
Poet reputation was the only significant effect: poems attributed to famous poets were rated higher ( $M=76$ ,  $SD=12$ ) than poems attributed to bogus poets ( $M=70$ ,  $SD=15$ ;  $F(1,273)=14.65$ ,  $p<.001$ ). Authenticity made no difference—both real and fake poems were rated 73 on average. Poet's gender was not significant, with women's poetry rated 74 ( $SD=12$ ) on average and men's poetry 72 ( $SD=15$ ) on average ( $F(1,273) = 3.16$ ,  $ns$ ). None of the

interactions was significant.

Figure 1 shows the effects of poem quality and poet reputation, collapsing over poet's gender and the individual poets.

We attempted to see whether "experts" would do better. We did not test professional experts, who likely would have recognized the authentic poems, rendering our test moot. Our "experts" were students who had some background in literature—66 had either taken (the Israeli equivalent of) Advanced Placement classes in Literature in high school (38), or majored in Literature at the university (35; 7 had done both). The experts were as influenced by the poet's name as the others (expert-by-poet interaction  $F(1,273) = .029$ ,  $ns$ ). Any discrimination shown by the group as a whole is due in its entirety to this subgroup (though experts are hardly more discriminating than laymen, 2-way interaction  $F(1,273) = .734$ ,  $ns$ ; the 3-way interaction is also not significant,  $F(1,273) = .047$ ). Moreover, we cannot rule out the possibility that some experts, even if unawares, recognized some of the poems.

Figure 1: Mean rating for real and for fake poems, when attributed either to famous poets or to bogus poets, for the entire sample; for “experts” only; and for “laymen” only.



### 3 Experiment 2

Our respondents showed no more appreciation for authentic poems than for fake poems. Would they distinguish between them better if both were presented together?

#### 3.1 Method

*Design.* Participants in 8 groups were given one of the 8 pairs of poems used in Experiment 1—the real thing and its imposter—and told as much, with the poet identified. They were asked to guess which poem is which, and indicate their confidence.

*Participants and procedure:* Respondents were 245 students (after discarding 3 who recognized one or more of the authentic poems), mostly undergraduates with various majors, mean age 26, 57% female, all fluent in Hebrew. They were recruited, instructed, and rewarded, as in Experiment 1. Questionnaires (which took only minutes to answer) were distributed at random, and respondents were promised participation in a lottery for a 200 NIS prize

#### 3.2 Results

Table 2 orders the pairs by decreasing rate of correct identifications. Authentic poems were correctly identified between 43% and 74% of the time (we chalk the Zach poem which is an exception to sampling error), with a mean

of 54% - hardly better than chance (binomial test, *ns*), and compatible with the results of Experiment 1. Mean confidence in the judgments was 67%, exhibiting the familiar pattern of overconfidence (67% vs. 54%, exact binomial test  $p < .003$ ) in forced choice tasks with difficult items (Lichtenstein, Fischhoff & Phillips, 1982). Respondents with an extended background in literature (“experts”,  $N=56$ ) did somewhat better than the rest, albeit not significantly (60% correct compared to 53%, Fisher’s exact test, *ns*), and expressed higher confidence (71 vs. 65,  $t=2.68$ ,  $DF=225$ ,  $p<.01$ ).

#### 3.3 Discussion

The results of Study 1 beg the question whether the fake poems might not have been as bad as we thought. Can a faked poem, deliberately devoid of any artistic intent, nonetheless be “good”? Artists who believe what they produce is good, while critics consider it bad, are commonplace. But can the opposite occur? Did we inadvertently produce good poems?

Not being philosophers or critics of art, our own opinions on this matter are of little merit. But we stress that it was never our intention to write poems with any artistic value—quite the opposite (we spent little more than 10–15 minutes per poem, giggling the while). It has been argued (e.g., Livingston, 2005) that artistic intention is a necessary condition for some human productions to be considered art (and, as in the case of Marcel Duchamp’s notorious urinal, even a sufficient one).

In a few notorious cases, a project designed to parody art or to forge art, rather than to actually *be* art, was so successful, that its esteem survived exposure. A notable example is the poetry of Ern Malley, a fictitious poet invented in the 1940s as a hoax by two Australian poets, whose own serious work was overshadowed with time by their parody (e.g., Heyward, 2003). Similarly, the forged paintings of Elmyr DeHory continued to command high prices and professional respect even after the truth about them emerged (e.g., Irving, 1969). These, however, are exceptional stories, and we have no reason to believe we possess the talent to have produced good poetry inadvertently. Indeed, for present purposes we are happier when friends deride our poems than when they praise them.

A second issue raised by our results feeds into the ongoing debate as to whether the merit of works of art is inherent or is a social construction; whether it is apparent without the signaling by various social cues or whether it totally depends on them. This debate is more important for art than it is for cognition, and in the present paper we will discuss it no further.

Table 2: Rates of correct identification.

Poet	Poem pair	N	% Correct	Confidence
Goldberg	Road to Granada / Road to Siberia	30	43	63
Amichai	Verb sonnet / Number sonnet	34	47	69
Rabikovitch	Blue lizard / Girl in garden	29	48	69
Amichai	Water surges / Earth quakes	30	53	66
Zach	On the curb / Museum visit	30	53	63
Rabikovitch	Praise of peace / True dream	30	57	68
Goldberg	Olive trees / Wheat fields	31	58	69
Zach	When it is late / When I watch TV	31	74	69
Overall		245	54	67

#### 4 Study 2—Do raters of poetry deliberately add points for a poet of note? Testing The Emperor’s New Clothes model

Study 1 established, at least for our respondents and our poems, that laymen cannot reliably distinguish good poetry from fake poetry, and their ratings can be swayed by changing the poem’s attribution. This raises the obvious question whether the effect is due to hypocrisy, or whether there is valid information in a poet’s name that justifies it.

Clearly, in some cases knowing who authored something gives information that not only alters judgment, but actually *improves* it. For example, a paper (or mathematical proof, or legal argument, etc.) may be hard to follow because it is deep and complex, or because it is confused and incoherent. Knowing who wrote it could resolve this ambiguity. Authorship sometimes even affects a text’s truth value, most notably in so-called indexical propositions (see, e.g., Perry, 1997).<sup>5</sup>

On the other hand, where one believes that “a rose by any other name would [or even *should*] smell as sweet”, rating the selfsame poem differently under different attributions could be awkward. The prevalence of blind tasting, blind auditioning, blind reviews, etc. suggests that biased judgments are considered normatively unwarranted and ethically objectionable. After all, a naked King cannot be clothed by the mere patter of his cunning tailors.

Goldin and Rouse (1997) showed that orchestra auditions carried out behind a screen that hides the candidate from the jury increase the probability of hiring women. A study done in the American Economic Review showed that when referees do not know the identity of the au-

thors of the papers they are reviewing, authors at near-top-ranked or nonacademic institutions have lower acceptance rates than when refereeing is not double blind (Blank, 1991). These results suggest the superiority of blind judgments insofar as they cannot be subject to discriminatory biases of dubious validity.

Study 2 tests the crudest form of bias, which we call the Emperor’s New Clothes effect (ENC). Specifically, we ask whether a public evaluation of an attributed poem consists of a private evaluation of the poem “in itself” that ignores the poet’s name, which is then consciously and deliberately adjusted to accommodate the reputation of the poet, perhaps due to various social considerations. The model,  $RJ = SJ + NM$ , states that Reported Judgment equals Sincere Judgment, plus Name Premium, and the latter is added consciously.

We did not deem it prudent to ask our respondents directly whether their Reported Judgment included a Name Premium added onto their Sincere Judgment, on the suspicion that insincere raters are unlikely to answer us sincerely. The challenge, then, was to elicit sincere judgments while finessing social desirability.

Lee et al. (2006) faced a similar challenge. Pub patrons tasted regular beer and “MIT brew” (beer laced with balsamic vinegar, which Lee et al. call “conceptually offensive”, p. 10). Some tasted the two beers blind. Others were informed before tasting. Blind tasters preferred the MIT brew. Informed tasters preferred the regular beer. Were the informed tasters expressing their sincere preference, or was their report shaped by social desirability? To answer this, a third group was given the tasting experience of the blind tasters, but the evaluation opportunity of the informed tasters (namely, they were informed of what they had drunk after the drinking, but before the evaluation). This group resembled the blind tasters, not the informed tasters. Apparently, their tasting experience was not altered retroactively by the “mildly unsettling news”

<sup>5</sup>E.g., “I was born in 1957.”

of the balsamic vinegar lacing (p. 1056). Moreover, they declined an opportunity to report a more socially desirable, albeit insincere, evaluation.

Our design is necessarily different<sup>6</sup>, although we also had three kinds of readers: “blind” readers who read an unattributed poem, “informed” readers who read it with the poet’s name, and readers who were blind when first reading the poem, and were informed of the poet’s name only after reading the poem.<sup>7</sup> We elicited ratings from the third, and critical, group in an indirect way. Rather than requesting them to first give a rating based on their blind reading, and then when informed of the poet’s name to give a second rating, after informing them we asked them to guess the rating of “other people like you”, hoping thereby to solicit more sincere evaluations.

Our rationale (which our results verified) was two-fold. First, we assumed that when one is guessing how a similar other rates a poem, one first asks oneself “How would I rate this poem?”, there being little else to draw upon. So asking about another is tantamount to asking about oneself. Second, we assumed that one feels less impelled to protect an anonymous other from an embarrassing admission (for a similar rationale see, e.g., Fischhoff, 1975). By asking participants how they think other people are affected by a poet’s name we compel them to introspect, while removing any reluctance to report their introspection sincerely (see Fisher, 1993).

#### 4.1 Method

*Materials.* Study 2 used a single poem by Yehuda Amichai, arguably Israel’s favorite poet. The poem chosen, *Infinite Poem*, was loose enough in form and structure that the poetic skill it required was not as apparent as when strict rhyme and rhythm constraints are imposed. This rendered its evaluation deliberately ambiguous. Some participants read the poem with, and some without, the poet’s name. We contend that when manipulated between-subjects, either heading (“*Infinite poem*, by Y. Amichai” vs. just “*Infinite poem*”) triggers no awareness that the independent variable of interest is presence or absence of the poet’s name.

Infinite Poem, by Yehuda Amichai (Translated from Hebrew by MBH)

Within a modern museum  
an old synagogue.  
Within the synagogue  
myself.  
Within me

my heart.  
Within my heart  
a museum.  
Within the museum  
a synagogue,  
within it  
myself,  
within me  
my heart,  
within my heart  
a museum.

*Participants:* A convenience sample of 511 Hebrew speakers participated in this study. All were graduates of Israeli high schools. They ranged in age from 17 to 74 (mean age=30), and 61% were female. Groups 1 and 2 were students who answered the questionnaire in a classroom. The rest were approached individually, and asked to answer a short questionnaire (up to 10 minutes), for a chance to win a monetary reward.

*Design and procedure:* Each participant received a questionnaire with *Infinite Poem* on its first page. Ratings of its “literary quality” were solicited on a scale from 0 (“total rubbish”) to 100 (“totally wonderful”). At the end of the task, they were asked to provide some personal details (e.g., gender, age, education). Two groups, G1, “blind readers” and G2, “informed readers”, read the poem with or without knowing who wrote it. The other four groups, after rating the poem themselves, were also asked to guess the mean rating of a group of other readers, described as “like themselves”. Two of these groups were asked to guess the mean rating of other readers holding the same authorship information as themselves (G3, “blind readers” guessed other “blind readers”; G4, “informed readers” guessed other “informed readers”). The fifth group, G5, read the poem blind, but were then told it was by Amichai, and asked to guess the rating of other readers who, unlike themselves, were informed at the time they rated it (similarly to Lee et al.’s third group). G6 read no poem and evaluated no poem, and will be described in the following results and discussion section.<sup>8</sup>

#### 4.2 Results and discussion

The first two groups establish the effect which we are trying to model. The blind readers gave the poem a lower mean rating, 54, than the informed readers, 63. This 9-point difference was significant ( $t = 2.09$ ,  $DF=150$ ,  $p<0.05$ ), and is the same order of magnitude as was found in Experiment 1.

<sup>8</sup>Besides the 6 groups reported in Table 3, 6 other groups were omitted from this paper, to make it easier to follow. They encumber, but do not alter, the reasoning behind Study 2. For a full description, the interested reader is referred to Bar-Hillel, Maharshak, Moshinsky & Nofech, (2010).

<sup>6</sup>Our study was conducted several years before Lee et al.’s.

<sup>7</sup>In Lee et al.’s study, the analogue of “giving poet’s name” was “doctoring the beer with balsamic vinegar”, so “informed” carries a different meaning here and there.

Table 3: Own ratings and guessed ratings of the experimental groups.

Group	Task	N	Own rating	Guess rating	Own SD	Guess SD
G1	Blind readers	69	54	–	30	–
G2	Informed readers	83	63	–	23	–
G3	Blind rating, then guess blind others	93	52	56	29	22
G4	Informed rating, then guess informed others	102	62	65	28	22
G5	Blind rating, then guess informed others	71	47	80	29	19
G6	“What is Amichai’s name worth?”	93	30	–	13	–

The next two groups constitute a manipulation check. The “guess-others” strategy assumes that when asked to guess the rating of someone else, our participants first introspect, and then project (namely, they first ask “What would I do?”, and then assume the other would do the same). Do the results support this assumption?

G3 read the poem without attribution, and rated it. They were then told that another group of people “like themselves” had previously evaluated the poem, and were asked to guess the mean evaluation given to the poem by those other people. Similarly, G4 rated the poem with Amichai’s name and then also guessed the mean of similar others. A reward of 100 NIS was promised to the most accurate guessers, with accuracy determined by comparison with the benchmark results of G1 and G2, respectively. The reward was intended to motivate participants to give the best—hence the most sincere—guess they could. If our guess-another manipulation is valid, both groups should be successful in their predictions.

Indeed, G3, the blind readers, rated the poem on average 52 themselves, and guessed a mean of 56 for the rating of other blind readers ( $t=1.59$ ,  $DF=92$ ,  $ns$ ). G4, the informed readers, gave the poem a mean rating of 62 themselves, and guessed a mean of 65 for other informed readers ( $t=1.41$ ,  $DF=101$ ,  $ns$ ). The slight upwards drift (even when combining G3 and G4) was not significant. Moreover, the modal difference between own rating and guessed rating in both groups was 0 (the SDs for both differences were between 19 and 20). Most importantly, the effect of the poet’s name is preserved. Thus, the results support our rationale.

We can now put ENC to an actual test. Recall that the ENC model,  $RJ = SJ + NM$ , states that Reported Judgment equals Sincere Judgment plus Name Premium, and assumes that raters are aware of this. We were concerned that our informed raters would deny adding a Name Premium, passing off their Reported Judgments as Sincere Judgments. To get around insincere self-reporting, we asked them about other people rather than about themselves, thereby removing any motive to enhance self-presentation. The participants of G5 were thus *in effect*

asked for the impact of the poet’s name on their own ratings, while *in fact* were asked to guess the impact of the poet’s name on other people’s rating (guesses here were rewarded similarly to before).

Under the ENC model, G5 participants should have been as successful in their guesses as were G4 participants. ENC predicts that raters have access by introspection to Amichai’s Name Premium (which we know from the earlier results to be about 9–10 points), and that they will add it to their own just-rendered Sincere Judgment, and report the outcome. In fact, however, G5 participants raised their own blind rating of 47<sup>9</sup> by a whopping 33 points ( $t=10.85$ ,  $DF=70$ ,  $p<0.0001$ ), guessing 80 for the mean rating of informed others.<sup>10</sup> This spectacular failure to guess G2 suffices to reject ENC.

We conclude that G2 participants were not rating the poem in the manner assumed by the ENC model, because that manner would, counterfactually, have been accessible to G5 participants as well.

If the 30+ points believed to have been added by the poet’s name did not come from introspection, where did it come from? The results of G6 can help us here. G6 participants were given no poem at all to read, and none to rate. They were told only: “Imagine people reading and evaluating a poem on a scale from 0 to 100. Some read it unattributed, and others know it is by Yehuda Amichai. What do you think would be the mean difference between the two groups?” Their mean guessed difference was 30 points.<sup>11</sup> Its inflated magnitude might well result from the focusing illusion.

“The idea of a focusing illusion involves hypotheses about two psychological processes, one in the subject whose experience is predicted [here G2, informed read-

<sup>9</sup>We cannot account for this unusually low mean, and chalk it to sampling error.

<sup>10</sup>This addition is also significantly higher than the true difference of about 9 points, even if we adjust the unusual 47 rating to G1’s 54. To test for significance, we calculated as if this adjusted 26 point difference were obtained between-S, and compared it to the 9-point difference between G1 and G2 ( $F$  interaction=9.2,  $DF=1, 338$ ,  $p<.003$ ).

<sup>11</sup>The difference between G6’s 30 points and G5’s 33 points is not significant, ( $t=1.199$ ,  $DF=162$ ,  $ns$ ).



ers], and the other in the judge who makes the prediction [here G6]" (Schkade & Kahneman, 1998, p. 340). Variables carry more weight for judges who focus on them than for those who do not (for evidence see also, e.g., Lowenstein & Frederick, 1997; Schwarz, 1996). From Table 3 (and from Experiment 1 in Study 1), we know that Amichai's name adds fewer than 10 points to the ratings of the subjects whose experience is actually measured. G6 participants, on the other hand, are the judges who predict that the addition could amount to 30 points or more.

The focusing illusion is mitigated when one has been personally exposed to the changes in the target variable, rather than having to guess their effect. If you have experienced a change, you will judge its effect from how it affected your experience, rather than from an (inflated) theory about its impact. For example, people who know paraplegics are not subject to the same overestimation of the impact of this misfortune on the paraplegics' happiness as those who do not (Schkade & Kahneman, 1998); likewise, people asked how they expect changes to affect their future well-being give higher estimates than when judging how such changes had affected them in the past (Lowenstein & Frederick, 1997).

Had G5 participants been able to project themselves into the shoes of G2 participants—a task that G4 participants performed with no difficulty, and that the ENC model assumes can be done with no difficulty—they could have drawn on this experience to assess the impact of the poet's name, and consequently would not have erred as they did. Since they could not do so (which is why we rejected the ENC model), they had to rely on their theory of the name's impact (as given by G6 participants), thereby greatly exaggerating it.

Since the 10 point difference between informed and uninformed readers was not added deliberately, where did it come from? We address this question at the very end of Study 3.

## 5 Study 3—Interpreting a poem in light of its author

If the difference between how informed and blind readers rate Amichai's Infinite Poem does not result from a deliberate addition of a Name Premium to an otherwise identically experienced poem, how can it be accounted for?

An intuition that contrasts with Juliet's is embodied in the aphorism: "*Beauty is in the eye of the beholder.*" Such is the power of suggestion that sometimes a naked King can look magnificent in his non-existent clothes, and a rose can smell like a rotten egg. The scent emitted by a rose depends, of course, on the rose's chemistry (bot-

tom up). Importantly, however, perceived scent also depends on top-down factors such as what is in the smeller's nose membranes, brain, and mind (e.g., de Araujo et al., 2005). The experience of stimuli can be altered without altering the physical stimuli themselves.

Wine, violins, poultry and poetry all yield better experiences when sporting reputation-enhancing labels. In Study 3, we study the possibility that knowing who wrote a poem alters the way the text is interpreted,<sup>12</sup> because different associations are primed thereby. Literary *mavens* we consulted pointed out, for example, that the motif of a synagogue appears frequently in Amichai's poetry. Among the erudite, the poem elicits associations to those other poems, which might not be elicited without Amichai's name. Similarly, in wine tastings, Morrot, Brochet and Dubourdieu (2001) found that when people tasted a white wine, they tended to describe its taste with white-wine adjectives such as "honey" and "lemon". When that same wine was dyed red with a flavorless dye, they switched to red-wine adjectives such as "cherry", "blackcurrant", etc.

We perused literature dealing in poetic criticism, extracting a list of adjectives commonly used when poetry is discussed or evaluated. Could Amichai's name have caused the attributed poem to be read differently than the unattributed poem with regard to some of these adjectives? If so, that would lend concrete meaning to the hypothesis that the poet's name altered the very experience of the poem, and not just its perceived, or reported, quality.

### 5.1 Method

*Participants and Procedure.* There were 324 participants, 56% of them female, ranging in age from 18 to 63, with a mean of 29. All were Israeli high-school graduates, and most were students, who were run in groups at the end of classes. They were asked to answer a short questionnaire (up to 10 minutes), and promised participation in a lottery for a 500 NIS prize.

*Stimuli and design.* We generated 24 pairs of adjective antonyms (albeit, with redundancies), as listed in Table 4. 165 respondents were asked to read Infinite Poem, either with Amichai's name ( $N=79$ ) or unattributed ( $N=86$ ). The poem was followed (on the next page) by a 7-point semantic differential, corresponding to these 24 paired adjectives, which respondents were asked to scale. For example:

short 1 2 3 4 5 6 7 long

<sup>12</sup>Jon Baron says: "I want to know the author of a paper I review and not review it blind, because I think the author is relevant. Knowing the author affects the way I interpret things that are said."

Table 4: 24 adjective pairs for evaluating poetry, and their ratings.

The adjective pairs	“Good Poetry”		Infinite Poem unattributed		Infinite Poem by Amichai		“Amichai’s poetry”	
	M	SD	M	SD	M	SD	M	SD
*1. rich-poor	5.7	1.2	3.6	1.6	4.4	1.5	5.7	0.9
*2. polychromatic-monochromatic	5.3	1.2	3.2	1.8	3.7	1.6	4.8	1.3
*3. connected-detached	5.2	1.3	4.1	1.7	4.6	1.5	5.2	1.1
*4. personal-general	5.0	1.2	5.1	1.8	5.7	1.4	5.0	1.4
5. colorful-gray	5.0	1.1	2.9	1.4	3.3	1.6	4.2	1.6
6. emotional-intellectual	4.9	1.1	4.5	1.5	4.8	1.5	4.8	1.4
7. optimistic-pessimistic	4.7	1.0	3.6	1.5	4.0	1.5	3.1	1.4
*8. mature-childlike	4.7	1.1	4.9	1.5	5.4	1.1	5.5	1.0
*9. sophisticated-unsophisticated	4.7	1.4	3.9	1.7	4.6	1.5	5.1	1.2
10. soothing-irritating	4.7	1.4	3.5	1.6	4.0	1.4	4.0	1.2
*11. modest-boastful	4.6	1.2	4.6	1.5	5.1	1.3	5.0	1.2
12. romantic-cynical	4.6	1.2	4.0	1.6	4.4	1.3	3.6	1.5
*13. refined-coarse	4.6	1.3	4.1	1.5	4.6	1.3	3.7	1.3
14. daring-conservative	4.5	1.0	3.7	1.4	3.8	1.4	4.6	1.4
15. revolutionary-conformist	4.5	1.1	3.9	1.5	4.1	1.5	4.7	1.4
16. secular-holy	4.4	1.2	3.9	1.5	3.8	1.4	4.5	1.4
17. fast-slow	4.4	0.9	4.2	1.7	4.6	1.7	3.9	1.2
18. modern-classical	4.3	1.2	4.6	1.6	4.8	1.5	5.0	1.3
19. clever-simpleminded	4.3	1.3	4.7	1.7	4.8	1.5	4.6	1.5
20. short-long	4.2	0.9	4.8	1.6	5.0	1.4	4.3	0.9
21. happy-sad	4.2	1.1	3.2	1.2	3.3	1.2	2.8	1.0
22. decisive-indecisive	4.1	1.1	4.0	1.8	4.2	1.9	4.5	1.5
23. unique-universal	4.1	1.4	4.5	1.9	4.6	1.7	4.1	1.7
24. direct-indirect	4.1	1.3	3.4	1.7	3.3	1.5	4.4	1.6
N	82		86		79		77	

A single order, randomly generated, was given to respondents (not the one in Table 4). Respondents were not asked to rate the poem’s overall quality. Indeed, neither the word “quality” nor any of its synonyms was ever mentioned at all.

The other respondents were not given any poem to read but were asked to characterize their idea either of “Good poetry” ( $N=82$ ), or of “Amichai’s poetry” ( $N=77$ ), using the same semantic differential.

## 5.2 Results and discussion

Table 4 orders the 24 paired adjectives according to the results of the group which characterized “Good Poetry”. Within each pair the first adjective is the one more closely associated, on average, with “Good Poetry” (hence necessarily rated higher than the midpoint, 4), and pairs are dis-

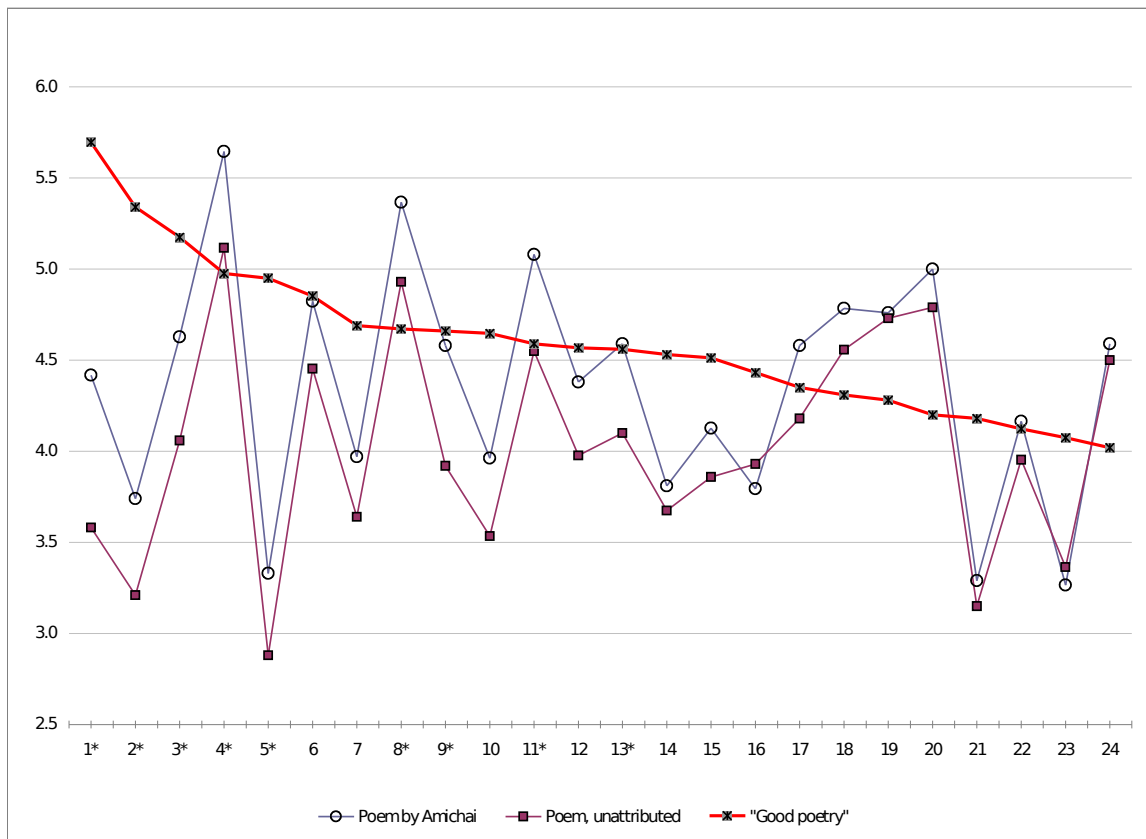
played from high to low in terms of the strength of their association with “Good Poetry”. Hence in the “Good Poetry” column the means are decreasing, and are always at least 4.

Figure 2 is a graphical presentation of the results in Table 4. The 24 adjective pairs<sup>13</sup> are on the abscissa, ordered as in Table 4. The ordinate shows the values on the semantic differential. The monotonically decreasing line is the “Good Poetry” profile, designed to be above the midpoint, 4.0, throughout. One jagged line is for the attributed poem (empty circles) and the other is for the unattributed poem (filled squares). Figure 3 (and Table 4) shows several things clearly.

First, the profiles of the attributed-poem and the unattributed-poem co-vary very closely. Their correlation is a remarkably high 0.93 (highly significant; all calcula-

<sup>13</sup>The adjectives sound somewhat better in the original Hebrew.

Figure 2: Profiles of attributed (circles) and unattributed (squares) poem, compared to “Good Poetry” (monotonically decreasing line).



tions are based on the unrounded numbers underlying the rounded-off numbers shown in Table 4), which is as high as the intra-group correlations, based on a Monte Carlo simulation.<sup>14</sup> In that sense, the 2 profiles look like 2 samples from the same population, in spite of the different conditions.

Second, the attributed-poem profile hovers above the unattributed-poem profile almost everywhere (excepting dimensions 16 and 24 only, exact binomial test,  $p < .0001$ ). This counters the possibility that the samples are derived from the same population. When testing whether any of the differences are significant, we found 8 dimensions on which the difference, considered on its own, would have been (1, 2, 3, 4, 8, 9, 11, and 13). The probability of getting as many as 8 significant results, at the .05 level, out of 24 possible trials, given the null hypothesis, is itself significant (exact binomial test,  $p < .001$ ). However, this calculation does not take into account that these are

simultaneous dependent multiple-comparisons. Applying the more conservative Bonferroni correction, only the first dimension, “rich-poor”, survives ( $t=3.46$ ,  $DF=163$ ,  $p=0.0007 < 0.05 / 24$ ). So it is not clear that the attributed poem profile can be said to be significantly higher than the unattributed profile on more than a single dimension. Be that as it may, our explanation for the upward drift is the same. Recall that Figure 2 was designed to show the “Good Poetry” line above the midline throughout. Hence, the higher the rating, the “better”, in some sense, it is; being rated higher is being judged a “better” poem.

Third, the attributed-poem profile and the unattributed-poem profile are usually on the same side of the midline (excepting 5 cases—1, 9, 12, 15 and 22; exact binomial test,  $p < .003$ ). In other words, inasmuch as the intensity of the rating for the attributed and unattributed poem differed, the directionality did not. For example, the “personal” unattributed poem became even more “personal” when attributed to Amichai (dimension 4), and the “sad” unattributed poem became less sad when attributed (dimension 21)—but a change such as from “conformist” to “revolutionary” (dimension 15) was rare.

<sup>14</sup>In the Monte Carlo simulation, both groups were randomly divided into two halves several thousands of times. Correlations were computed both within the group halves and between the group halves. The 3 resulting distributions were practically indistinguishable.

Table 5: Pairwise correlations between all targets.

	Amichai's poetry	Attributed poem	Unattributed poem
"Good poetry"	0.47	0.10	-0.20
Amichai's poetry		0.48	0.37
Attributed poem			0.93

A telling picture emerges from considering various correlations between the profiles. We correlated an adjective's mean rating on "Good poetry" with the mean advantage Amichai's name gave the poem on that dimension (namely, the difference between the attributed and unattributed poem). The same was done with regard to "Amichai's Poetry". Pearson's correlations were 0.78 and 0.36, respectively,<sup>15</sup> indicating that the poet's name contributed more to dimensions more closely associated with "Good poetry" (and to a lesser extent with "Amichai's poetry"). Indeed, note that the eight 8 dimensions on which the attributed poem differs most from the unattributed poem (marked by an asterisk) are concentrated in the top part of the 24 dimensions, as ordered by "Good poetry". The 9 dimensions on which the attributed poem differs least from the unattributed poem (14, 16, 18, 19, 20, 21, 22, 23, 24) are concentrated in the bottom of Table 4.

We conclude that the attributed poem, even though significantly different from the unattributed poem on almost none of the dimensions (except for being rated as significantly "richer"), is nonetheless perceived overall as consistently "better", and the more so the closer a dimension is related to "Good poetry".

Table 5 shows Pearson's correlations between the mean ratings of every pair of the four experimental groups across the 24 attributes (all correlations are highly significant; see footnote 15).

The correlations seem to be telling the following story: i. "Amichai's poetry" and "Good poetry" are correlated, but only weakly ( $r=0.47$ ). This is as it should be: since good poets have individual styles, not all "good poetry",

of course, is the same. ii. The attributed-poem correlates with "Amichai's poetry" ( $r=0.48$ ), as would be expected if a poet has a distinct individual style; but only weakly, since not all of Amichai's poems are the same. iii. Even with the poet's name withheld, some correlation between the unattributed poem and Amichai's poetry remains ( $r=0.37$ ), indicating that *Infinite Poem* carries some recognizable elements of Amichai's style even when unaccompanied by his name. iv. Both the attributed poem and the unattributed poem have negligible correlations with "Good poetry" ( $r=0.10$  and  $r=-0.20$ , respectively). This too makes sense, because whereas one might expect a particular style to characterize a particular poet's poetry, it is ludicrous to expect any particular style to characterize all good poetry (the task of the group that gave the "Good poetry" line notwithstanding). v. Despite these differences in how the two presentations of the poem correlate with "Good poetry" and with "Amichai's poetry", their correlation with each other, as noted before, is a remarkably high 0.93.

This overall pattern of correlations and distances can be reconciled by assuming that knowing that the poem is by Amichai creates a partly self-fulfilling expectation that the poem would be good,<sup>16</sup> priming a small but significant drift in the adjectives towards these expectations, but hardly altering the overall profile of the poem. Priming is an effect in which exposure to a stimulus lowers the threshold for responding to a later, associatively related, stimulus. In particular, it can occur between semantically related words. Inasmuch as in the eyes of our respondents some of the 24 words in our semantic differential, such as "rich", "sophisticated" and "connected", are semantically related to a poem's quality (and all are related to quality more closely than their antonyms), they are primed by the mention of Amichai's name, a poet recognized by our respondents as a fine and beloved poet of note. The threshold for attributing the primed adjectives to the poem decreases, and the mean rating on these adjectives increases. The effect almost always moves the poem's rating upwards, to the "good poetry" domain.

We believe that precisely the same thing occurred in Study 2. The 10 point difference between the informed rating of *Infinite Poem* and the uninformed rating is an unaware priming effect, where Amichai's name primed readers to read a "better" poem.

<sup>15</sup>Since the means that were correlated are based on about 80 Ss each, their reliability is not well-represented by taking N to be just 24. Instead, the statistical significance of the correlation between vector X and vector Y was calculated as follows: i. The vector was centered (namely, its mean was subtracted from each value). ii. The inner-product was calculated (namely,  $\sum_{i=1}^{24} X_i Y_i$ ). iii. Its SD was calculated using the formula  $SD = \sqrt{\text{Trace}(\Sigma_X \cdot \Sigma_Y)} / (24 \cdot 24)$  where  $\Sigma_X$  is the matrix that has X's 24 variances on the diagonal, and off diagonal has the covariances between the 24 judgments, and the trace is the sum of the terms in the diagonal of the product matrix. iv. The inner product divided by SD is distributed like Z. These 2 correlations, the 6 correlations in Table 5, and the pairwise differences to be discussed later, are all highly significant.

<sup>16</sup>MBH reports a striking and insightful moment: A friend recently gave me a jacket which she had bought and rarely wore. As I was admiring it on myself in the mirror, she happened to mention that it was designed by X, a famous designer. I distinctly remember how in front of my very eyes, the jacket mutated into a better looking jacket than it had been just a moment before.

## 6 General discussion

### 6.1 When do expectations have “real” effects?

We noted in the introduction that the studies showing that expectations influence ratings rarely give a process account for how this comes about. But some studies did show that the effect extends beyond ratings, and in that sense is “real”. These studies come in two kinds. One supplements behavioral data with brain scans. For example, Plassman et al. (2008) and McClure, Tomlin, Cypert, Montague, & Montague, (2004) showed that subjects’ changes in ratings or in choice were accompanied by changes in fMRI data. Alas, this doesn’t answer Lee et al.’s (2005) question about whether the gustatory experience of the wine or the cola was changed, because it attests only to the genuine enhancement of the subjects’ pleasure at the time of consumption, a pleasure that can derive from knowing what is being consumed rather from affecting the taste.

The second kind goes directly to performance measures. Although performance is a behavioral variable, if a given object leads to better performance when it is more expensive or more prestigiously branded, we know that it isn’t just expected to be better, or rated as better—it actually becomes better. Shiv et al. (2005) showed that discounting the price of a drink purporting to increase mental acuity reduces performance on solving word puzzles compared to drinking the drink at its regular price; Amar, Ariely, Bar-Hillel, Carmon & Ofir (2011) showed that participants “wearing sunglasses tagged Ray-Ban made fewer errors, yet read more quickly, than those wearing the identical pair of sunglasses when tagged Mango... Similarly, ear-muffs blocked noise more effectively, and chamomile tea improved mental focus more, when otherwise identical target products carried more reputable names” (p. 1). These data prove that products that are expected to be better sometimes actually become better through the expectation.

Relatedly, Lee, Linkenauger, Bakdash, Joy-Gaba and Proffitt (2011) showed that amateur golfers who believed they were using a professional golfer’s putter perceived the size of the golf hole to be larger, and sank more putts; Crum and Langer (2007) showed that informing hotel room attendants in a thorough and scientific manner that their work is good exercise reduced their weight, blood pressure, body fat, and other similar measures, compared to uninformed controls.

This evidence of “real” effects of expectations is very compelling, though some of the effects are harder to explain than others. When the dependent variables are physiological (e.g., blood pressure), what we know about medical placebos comes to bear. Regarding behaviors

that are under one’s control (e.g., golf putting; puzzle solving) the effect may be mediated by motivation (see, e.g., Irmak, Block & Fitzsimons, 2005). Other effects (e.g., Amar et al., 2011) are more mysterious.

### 6.2 The problem with subjective ratings

For stimuli like poetry, no “performance” can substitute for verbal ratings. However, one could use other measures, that are supposedly more “objective”, such as observing our readers’ brains as they were reading Infinite Poem—attributed or unattributed. We also could have measured physiological indicators of their emotional reactions, or tracked eye-movements, or measured reaction times. These could have confirmed (or not) “objectively” that the informed reader and the blind reader were in different cognitive states. But in the present context, they would not have been superior in helping us understand the nature of this difference beyond the simple expedient of asking for subjective ratings, as we did. Invasive and expensive techniques are not the only way to delve into the “black box”. The right kind of old-fashioned paper-and-pencil subjective ratings can still go a long way.

A possible artifact of rating scales that can be dismissed here is that the change in the ratings received by the attributed versus unattributed Infinite Poem is due to a change in scale (see, e.g., Frederick & Mochon, 2011). Such a change can occur if the unattributed poem is judged as a poem, whereas the attributed poem is judged as an Amichai-poem. Numbers are not comparable when scales are not comparable. After all, a small elephant is still much bigger than a large mouse (Stevens, 1958). Might the attributed poem have merited a 63 rating among Amichai’s poems, and a 53 rating among all poems? Commonsense argues against it. Amichai is a highly regarded poet (see the results of G6), which means that his poetry is regarded on average as better than the average poem. Rescaling would thus have led to an opposite result: Infinite Poem’s rating should have gone down, not up (where is Michael Jordan perceived as taller—compared to the population at large, or compared to other basketball players?).

### 6.3 Is priming a bias? Ethical considerations

It is interesting to ponder whether an effect such as the one we found in the present series of studies should be regarded as an undesirable bias. Recall that, while the effect of the name occurs out of consciousness and is not deliberate, it did not fall into the category of instances where the information imparted by the name serves to change the object being evaluated. All it did was pull the evaluations in the direction of the expectations set up by

the name. One might call this a halo effect, or a self-fulfilling expectation, a confirmation bias, etc. If knowledge extrinsic to an object helps in evaluating it more accurately, the arguments for informed judgment are quite different than if all it does is just to pull the judgments generally in the expected direction. It is not ethically objectionable if people enjoy some products or experiences more when their expectations are raised, because these products and experiences are often bought, among other reasons, for the enjoyment they can bring. If price or brand brings one pleasure—why not? It is also not problematic if people who expect a cartoon to be funny find it funnier than people without this prior expectation (Wilson et al, 1993)—what’s to deplore if people find a cartoon funny? However, in the context of, say, a competition for “Funny cartoon of the year”, it seems that blind judging is ethically better. Not all cartoonists enjoy the same reputation, and it is unfair if the identified winner of last year’s competition enjoys the kind of ineffable advantage that our study discovered in this year’s competition. Reputations clearly feed upon themselves, and can snowball on their own weight. But where fairness is a concern, some advantages should be blocked.

The question of whether judgments are better when performed blind or when they are informed is thus seen to depend not only on how, in each context, the information affects the judgments (sinisterly, as when it is abused; usefully, as when it clarifies ambiguities; recreationally, as when it enhances pleasure; manipulatively, as when it promotes sales; beneficially, as when it improves performance; etc.), but also on the uses to which the judgments will be put.

## References

- Allison, R. I. & Uhl, K. P. (1964). Influence of beer brand identification on taste perception. *Journal of Marketing Research*, 1, 36–39.
- Amar, M., Ariely, D., Bar-Hillel, M., Carmon, Z. & Ofir, C. (2011). Brand names act like marketing placebos. Discussion Paper 566, The Center for the Study of Rationality. [http://www.ratio.huji.ac.il/dp\\_files/dp566.pdf](http://www.ratio.huji.ac.il/dp_files/dp566.pdf).
- Bar-Hillel, M., Maharshak, A., Moshinsky, A., & Nofech, R. (2010). Does a rose by any other name smell as sweet? A cognitive perspective on poets and poetry. Discussion Paper 549, The Center for the Study of Rationality. [http://www.ratio.huji.ac.il/dp\\_files/dp549.pdf](http://www.ratio.huji.ac.il/dp_files/dp549.pdf).
- Blank, R. M. (1991). The effects of double-blind versus single-blind reviewing: Experimental evidence from The American Economic Review. *The American Economic Review*, 81, 1041–1067.
- Bohannon, J., Goldstein, R., & Herschkowitsch, A. (2009). Can people distinguish P ate from dog food? *Chance*, 23 (2), 43–46.
- Crum, A. J., & Langer, E. J. (2007). Mind-set matters: Exercise and the placebo effect. *Psychological Science*, 18, 165–171.
- de Araujo, I. E., Rolls, E. T., Velazco, M. I., Margot, C., & Cayeux, I. (2005). Cognitive modulation of olfactory processing. *Neuron*, 46, 671–679.
- Fischhoff, B. (1975). Hindsight  $\neq$  Foresight: The effect of outcome knowledge on judgment under uncertainty. *Journal of Experimental Psychology: Human Perception and Performance*, 1, 288–299.
- Fisher, R. J. (1993). Social desirability bias and the validity of indirect questioning. *Journal of Consumer Research*, 20, 303–315.
- Fritz, C., Curtin, J., Poitevineau, J., Morrel-Samuels, M., & Tao, F. (2012). Player preferences among new and old violins. *Proceedings of the National Academy of Sciences*, 109, 760–763.
- Frederick, S., & Mochon, D. (2012). A scale distortion theory of anchoring. *Journal of Experimental Psychology: General*, 141, 124–133.
- Goldin, C., & Rouse, C. (2000). Orchestrating impartiality: The impact of “blind” auditions on female musicians. *American Economic Review*, 90, 715–741.
- Goldstein, R. (2008). *The wine trials: 100 everyday wines under \$15 that beat \$50 to \$150 wines in brown-bag blind tastings*. Austin, TX: Fearless Critic Media.
- Goldstein, R., Almenberg, J., Dreber, A., Emerson, J. W., Herschkowitsch, A., & Katz, J. (2008). Do more expensive wines taste better? Evidence from a large sample of blind tastings. *Journal of Wine Economics*, 3 (1), 1–9.
- Heyward, M. (2003). *The Ern Malley Affair*. Faber and Faber.
- Irmak, C., Block, L. G., & Fitzsimons, G. J. (2005). The placebo effect in marketing: Sometimes you just have to want it to work. *Journal of Marketing Research*, 42, 406–409.
- Irving, C. (1969). *Fake! The Story of Elmyr De Hory, the Greatest Art Forger of Our Time?* New York: McGraw-Hill.
- Kahneman, D., & Riis, J. (2005). Living, and thinking about it: Two perspectives on life. In F. Huppert, N. Baylis and B. Keverne (Eds.) *The Science of Well-Being*, pp. 285–304. Oxford University Press.
- Lee, L., Frederick, S., & Ariely, D. (2006). Try it, you’ll like it: The influence of expectation, consumption and revelation on preferences for beer. *Psychological Science*, 17, 1054–1058.
- Lee, C., Linkenauger, S. A., Bakdash, B. Z., Joy-Gaba, J. A. & Proffitt, D. R. (2011). Putting like a pro: The role

- of positive contagion in golf performance and perception. *PLoS ONE* 6(10), e26016.
- Levin, I. P., & Gaeth, G. J. (1988). How consumers are affected by the framing of attribute information before and after consuming the product. *Journal of Consumer Research*, 15, 374–378.
- Lichtenstein, S., Fischhoff, B., & Phillips, L. D. (1982). Calibration of subjective probabilities: The state of the art up to 1980. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under Uncertainty: Heuristics and Biases*, pp. 306–334. Cambridge, UK: Cambridge University Press.
- Livingston, P. (2005). *Art and Intention: A Philosophical Study*, Oxford University Press.
- Lowenstein, G., & Frederick, S. (1997). Predicting reactions to environmental change. In M. Bazerman, D. Messick, A. Tenbrunsel, & K. Wade-Benzoni (Eds.), *Environment, Ethics, and Behavior* (pp. 52–72). San Francisco: New Lexington Press.
- Makens, J. C. (1965). Effect of brand preference upon consumers perceived taste of turkey meat. *Journal of Applied Psychology*, 49(4), 261–263.
- McClure, S. M., Li, J., Tomlin, D., Cypert, K. S., Montague, L. M., & Montague, P. R. (2004). Neural Correlates of Behavioral Preference for Culturally Familiar Drinks. *Neuron*, 44, 379–387.
- Morrot, G., Brochet, F., & Dubourdieu, D. (2001). The color of odors. *Brain and Language*. 79, 309–320
- Perry, J. (1997). Indexicals and demonstratives. In R. Hale & C. Wright (Eds.), *Companion to the philosophy of language*. Oxford: Blackwells Publishers.
- Plassman, H., O'Doherty, J., Shiv, B., & Rangel, A. (2008). Marketing actions can modulate neural representations of experienced pleasantness. *Proceedings of the National Academy of Sciences*, 105, 1050–1054.
- Schkade, D., & Kahneman, D. (1998). Does living in California make people more happy? A focusing illusion in judgments of life satisfaction. *Psychological Science*, 9(5), 340–346.
- Schwarz, N. (1996). *Cognition and communication: Judgmental biases, research methods and the logic of conversation*. Hillsdale, NJ: Erlbaum.
- Shiv, B., Carmon, Z., & Ariely, D. (2005). Placebo effects of marketing actions: Consumers may get what they pay for. *Journal of Marketing Research*. 42,383–393.
- Stevens, S. S. (1958). Adaptation-level vs. the relativity of judgment. *American Journal of Psychology*, 71, 633–646.
- Wilson, T. D., Lisle, D. J., Kraft, D., & Wetzell, C. G. (1989). Preferences as expectation-driven inferences: Effects of affective expectations on affective experience. *Journal of Personality and Social Psychology*, 56, 519–530

## **Appendix: Authentic poem by Emily Dickinson and fake counterpart**

Wild Nights — Wild Nights!  
Were I with thee  
Wild Nights should be  
Our luxury!

Futile — the winds —  
To a heart in port —  
Done with the compass —  
Done with the chart!

Rowing in Eden —  
Ah, the sea!  
Might I moor — Tonight —  
In thee!

Blue Dawns — Blue Dawns!  
For my babie  
Blue Dawns decree  
A fantasy!

Empty — the crib —  
No nursery —  
No bibs, no nappies  
No luxury!

Roaming the gardens  
Oh, the grass!  
Seeking a laddie  
Seeking a lass!