Learning and Individual Differences xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

Learning and Individual Differences

journal homepage: www.elsevier.com/locate/lindif

More than the eye of the beholder: The interplay of person, task, and situation factors in evaluative judgements of creativity

Damian P. Birney [a,*], Jens F. Beckmann [b], Yuan-Zhi Seah [a]

[a] University of Sydney, Australia
[b] Durham University, UK

ARTICLE INFO

Article history:
Received 19 November 2014
Received in revised form 10 June 2015
Accepted 14 July 2015
Available online xxxx

Keywords:
Creativity
Judgement
Person–Task–Situation
Evaluation accuracy

ABSTRACT

Judging creativity accurately is difficult. Individuals who are involved in product creation tend to overestimate the creativity of their work; individuals not involved lack understanding of the creative process that led to the product under scrutiny. We studied creativity judgements in a tripartite person–task–situation framework. Under high, medium, or no structure conditions and different orders of evaluation, participants ($N = 90$) rated the creativity and purchase appeal of products created by themselves and others. Accuracy was defined as differences from consensus evaluations of participants not involved in production ($N = 30$). Moderator analyses suggest that externally set structure of the evaluation process (e.g., using a set of criteria) facilitates the quality of creativity judgement. In unstructured conditions, evaluating one's own product before evaluating a peer's leads to low accuracy, but higher levels of conscientiousness seem to mitigate potentially deleterious effects of lack of structure. Higher levels of openness facilitated accurate creativity judgements of peer-produced products, but not self-produced products. A person–task–situation approach is needed to fully unpack the complexity of processes underlying accurate evaluation of creativity.

1. Introduction

Creativity, when viewed from a product approach, incorporates the conceptual dimensions of novelty and originality and focuses on outcomes that are both useful and appropriate (Barron, 1988; Bleakley, 2004; Nickerson, 1999; Ruscio, Whitney, & Amabile, 1998; Torrance, 1988). Innovation, broadly speaking, is the successful implementation of creativity (Hirst, van Knippenberg, & Zhou, 2009; Hülsheger, Anderson & Salgado, 2009; Klein & Knight, 2005). There has been considerable work investigating the characteristics that make for a creative product and its successful implementation. However, considerably less attention has been given to the processes and structures that people employ in the evaluation of creativity. Evaluation is crucial to both creativity (Mumford, 1999) and innovation (Klein & Knight, 2005), in that it dictates which products to develop further and which to discard. History is littered with artistic and literary works, technological inventions, and scientific discoveries that were initially ignored or even disregarded because of poor or inaccurate evaluations (Elsbach & Kramer, 2003). The current paper contributes to research in the area by investigating factors that underlie accurate evaluation.

Evaluative judgement, as for the study of most human behaviour in psychological research, takes place in the tripartite context of the person, the task, and the situation. The *Person* dimension comprises all that can be subsumed under person-related psychological variables, such as attitudes, skills, abilities, and knowledge. A *Task* is defined as a specified requirement of behaviour (e.g., to solve a problem, to acquire knowledge, or make decisions). Behaviour in this regard is not limited to observable physical acts, it also includes cognition linked to processing information (Ferguson, 1956; Hackman, 1969; McGrath, 1984) — or in our case, evaluation. The *Situation* encompasses the circumstances or the situational context in which the task is to be performed. A conceptual demarcation between task and situation seems challenging, mainly because experientially every task is linked to a particular situation (i.e., we cannot describe a task without any circumstantial reference). However, because the same task can be presented in different ways and contexts, tasks and situations need to be treated as conceptually independent when studying behaviour (Beckmann, 2010; see also the notion of "task environment" described by Newell & Simon, 1972, p. 55).

In the present study we focus on whether and how *person* and *contextual* factors interactively contribute to creativity judgements. Person factors considered include individual differences in *divergent thinking* skills and various dimensions of *personality*. Situational factors include the *level of involvement* in the actual creation of the product that is to be evaluated. The imposed *structure of the judgement* process and the *order* of judging (whether one evaluates one's own product first or

* Corresponding author at: School of Psychology, University of Sydney, Sydney, New South Wales 2006, Australia.
E-mail address: damian.birney@sydney.edu.au (D.P. Birney).

http://dx.doi.org/10.1016/j.lindif.2015.07.007
1041-6080/

not) are contextual variables also considered a part of the judging situation. In the following section we describe our investigative framework.

1.1. Creativity criteria and evaluation accuracy

Contemporary research has for the most part adopted a product approach to creativity in lieu of focusing directly on the person, the process, or the environment. This has been because, first, creative people are typically judged to be creative by what they produce (Kaufman, Christopher & Kaufman, 2008). Second, product characteristics explain the most variance in evaluations of creativity, far more than person or process dimensions (Demirkan & Hasirci, 2009); and most importantly, third, because a product approach has been seen for some time to provide access to what are considered to be the main contributors of creativity: environmental factors, processes used, and attributes of the individual producing the creative product (Amabile, 1988).

Derived from this view, the most common method to ascertain product creativity is through some form of expert judgement (e.g., Dailey & Mumford, 2006). This approach has been applied to the evaluation of real-world creative products ranging from TV shows of the "got talent" variety to Nobel Prizes. Various guidelines have been developed on how to use consensus judgement systematically and rigorously (e.g., Amabile, 1982; Baer, Kaufman, & Gentile, 2004). The consensus-based expert judgement regarding the creativity of a given product serves as the reference to determine the 'accuracy' of a specific evaluation. That is, accuracy is defined by the degree of correspondence or alignment between the judgement provided by an individual rater and what a consensus group has agreed upon. In the current research, the consensus group is defined as typical, potential consumers of the created product. We include willingness to purchase the product (*purchase appeal*) as an additional indicator of the *utility* criterion of creativity.

Our study of how individuals deal with the *task* of evaluating a creative product focuses on three *situation* factors, structure, involvement, and order, and two categories of *person* factors, ability and personality.

1.1.1. Structure of evaluation

Structure, in the form of a prescribed set of evaluation criteria, is a situational factor that is expected to impact upon how the evaluation task is performed and consequentially, the quality (e.g., accuracy) of the judgement (Gary, Birney & Wood, submitted for publication). The most unstructured approach is simply to ask people to provide a summary rating of how creative they believe a product is. This is often the basis of consensus scores where the unstructured, 'naturalistic', or intuitive evaluations of experts are obtained (without using a scoring rubric, Kaufman, Baer, Cole & Sexton, 2008) and then aggregated (Amabile, 1982).

Criterion-bound methods are often developed as alternatives to naturalistic ratings and commonly used to structure evaluation (e.g., O'Quin & Besemer, 2006). Besemer (2000) argues that while natural, intuitive judgements are useful, they can result in snap judgements and less considered processing. Structured evaluation methods enforce a more conscious and deliberative evaluation process (Wood, Beckmann, & Birney, 2009; Beckmann, Beckmann, Birney, & Wood, 2015). Gary et al. (submitted for publication) have argued for the efficacy of using similar approaches for structuring analogical reasoning. Positioned between a naturalistic, intuitive evaluation process and a structured approach is the *implicit criteria evaluation method*, which relies on individual evaluators explicating their own implicit judgement standards and using these as criteria for summary judgements (Weinstein, 1980).

The general expectation is that structured evaluation methods lead to more effective and generally more accurate judgements (Beckmann & Schumacher, 2004; Meehl, 1954). We therefore hypothesise that all else equal, structured evaluation methods will result in more accurate evaluations for creativity and purchase appeal of a product.

1.1.2. Involvement

Another situational variable that is in the focus of this study is the level of *involvement* in the creation of the product to be judged. *Product-involved evaluators* are likely to be more knowledgeable about the product and make evaluations cognizant of the idiosyncratic features of the creation process. Involvement in product development may also create higher levels of vested interests in favourable evaluations than would be expected of uninvolved, more impartial judges. Product-involved evaluators can be *self-evaluators* who are directly involved or chiefly responsible for the creative output under scrutiny, or *peer-evaluators* who are either only marginally involved in the creation of the product or have been involved in the creation of a similar product but not the one under scrutiny.

Domain experts seem to use different implicit criteria than laypeople for evaluating creativity (Runco & Bahleda, 1986; Sternberg, 1985). Evaluation criteria not only differ as a function of expertise, they also tend to vary intra-individually (Charles & Runco, 2001; Runco & Chand, 1994). In our study we specifically compare the evaluations made by product creators (self) with those provided by others who were engaged in the same task but not having produced the presented evaluation target (peers). Within social comparison research, the consistently documented above-average effect, where people rate themselves to be above average on an assortment of traits and attributes, is hinged on a difference in perspective between people evaluating themselves versus evaluating others (e.g., Klar & Giladi, 1999; Chambers & Windschitl, 2004). One explanation proposed for differences in creativity judgements is that unequal involvement and familiarity with the creation process and the product leads to biases and/or differences in cognitive processing (Chambers & Windschitl, 2004). Runco and Smith (1992), on the other hand, found that people were more accurate at evaluating the originality (in terms of statistical rarity) of their own responses in a divergent thinking test, than of responses provided by others.

Product-uninvolved evaluators can be sub-divided into *judges* and *consumers*. Judges are 'Appropriate Observers' (Amabile, 1982) or 'Domain Gatekeepers' (Csikszentmihalyi, 1990) with established expertise to (arguably rightly) determine whether the output is deemed to be creative or not. Consumers on the other hand are those individuals directly affected by, or are the intended target group for the product in question. They are not necessarily experts and are likely to lack the breadth of experiences that judges possess. We hypothesise that the level of involvement in creation has an impact on the evaluation of a creative product. Products will be rated as being more *creative* and as having higher *purchase appeal* by product-involved raters than by product-uninvolved raters. Also, product-involved individuals will rate their products consistently more favourably than they will rate the products of others who worked on a similar task.

1.1.3. Order of evaluation

The third situational variable included in this study refers to the effects of the chronological context (i.e., *order*) in which judgements are made. Specifically, we are interested in the potential effects of evaluating one's own product after or before evaluating a peer's product. Two person-related constructs relevant to how individuals deal with the situational factor of evaluation order are (1) *egocentrism* and (2) *trait underestimation of others*. Egocentrism refers to instances when thoughts about the self loom larger than thoughts about others, which in turn results in a disproportionate weighing of self-referent information (Chambers & Windschitl, 2004). In a similar vein, Klar and Giladi (1999) argue that *trait underestimation of others* reflects a lack of awareness about the level of the trait or ability in others. Both lines of social comparison research converge to suggest that someone who self-evaluates first would tend to focus heavily on their own work and rate themselves with limited awareness of the abilities or creativity of the works of others. Conversely, evaluation of a peer's product first would lead to a clearer awareness of the creativity of products produced by others, self-

Please cite this article as: Birney, D.P., et al., More than the eye of the beholder: The interplay of person, task, and situation factors in evaluative judgements of creativity, *Learning and Individual Differences* (2015), http://dx.doi.org/10.1016/j.lindif.2015.07.007