HAL
open science

A variational model for context-driven effects in perception and cognition

Giorgio Gronchi, Edoardo Provenzi

To cite this version:

Giorgio Gronchi, Edoardo Provenzi. A variational model for context-driven effects in perception and cognition. Journal of Mathematical Psychology, 2017, 77, pp.124 - 141. 10.1016/j.jmp.2016.10.005 . hal-01386167

HAL Id: hal-01386167

https://hal.science/hal-01386167v1

Submitted on 23 Oct 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A variational model for context-driven effects in perception and cognition

Giorgio Gronchi[a], Edoardo Provenzi[b]

[a]*Department of Neurosciences, Drug Research, and Child Health (NEUROFARBA)*
Università degli Studi di Firenze
via di San Salvi n. 12, Complesso di San Salvi, Padiglione 26
50135 Firenze (FI), Italy
[b]*Laboratoire MAP5 (UMR CNRS 8145)*
Université Paris Descartes
45 rue des Saints-Pères, 75006, Paris, France

Abstract

Starting with a computational analysis of brightness matching, we develop a novel variational framework able to model perceptual context-driven effects that may be extended to non-physical judgments as well. The most important feature of the variational framework is the description of these phenomena as a suitable balance between contrast and dispersion. The optimal balance is defined through the simultaneous minimization of functionals characterized by two terms in opposition to each other. When the minimum is reached, the equilibrium between contrast and dispersion is attained. To show the flexibility of the proposed framework, we discuss several examples of such functionals in the field of color perception and cognition which show adherence between theoretical predictions and empirical results. With regard to social cognition theories, the simultaneous occurrence of contrast and dispersion conflicts with sequential models, thus supporting the idea of a concurrent presence of both effects in each judgment. The variational framework can serve as a view from above on perceptual and cognitive phenomena that may help in deriving new constraints for disambiguating alternative theories.

Keywords:
Context effects, contrast, assimilation, variational models, psychophysics.

1. Introduction and state of the art

Context-driven effects are one of the most frequent observations in psychology. We can define a context-driven effect as an over- or under-estimation of a stimulus embedded in a given context compared to the same evaluation task performed in isolated conditions. Within the field of visual perception, a typical example is the contrast effect observed in the brightness matching

Email address: `edoardo.provenzi@parisdescartes.fr` (Edoardo Provenzi)

experiment performed by Wallach (1948). Using Rudd and Zemach's reinterpretation of Wallach's experiment (Rudd & Zemach, 2004), we developed a general variational model where context-driven effects can be described. Formulating a problem in terms of variational principles is a common strategy in practically every scientific discipline. The main purpose is to obtain a broader view of the problem, thereby enabling us to derive a higher level explanation of the phenomenon and to detect its underlying functional constraints. Furthermore, by using this approach it is possible to highlight new constraints that may help to disambiguate alternative interpretations.

When the brightness matching experiment is seen as a calculus of variations problem, judgments of a stimulus embedded in a context can be interpreted as the result of the balancing of two opposing processes. The first (that we will call with the standard term *contrast*) tries to emphasize differences in the final percept, whereas the second (that we will call *dispersion*) tries to emphasize similarities. We attempt to present evidence that a variety of contextual effects observed in non-physical judgments (such as those of social cognition) can be analyzed using the same variational framework.

1.1. Contextual effects for physical judgments

In the domain of perception, as early as in the seventeenth century the philosopher John Locke (1690) had described the contrast effect by noticing that a hand's contact with tepid water can produce either the sensation of cold or hot if the hand had been previously placed in hot or cold water, respectively. The contrast effect has been considered since the earliest days of psychophysics (Chevreul, 1855; Wundt, 1896) and extensively investigated for judgments in the context of physical dimensions, such as the loudness of a tone (Melamed, 1971), the brightness of light (Wallach, 1948), weight (Heintz, 1950; Sherif, Taub, & Hovland, 1958), length of lines (Krantz & Campbell, 1961), and so on. In parallel, since von Bezold (1876) described a phenomenon "*in which a colored surface appears lighter when overlaid by thin white lines or small white dots and appears darker if the lines or dots are black,*" assimilation effects (after the name given to them by Evans, 1948) were also studied (Blakeslee & McCourt, 2004; De Weert & Spillman, 1995; Festinger, Coren, & Rivers, 1970; Helson, 1963; Kingdom & Moulden, 1991), especially in visual phenomena such as the white effect (White, 1979), Bressan's dungeon illusion (Bressan, 2001) or the *cube* illusion of Agostini and Galmonte (2002). Within the domain of brightness perception, assimilation is considered the opposite of contrast. However, given the variety of models, theoretical approaches, and empirical results, such a notion is rather controversial and it has been employed as a convenient catch-all in which to place anti-contrast effects (Gilchrist, 2006). In particular, when referring to brightness judgments, the term 'assimilation' has a special and different meaning from the concept described in this paper. So, as we wrote before, we will employ the term dispersion instead to indicate a positive correlation between the judgment and the context. The use of this generic term will also be helpful because we are going to refer the corresponding context effects in the social cognition and cognitive psychology fields where the word assimilation may have different meanings.

The debate overt the variables (and the underlying mechanisms) that determine different context effects is still open for discussion. Indeed, there is no consensus on how surface lightness is processed by the brain (Gilchrist, 2015) and thus under which conditions different context effects are observed. Within this debate, Agostini and Galmonte (2002) proposed that perceptual belongingness may determine the kind of context effect. According to it, gestalt laws (proximity, similarity, good continuation, common fate, closure, and pragnänz) (Koffka, 1935; Wertheimer, 1923) can explain the tendency of the visual system to aggregate discrete stimuli within larger wholes and thus determine if a stimulus phenomenologically belongs to a larger object or not. The basic idea is that if two elements belong to different perceptual groups, their colors are contrasted with the color of the group to which they belong (Agostini & Profitt, 1993; Agostini & Galmonte, 2000). On the contrary, when an element is intentionally organized into one or another of two groups, its color is assimilated to the color of the group to which it belongs (so, in our terminology, a dispersion effect will be observed). For the sake of simplicity, in this paper we will assume Agostini and Galmonte's (2002) perspective because it is compatible with the proposed formal framework and it addresses an important open question in the literature. However, our analysis does not depend on this assumption and it can be compatible with other theoretical proposals.

1.2. Context effects for non-physical judgments

The results obtained for basic perceptual judgments have suggested investigating the influence of the context for non-physical judgments as well. Within social cognition, the word contrast is employed to define the case of a judgment negatively correlated with the contextual information, whereas the term assimilation refers to a positive correlation between the judgment and the contextual information. Those effects have been observed for non-physical judgments related to moral evaluations (Parducci, 1968; Pepitone & DiNubile, 1976), pleasantness of music (Parker, Bascom, Rabinovitz, & Zellner, 2008), friendliness of a person (Stapel, Koomen, & van der Pligt, 1997), attractiveness (Kenrick & Gutierres, 1980), prices of objects (Matthews & Stewart, 2009b), and a wide variety of social judgments and evaluations (Biernat, 2005; Moskowitz, 2005). For example, when investigating the contrast effect in moral judgment, Parducci (1968) asked respondents to rate the seriousness of a number of acts, such as *poisoning a neighbor's dog*, alongside trivial acts, such as *keeping a dime you find in a telephone booth*, and very serious bad acts, such as *murdering your mother without justification or provocation*. He found that in the first case, the sentence *poisoning a neighbor's dog* was judged as more serious when compared to the second case. Different terms have been employed to indicate a judgment biased towards the context. Within cognitive psychology, Tversky and Kahneman (1974) used the word anchoring to indicate the bias of a numeric judgment towards a previously considered standard. In line with the observations made previously about physical judgments, given such terminological ambiguities, we employ the word dispersion to mean such kind of bias.

Several studies have identified many factors that can induce dispersion or contrast in judgment related to person perception (Higgins & Lurie, 1983) and self-evaluation (Festinger, 1954). For example, broad contextual categories (such as traits) are likely to produce dispersion whereas

in the case of a context represented by narrow categories (such as exemplars), contrast effect will be observed (Stapel, Koomen, & van der Pligt, 1996). Other factors include processing goals (memorization vs. impression formation where the first induces assimilation and the second contrast, Moskowitz & Roman, 1992), distinctness of the context (in the case of high distinctness that is more likely to observe contrast, see Wedell, Parducci, & Geiselman, 1987), temporal distance between events (distant contextual events are more likely to induce contrast, see Strack, Schwarz, & Gschneidinger, 1985), and many others (Biernat, 2005).

Given the factors that can induce cognition-related context-driven effects, several theoretical models have been developed to furnish parsimonious and effective predictions about how and when dispersion or contrast occurs in a given situation. Among them, we can cite the set-reset model (Martin & Achee, 1992), the inclusion-exclusion model (Schwarz & Bless, 1992), the flexible correction model (Petty & Wegener, 1993), the interpretation-comparison model (Stapel & Koomen, 1998) and the selective accessibility model (Mussweiler, 2003). Those models differ in terms of the assumed degree of effort involved in the effects (automatic or controlled), if the two effects are simultaneous or sequential, the specific variables involved in the processes, and the assumed default process (either contrast or dispersion). However, the majority of them agree with the notion that the factors that make the context less distinct from the stimulus (in other terms, factors suggesting an inclusion of the stimulus in the context) induce a dispersion effect, whereas factors that make the context distinct from the stimulus (so, suggesting an exclusion from the context) induce a contrast effect. Such interpretation is coherent with Agostini and Galmonte's (2002)'s perspective about the factors determining context effect in physical judgments.

1.3. Rationale for this article

In this paper, we develop a general, formal framework for context-driven effects in both perception and cognition. In the psychophysical literature about contrast, it is possible to find attempts that provide mathematical formalizations of this effect. The first, and still one of the most famous, is provided by Fechner's formalization (Fechner, 1860) of Weber's findings about the differential threshold of sensation by sense organs (Weber, 1846). Even more remarkably, it is possible to find attempts to build mathematical models that try not only to formalize measurements, but also to predict new phenomena. We will present a pertinent example of such a model in section 2 related to visual induction (i.e., perception of light stimuli in a non-isolated context). Recently, this model has been embedded in a variational setting (Provenzi, 2013).

The use of a variational principle is not yet widespread in mathematical psychology, even though a few examples can be found (Ehm & Wackermann, 2012; Noventa & Vidotto, 2012). In other disciplines (e.g., physics and signal processing), variational principles have been extensively employed. One of the main reasons for this success is that variational calculus allows the reinterpretation of a model in terms of the hidden basic mechanisms underlying the model itself. Crucially, these basic mechanisms are also shared with other models and this allows for a demonstration of similarities and differences that are often very difficult to discover without a variational approach.

We stress that the main purpose of variational calculus is not to provide new predictions compared to already existing models, but to offer a more general and profound view about the phenomena described by these models to possibly build bridges between different theories. Once the variational interpretation of a model has been established, it is possible to determine general constraints about the phenomena and it is also possible to slightly modify the analytical expression of the variational model in order to generate new equations to be tested.

To illustrate our approach, we are going to show that the variational reinterpretation of Rudd-Zemach's model can also be applied to the comprehension of mechanisms underlying non-isolated judgments of non-physical stimuli. As stressed above, the core contribution of this variational framework is the possibility of describing context-driven judgments as a suitable balance between two opposing mechanisms: contrast and dispersion. On the one side, the contrast mechanism tends to maximize as much as possible the difference between the stimulus and its surround, while on the other side, simultaneously, the dispersion mechanism tends to absorb this difference, integrating the stimulus into its surround. Finally, we will strengthen our hypothesis with a quantitative discussion of previous studies about context effects for non-physical judgments.

The paper is organized as follows: In section 2 we introduce Rudd-Zemach's analysis of a context-driven perceptual match experiment; after a brief summary of variational principles in section 3, we reformulate Rudd-Zemach's model in section 4 and interpret brightness matching as an optimal balance between dispersion control and contrast enhancement. This concept is further generalized in section 5, in which a fully general variational framework for context-driven perceptual and cognitive phenomena is provided and discussed with examples. The simple, yet already significant, variational interpretation of linear equations in the linear and logarithmic domain is discussed in great detail. Section 6 is devoted to a thorough discussion of the proposed model and several case studies. The paper ends with conclusions and an appendix showing the proof of the variational result of section 4.

2. A significant example of modeling of context-driven physical judgments: Achromatic induction

In this section, we will discuss in detail a successful model for achromatic induction. In the next section, we will provide a variational interpretation of this model and this will serve as a concrete example for the formulation, in section 5, of a more general variational model suitable for non-physical judgments as well.

Human perception of a color patch is not determined by its reflectance properties alone, but also by those of the surrounding patches. This phenomenon is called *chromatic induction*, to stress the fact that color perception is induced (and altered) by the surround. This is also true for the so-called achromatic colors (i.e., perceived shades of gray), in which case one talks about *achromatic induction*. Following a common nomenclature, we will call *lightness* the *perceived reflectance* of a non self-luminous patch, while *brightness* will refer to the *perceived luminance*

emitted by a source of light. Thus, the terms lightness and brightness already incorporate the potential effect of induction in their definition. The most elementary example of induction is the *simultaneous contrast* phenomenon, depicted in Fig.1: The two inner squares have exactly the same luminance, however we perceive them very differently because we are strongly influenced by their distinct surrounds.

Figure 1: Achromatic simultaneous contrast: The inner gray squares have the same luminance but the left one, surrounded by a dark gray patch, looks lighter compared to the right one that is surrounded by a light gray patch.

Induction can be measured through psychophysical experiments. The first quantitative measure of achromatic induction was performed by Wallach (1948). In his classical experiment (Fig.2), Wallach considered two disks, D_T and D_M for *Target* and *Match*, surrounded by two rings R_T and R_M, embedded in a uniform background B. Let us denote with L_{D_M}, L_{R_M}, L_{D_T}, L_{R_T}, and L_B the luminance values of D_M, R_M, D_T, R_T, and B, respectively. He showed this configuration to a set of observers adapted to the light conditions of a dimly illuminated room, keeping L_{D_T} and L_{R_M} fixed, using L_{R_T} as an independent variable that he could fix in every experiment, and L_{D_M} as a dependent variable that the observers could adjust in order to achieve a perceptual match between the two disks T and M. The stimuli presented to the observers did not have chromatic components.

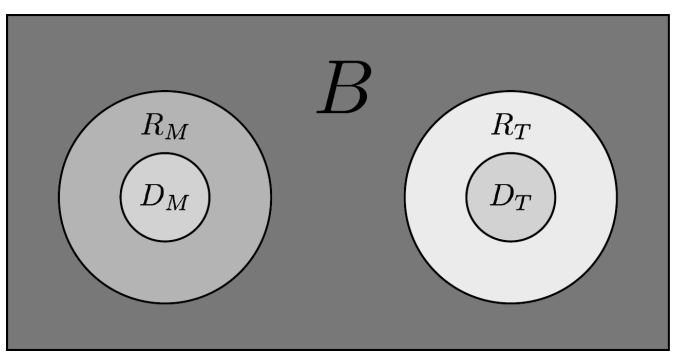

Figure 2: Wallach's classical experiment. Over a uniform background B, there are two inner disks, D_T and D_M (T and D for *Target* and *Match*, respectively), surrounded by two external rings R_T and R_M.

If the luminance of the surrounding rings failed to influence the perception of the achromatic color of the disks, then the match between the two disks would simply be the *photometric* one, i.e., $L_{D_M} = L_{D_T}$; instead, Wallach found that a fairly good match among the achromatic color of the two disks was obtained when the ratios between the disk and the ring luminances were

identical on the two sides of the display, i.e.,

$$\frac{L_{D_M}}{L_{R_M}} = \frac{L_{D_T}}{L_{R_T}}, \tag{1}$$

a formula called *Wallach's Ratio rule*. By taking the logarithms at both sides and solving for L_{D_M} we find:

$$\log L_{D_M} = \log L_{D_T} + \log L_{R_M} - \log L_{R_T}, \tag{2}$$

thus, according to Wallach's Ratio rule, the plot of the perceptual match in the plane of coordinates $(x, y) = (\log L_{R_T}, \log L_{D_M})$ should be a straight line with slope -1, against the slope 0 that a photometric match would measure. More recent measures using the classical Wallach's experiment have shown that this slope is actually between -1 and 0, as can be seen in Fig.3 (adapted from Rudd & Zemach, 2004).

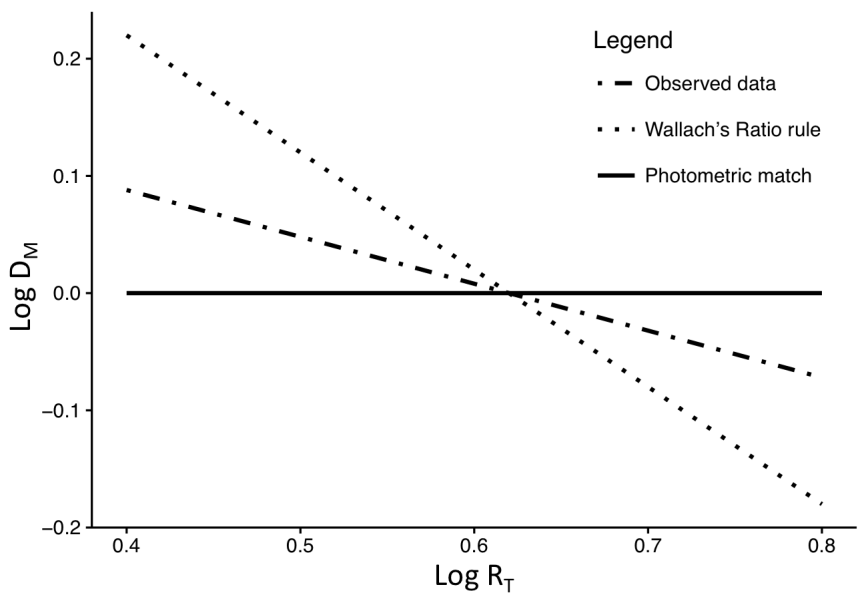

Figure 3: Quantitative measures of Wallach's achromatic color induction experiment for four observers performed in Rudd & Zemach (2004). The best-fit regression line slopes and associated 95% confidence limits observed by Rudd & Zemach (2004) for the four subjects of the experiment are the following: -0.639±0.033, -0.791±0.034, -0.723±0.047, and -0.657±0.042.

To account for these new psychophysical data, Rudd and Zemach (2004) have proposed a more sophisticated model than Wallach's. They repeated Wallach's experiment adding a non-black background B. As in Wallach's experiment, L_{D_T} and L_{R_M} are fixed and the observer's task is to adjust L_{D_M} to achieve an achromatic color match to the test disk as a function of L_{R_T}. L_{R_T} is varied from trial to trial by sampling from a set of six luminance values spaced equally in RGB units from 2.54 to 6.31 cd/m^2 (notice that Rudd and Zemach used the base 10 for their logarithmic values, so that the logarithmic range goes from 0.405 to 0.800 cd/m^2).

Rudd and Zemach also pointed out some similarities between their model of achromatic induction and the Retinex theory of color perception of Land and McCann (1971) without the

so-called threshold and reset mechanisms (see Provenzi, De Carli, Rizzi, & Marini, 2005, for more details). Rudd and Zemach's model can be described as follows. Let L_i and L_j be the luminance of two points i and j in an image, the ratio $\frac{L_i}{L_j}$ can be decomposed as a sequential multiplication of the local luminance ratios at borders encountered along a path connecting j and i, for istance:

$$\frac{L_i}{L_j} = \prod_{k=i}^{j-1} \frac{L_k}{L_{k+1}}, \tag{3}$$

by taking the logarithm at both sides we get

$$\log \frac{L_i}{L_j} = \sum_{k=i}^{j-1} \log \frac{L_k}{L_{k+1}}. \tag{4}$$

Rudd and Zemach introduced *induction strength weights* w_k in order to take into account the *locality of vision*, i.e., the fact that patches that lie in a nearby surround have a stronger influence on the induced perception than those that are far away. This point is somewhat delicate and it will be discussed in the next section. Notice that in the configuration shown in Fig.2, we have $i = D_M$, $i+1 = j-1 = R_M$, $j = B$ on the left part of the visual field, and $i = D_T$, $i+1 = j-1 = R_T$, $j = B$ on the right part. The *logarithmic brightness of* i, which we denote with $\log \Phi(i)$, that can be inferred by Rudd and Zemach's model, is the following:

$$\log \Phi(i) \equiv \sum_{k=i}^{j-1} w_{k-i+1} \log \frac{L_k}{L_{k+1}} + \mu, \tag{5}$$

where $\mu \in \mathbb{R}$ is an arbitrary constant that will be eliminated by the matching procedure and that we introduced to underline the fact that brightness perception is relative to a context and not absolute. It can be seen that, if the luminances L_k and L_{k+1} are equal, then their ratio does not give any contribution to $\Phi(i)$. A meaningful contribution to $\Phi(i)$ is given only by the luminances of points lying at the border of an edge. So, $\Phi(i)$ represents the summed influence of all the edges present within the spatial surround of the target point, suitably weighted. The weights index is $k - i + 1$, which means that small values of the index refer to patches close to i, and vice versa. With this convention, and invoking the fact that induction strength decreases with the distance, as proven by Wallach (1963), we have that $w_1 > w_2 > \ldots$, i.e., $\frac{w_2}{w_1} < 1$, and so on.

Rudd and Zemach called their model of achromatic induction 'Weighted Log Luminance Ratio', or WLLR for short. WLLR predicts that the brightness match between L_{D_M} and L_{D_T} is attained when $\Phi(D_M) = \Phi(D_T)$, i.e.,

$$w_1 \log \frac{L_{D_M}^{\text{Match}}}{L_{R_M}} + w_2 \log \frac{L_{R_M}}{L_B} + \mu = w_1 \log \frac{L_{D_T}}{L_{R_T}} + w_2 \log \frac{L_{R_T}}{L_B} + \mu, \tag{6}$$

solving this equation w.r.t. $\log L_{D_M}$ we have

$$\log L_{D_M}^{\text{Match}} = \log L_{D_T} + \left(1 - \frac{w_2}{w_1}\right) \log L_{R_M} - \left(1 - \frac{w_2}{w_1}\right) \log L_{R_T}, \tag{7}$$

$L_{D_M}^{\text{Match}}$ is the luminance value of D_M selected by the observer to match L_{D_T}. If we set $u = \log L_{D_M}^{\text{Match}}$, $\alpha = \log L_{D_T} + \left(1 - \frac{w_2}{w_1}\right) \log L_{R_M}$, $\beta = -\left(1 - \frac{w_2}{w_1}\right)$ and $v = \log L_{R_T}$, then the WLLR model predicts the following *linear behavior in the logarithmic domain*:

$$u = \alpha + \beta v, \tag{8}$$

with a slope $\beta = -\left(1 - \frac{w_2}{w_1}\right) \in (-1, 0)$, which is coherent with Rudd and Zemach's empirical observations. In fact, the estimations of the ratio $\frac{w_2}{w_1}$ from their interpolated data for the four observers are: 0.361, 0.209, 0.277, and 0.343. A key assumption of *edge integration models*, like WLLR, is that the total achromatic color induction produced by a complex surround is *the sum of the individual induction effects* produced by the luminance borders comprising that surround. Rudd and Zemach performed experiments to directly test this assumption by predicting and then measuring the magnitude of the total induction effect produced by combining three circular edges located at different distances from the test disk. This was done after first measuring the magnitudes of the induction effects produced by the individual edges. Results were in accordance with the predictions of the model.

In section 4, we will re-interpret the WLLR model in terms of variational principles; this will give us the possibility of introducing in a clearer way the more general variational framework of section 5. For the sake of clarity, and to introduce the usual nomenclature and notation of variational calculus, we summarize the basic information about variational principles in the next section.

3. Overview of variational principles

Calculus of variations is a generalization of ordinary calculus in $\mathbb{R}^n$. In the latter, we deal with functions $f : D \subseteq \mathbb{R}^n \to \mathbb{R}^m$, n, m integers ≥ 1, while in variational calculus we work with functions acting on functional spaces. More precisely, a functional space is a vector space whose elements are functions having some specified features. To provide a concrete example, let us consider two very well-known and useful functional spaces:

- $C^n(D)$, $D \subseteq \mathbb{R}^n$, D open, is the space of functions $f : D \to \mathbb{R}$ which are n-times differentiable, with continuous derivatives on the whole D, the case $n = 0$ corresponds simply to continuous functions on D;
- $L^2(\mathbb{R})$ is the space of square-integrable functions on $\mathbb{R}$, i.e., $f : \mathbb{R} \to \mathbb{R}$ such that $\int_{\mathbb{R}} f(x)^2 \, dx < +\infty$. These functions are also said to be finite-energy functions.

Given an abstract functional space $\mathcal{F}$ over the field $\mathbb{K}$ (in general $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$), the linear operations in $\mathcal{F}$ are defined point-wise, i.e., given $f, g \in \mathcal{F}$ and $\alpha, \beta \in \mathbb{K}$, the function $h \in \mathcal{F}$ defined by the linear combination $h = \alpha f + \beta g \in \mathcal{F}$ acts as follows on the arguments x of f and g: $h(x) = \alpha f(x) + \beta g(x)$. A functional φ acting on the abstract functional space $\mathcal{F}$ is a linear form

over $\mathcal{F}$, i.e., a linear function from $\mathcal{F}$ to the field $\mathbb{K}$:

$$\begin{array}{rccc} \varphi: & \mathcal{F} & \longrightarrow & \mathbb{K} \\ & f & \longmapsto & \varphi(f). \end{array}$$

Let us also recall that, in ordinary calculus, a great deal of effort is dedicated to finding the extrema of functions. To fix ideas, let us consider a function $f \in C^n(D)$, $\vec{x} \in D$, then:

- we call $\underline{\vec{x}} = \underset{D}{\operatorname{argmin}}\ f$, if $f\left(\underline{\vec{x}}\right) = \min\limits_{\vec{x} \in D} f(\vec{x})$;
- we call $\overline{\vec{x}} = \underset{D}{\operatorname{argmax}}\ f$, if $f\left(\overline{\vec{x}}\right) = \max\limits_{\vec{x} \in D} f(\vec{x})$.

We recall that, given a function $f : D \subseteq \mathbb{R}^n \to \mathbb{R}$, $f \in C^n(D)$ and any unit vector $\vec{v} \in D$, the *directional derivative* of f along $\vec{v}$, calculated in $\vec{x}$, is defined by:

$$D_{\vec{v}} f(\vec{x}) = \lim_{\varepsilon \to 0} \frac{f(\vec{x} + \varepsilon \vec{v}) - f(\vec{x})}{\varepsilon}, \tag{9}$$

the partial derivatives $\partial_i f(x)$, $i = 1, \ldots, n$, of f are simply the n directional derivatives computed by choosing $\vec{v} = \vec{e}_i$, the i-th unit vectors of the canonical basis of $\mathbb{R}^n$. Finally, the gradient $\vec{\nabla} f(x)$ is the n-dimensional vector whose components are the partial derivatives of f in x.

By virtue of Fermat's interior extrema theorem, the gradient of f (and its directional derivatives in every direction) must be null when computed in the argmin or argmax of f. This necessary condition also becomes sufficient when D and the function f are convex. The computation, either analytical or approximated, of the extrema of a function $f \in C^n(D)$ belongs to a field called *optimization in* $\mathbb{R}^n$. Contrary to ordinary calculus, in variational calculus *the argmin and argmax of a functional are functions*, more precisely, for an arbitrary functional $E : \mathcal{F} \to \mathbb{K}$:

- we call $\underline{f} = \underset{\mathcal{F}}{\operatorname{argmin}}\ E$, if $E\left(\underline{f}\right) = \min\limits_{f \in \mathcal{F}} E(f)$;
- we call $\overline{f} = \underset{\mathcal{F}}{\operatorname{argmax}}\ E$, if $E\left(\overline{f}\right) = \max\limits_{f \in \mathcal{F}} E(f)$.

The possibility of progressing from an *extrema of functions*, represented by points of $\mathbb{R}^n$, to *extrema of functionals*, represented by functions allows us to examine, in variational calculus, much more general problems than in ordinary calculus. Of course, this comes at the expense of a greater mathematical difficulty. The computation (in some cases analytical but, most of the time, approximated) of the extrema of a functional E is called *variational* (or *functional*) *optimization* and it has been the subject of research by many mathematicians, physicists, and engineers in the past two centuries (Boyd & Vandenberghe, 2004). Later in this section, we will provide some justification for the name 'optimization.'

A basic tool in variational optimization is the concept of the *first variation* of a functional, which is a direct generalization of the directional derivative of a function. More precisely, given

a functional $E : \mathcal{F} \to \mathbb{R}$ and any function $g \in \mathcal{F}$, called *perturbation*, the first variation (or Gâteaux derivative) of E along g, calculated in f, is defined by

$$\delta E(f, g) = \lim_{\varepsilon \to 0} \frac{E(f + \varepsilon g) - E(f)}{\varepsilon}, \tag{10}$$

here, the perturbation g plays the role of the vector $\vec{v}$ in the definition of directional derivative. The generalization of Fermat's interior extrema theorem to variational calculus states that the first variation of a functional computed in any extreme (argmin or argmax) must be null for every perturbation. Moreover, this necessary condition also becomes sufficient under suitable convexity hypotheses or when it is associated with some properties of the second variation, i.e., the first variation of the first variation interpreted as a functional.

The equations

$$\delta E(f, g) = 0 \qquad \forall g \in \mathcal{F}, \tag{11}$$

are called Euler-Lagrange equations. For a better comprehension of the variational framework that we will develop in sections 4 and 5, it is worthwhile to complete this overview by showing explicit examples of functionals given by one or two terms and by explaining why the search for their extrema is called optimization. Let us start with the problem of finding the extremal function $y = f(x)$ whose graph gives the shortest curve that connects two points (x_1, y_1) and (x_2, y_2) in $\mathbb{R}^2$. To find the variational principle associated with this problem, let us recall that the arc length A of the curve represented by the graph of a differentiable function $y = f(x)$ between the points $(x_1, y_1 \equiv f(x_1))$ and $(x_2, y_2 \equiv f(x_2))$ is given by the integral $\int_{x_1}^{x_2} \sqrt{1 + [y'(x)]^2}\, dx$. The solution to our problem will therefore be the argmin of the functional

$$\begin{array}{rccl} A: & C([x_1, x_2]) & \longrightarrow & \mathbb{R} \\ & f & \longmapsto & A(f) = \int_{x_1}^{x_2} \sqrt{1 + [f'(x)]^2}\, dx. \end{array}$$

It can be easily proven that, as expected, the argmin of $A(f)$ is given by the straight line function $\underline{f}(x) = mx + b$, where $m = (y_2 - y_1)/(x_2 - x_1)$ and $b = (x_2 y_1 - x_1 y_2)/(x_2 - x_1)$.

In this case, it is very easy to understand why the argmin of $A(f)$ represents the optimal solution to the problem of finding the shortest curve, because of the *direct interpretation* of $A(f)$ as the arc length functional. It is more difficult to understand what optimality means when the functional associated with a problem is given by two or more terms. To help understand this, let us consider the problem of determining, with a variational principle, the trajectory of a particle moving into space between time t_0 and time t_1 in a conservative physical system[1]. The Italian-French mathematician and physicist Lagrange solved this problem by considering the following functional, that nowadays we call *Lagrangian* in his honor:

$$L(\vec{q}) = \int_{t_0}^{t_1} \mathcal{L}(\vec{q}(t))\, dt \tag{12}$$

[1] A physical system is called conservative if the forces acting on it can be expressed as minus the gradient of a potential energy function V.

where $\mathcal{L}(\vec{q}) = T(\vec{q}) - V(\vec{q})$, $\vec{q} : [t_0, t_1] \to \mathbb{R}^3$ being the time-dependent position function of a moving particle in $\mathbb{R}^3$ and T, V being the kinetic and the potential energy functions of the physical system, respectively.

Lagrange's outstanding achievement is that the argmin of the functional in (12) is the function $\underline{\vec{q}}(t)$ which solves Newton's equation of motion $\vec{F}(t) = m\ddot{\vec{q}}(t)$, $\forall t \in [t_0, t_1]$, where $\vec{F}$ is the result of forces acting on the particle and $\ddot{\vec{q}}(t)$ is the second temporal derivative of $\vec{q}$ (i.e., its acceleration). Since it is well-known that the trajectory of a particle in a (non-relativistic) physical system satisfies Newton's second law of dynamics, Lagrange proved that this is equivalent to searching for the argmin of the functional (12). In other words, the minimization of functional (12) gives the optimal result because it coincides with the solution of Newton's second law of dynamics. Since $\mathcal{L}(\vec{q}(t))$, the integrand function of (12) is given by two terms with opposite signs, i.e., the kinetic energy $T(\vec{q}(t)) = \frac{1}{2}m\|\dot{\vec{q}}(t)\|^2$ and the potential energy $V(\vec{q}(t))$, we can interpret Lagrange's result by saying that in every instant $t \in [t_0, t_1]$, the particle moves along a trajectory which minimizes the difference between the energy that the particle actually has due to its motion (i.e., T) and the energy that the particle could potentially attain (i.e., V). This interpretation is commonly summarized by saying that the trajectory of motion of a particle is given by the *optimal balance* between its kinetic and potential energy in every instant. The balance between them is always present, but it is optimal to describe the trajectory of motion only for the argmin of the functional (12). Of course, when we deal with another problem, not necessarily related to motion of particles or length of curves, optimality will refer, more generally, to the match between the argmin of the functional with an empirical or theoretical law describing a phenomenon or a property. Consistent with the nomenclature just recalled, also in those cases we will say that the argmin is characterized by the optimal balance between the functional terms.

An illuminating example in this sense is the very deep variational interpretation of the histogram equalization of digital images provided in Sapiro and Caselles (1997) that we will discuss in section 5.1. We will see that a digital image with an equalized histogram[2] can be interpreted as the argmin of a functional characterized by the difference between a functional term that describes adjustment to the middle gray-level of the image and another that gives a global measure of contrast carried by the image. Coherently with the considerations above, we will say that histogram equalization is given by the optimal balance between control of the dispersion around the middle gray and contrast intensification.

The possibility of arriving at these highly non-intuitive interpretations of known phenomena and also to predict new ones is what makes the use of variational calculus so prevalent among many different disciplines as a sort of unifying principle. This is one of the main reasons why, in this paper, we analyze the possibility of using variational principles as a bridge between the description of perceptual and non-physical judgments.

Remark 1: Optimization of functionals is conventionally associated with the search for their

[2]We recall that the intensity of a digital image is quantized and bounded in the set of values $\{0, 1, \ldots, 255\}$. A digital image is said to be equalized if each intensity level has the same occurrence probability in the image.

minima. Of course, a minimization problem can be transformed into a maximization one simply by changing the sign of the functional under analysis.

Remark 2: When parameters are involved in the definition of a functional, the interpretation of optimality can be a little more difficult. In fact, functional minimization in this case generates a whole family of optimal solutions that depend on the selection of the parameters appearing in the equations. In these cases, a suitable tuning procedure that may vary from case to case, must be used to set the parameters once and for all, thus providing the (only) optimal solution to the problem under consideration. We will return to this subtle matter in section 5.1 during the variational analysis of histogram equalization.

4. Variational interpretation of brightness matching

In this section, we will discuss an alternative version of the Rudd-Zemach model which leads to analogous predictions and that has the advantage of being understandable in terms of variational principles. As underlined in section 2, the choice made by Rudd and Zemach to pass from eq. (4) to eq. (5) during the description of their WLLR model is questionable. The reason is that the weights of spatially local induction $w_1, w_2, \ldots$ should decrease with the distance between two point of the visual scene; however, in the decomposition of the chain of ratios in eq. (3), the points corresponding to the indexes $k-1$ and k and those corresponding to k and $k+1$ have exactly the same distance. Instead, if we keep fixed the target point i and consider points at an increasing distance from it, then it is perfectly correct to consider weights of decreasing strength.

These considerations have been thoroughly analyzed in previous works on the interpretation of Retinex (Bertalmío, Caselles, & Provenzi, 2009; Provenzi et al., 2005, 2007; Provenzi, Gatta, Fierro, & Rizzi, 2008). Starting from this more coherent Retinex interpretation, we propose the following alternative definition of logarithmic brightness of i, denoted with $\log \Psi(i)$:

$$\log\ \Psi(i) \equiv \sum_{k=i}^{j-1} w_{k-i+1} \log \frac{L_i}{L_{k+1}} + \mu, \tag{13}$$

where, as for $\log\ \Psi(i)$, $\mu \in \mathbb{R}$ is an arbitrary constant. If we compare the formulae of $\log\ \Phi(i)$ and $\log\ \Psi(i)$, we see that in the latter the numerator of each ratio is L_i, this means that the contribution to $\Psi(i)$ is given by the logarithmic ratios between $L(i)$ and the luminance of all the other patches in the visual field, weighted by the distance between i and the patches.

It is easy to see that this model also predicts a linear relationship between $\log L_{D_M}^{\text{Match}}$ and $\log L_{R_T}$ with slopes in $(-1, 0)$ for the visual match in Wallach's experiment. In fact, developing eq. (13) for both sides of the visual field and matching the brightness, we get:

$$w_1 \log \frac{L_{D_M}^{\text{Match}}}{L_{R_M}} + w_2 \log \frac{L_{D_M}^{\text{Match}}}{L_B} + \mu = w_1 \log \frac{L_{D_T}}{L_{R_T}} + w_2 \log \frac{L_{D_T}}{L_B} + \mu, \tag{14}$$

solving this equation w.r.t. $\log L_{D_M}^{\text{Match}}$ we have

$$\log L_{D_M}^{\text{Match}} = \log L_{D_T} + \frac{w_1}{w_1 + w_2} \log L_{R_M} - \frac{w_1}{w_1 + w_2} \log L_{R_T}. \tag{15}$$

We stress that the absolute value of the slope $w_1/(w_1 + w_2) = \beta_1$ is naturally bounded between -1 and 0, so it also accounts for Rudd-Zemach's observations, even without the hypothesis that $w_2 < w_1$. The weight w_2 can be expressed in terms of the measured value of β_1 as follows:

$$w_2 = \frac{1 - \beta_1}{\beta_1} w_1. \tag{16}$$

If we now add two other rings in both the match and target bipartite field, and again use the outermost ring in the target field as an independent variable, then, by direct computation, it can be proven that expression of the induction weight w_3 corresponding to this new, and more distant, ring predicted by our model is the following:

$$w_3 = \frac{1 - \beta_2}{\beta_2} w_2 - w_1, \tag{17}$$

where $\beta_2 \in (0, 1)$ is the absolute value of the measured slope of the linear relationship in the logarithmic domain between the $L_{D_M}^{\text{Match}}$ and the logarithmic luminance of the new outermost target ring. By iterating the process we find the following formula for the n-th induction weight (corresponding to the configuration given by one disk and $n - 1$ rings):

$$\begin{cases} w_2 = \frac{1-\beta_1}{\beta_1} w_1 \\ w_n = \frac{1-\beta_{n-1}}{\beta_{n-1}} w_{n-1} - \sum_{k=1}^{n-2} w_k \qquad n \geq 3, \end{cases} \tag{18}$$

$\beta_{n-1} \in (0, 1)$ is the absolute value of the measured slope of the linear relationship in the logarithmic domain between the $L_{D_M}^{\text{Match}}$ and the logarithmic luminance of the outermost $(n - 1)$-th target ring.

It is natural to search for a generalization of Rudd-Zemach's model that is valid for arbitrary spatial configurations and not just for the special one discussed in their experiments. To do that, we will distinguish between a discrete and a continuous context. In a discrete context, we will denote the discrete visual field with the lattice $\Omega \subset \mathbb{Z}^2$, the coordinates of two arbitrary points in Ω as $x = (x_1, x_2)$, $y = (y_1, y_2)$, and the corresponding luminance values as $L(x)$ and $L(y)$, respectively. The equivalent of the logarithmic brightness of x defined in eq. (13) in this case is:

$$\log \Psi(x) = \sum_{y \in \Omega} w(\|x - y\|) \log \frac{L(x)}{L(y)} + \mu, \qquad \text{Discrete context,} \tag{19}$$

$w(\|x - y\|)$ being a weight function which decreases with the distance $\|x - y\|$ and μ is an arbitrary constant. In a continuous context, Ω is a subset of $\mathbb{R}^2$, and of course the discrete sum must be replaced by an integral:

$$\log \Psi(x) = \int_{\Omega} w(\|x - y\|) \log \frac{L(x)}{L(y)} \, dy + \mu, \qquad \text{Continuous context.} \tag{20}$$

The advantage of this formulation, over that of Rudd and Zemach, is that it is possible to provide a variational interpretation of formulae (19), (20), as the following proposition states.

Proposition 4.1. *The achromatic logarithmic brightness $\Psi(x)$ of an arbitrary point $x \in \Omega$ is the argmin of the functional*

$$E_w(\log\ L) = \frac{1}{2}\sum_{x\in\Omega}\left(\log\frac{L(x)}{\mu}\right)^2 - \frac{1}{4}\sum_{x\in\Omega}\sum_{y\in\Omega} w(\|x-y\|)\left(\log\frac{L(x)}{L(y)}\right)^2, \tag{21}$$

in the discrete scenario, and of the functional

$$E_w(\log\ L) = \frac{1}{2}\int_{\Omega}\left(\log\frac{L(x)}{\mu}\right)^2 dx - \frac{1}{4}\iint_{\Omega^2} w(\|x-y\|)\left(\log\frac{L(x)}{L(y)}\right)^2 dxdy. \tag{22}$$

in the continuous scenario.

The proof of this proposition is provided in the Appendix. Here, we are more interested in its interpretation. To this aim it is convenient to write the terms appearing in the functional as follows:

- $D(\log L) \equiv \frac{1}{2}\sum\limits_{x\in\Omega}\left(\log\frac{L(x)}{\mu}\right)^2$: *discrete quadratic dispersion term*,
- $C_w(\log L) \equiv \frac{1}{4}\sum\limits_{x\in\Omega}\sum\limits_{y\in\Omega} w(\|x-y\|)\left(\log\frac{L(x)}{L(y)}\right)^2\ dxdy$: *discrete local quadratic contrast term*,

in the discrete domain, and

- $D(\log L) \equiv \frac{1}{2}\int\limits_{\Omega}\left(\log\frac{L(x)}{\mu}\right)^2\ dx$: *continuous quadratic dispersion term*,
- $C_w(\log L) \equiv \frac{1}{4}\iint\limits_{\Omega^2} w(\|x-y\|)\left(\log\frac{L(x)}{L(y)}\right)^2\ dxdy$: *continuous local quadratic contrast term*,

in the continuous domain.

As we specified in the introduction, in order to avoid confusion, we chose to use the word dispersion instead of assimilation because of the different meanings that the latter term can have in perception research (in particular brightness perception) and in the cognition field. From a general point of view, in our work the term 'dispersion' is employed to mean a tendency that simply weakens contrast (as it happens in the Rudd-Zemach induction model) and makes the target appear more like the other members of its contextual set. The key observation is that the minimization of $E_w(\log L) = D(\log L) - C_w(\log L)$ is reached through the *simultaneous minimization of* $D(\log L)$ *and maximization of* $C_w(\log L)$ (because of the minus sign appearing in front of the contrast term). $D(\log L)$ is minimized for values of $L(x)$ close to μ, and this holds for

all $x \in \Omega$. The numerical value of μ is not important; what really matters is that μ is a constant, thus we infer that minimizing $D(\log L)$ corresponds to inducing a uniform stimulus, which has the same value in every point x. This is the reason why we called $D(\log L)$ the dispersion term: If only this term existed, then all the stimuli coming from different points of the scene would be blended into a unique stimulus, uniform across the visual field. Of course, this is not what happens when we look at a scene, so there must be an opposing mechanism to dispersion, which is defined precisely by the action of maximizing $C_w(\log L)$. In fact, $C_w(\log L)$ is maximized when the terms of the integral, i.e., $\left(\log \frac{L(x)}{L(y)}\right)^2 = (\log\, L(x) - \log\, L(y))^2$ are intensified, but that means that the differences $\log\, L(x) - \log\, L(y)$ must be amplified as much as possible, which, in turn, corresponds to maximizing the contrast of the image content. The presence of the weight function $w(\|x-y\|)$ guarantees that contrast amplification respects the locality of visual perception in the sense discussed above, which is a context-driven effect. All these considerations explain why we named $C_w(\log L)$ local contrast.

To summarize, the logarithmic brightness values $\Psi(x)$ can be interpreted as being *the optimal balance between two opposite mechanisms*: one that tends to adjust all stimuli to a constant, uniform value, and the other that tends to do the opposite, i.e., to amplify as much as possible the differences among all stimuli in a local, or context-driven, way. The results just obtained have many similarities with the theory of perceptually-inspired color correction (Bertalmío, Caselles, Provenzi, & Rizzi, 2007; Bertalmío, Caselles, & Provenzi, 2009; Palma-Amestoy, Provenzi, Bertalmío, & Caselles, 2009; Provenzi & Caselles 2014). The functionals considered in these papers balance dispersion and local contrast enhancement to perform perceptual color correction, i.e., to modify the intensities of pixels in digital images to approach the sensation produced by the real-world scene.

5. A general variational framework for context-driven effects

In this section, we generalize the concepts introduced in the previous section. The variational interpretation of brightness matching refers to the simultaneous contrast phenomenon as initially investigated by Wallach. Until now, we have taken into account the influence of context on psychological judgments along the physical dimension of luminance and for matching experiments. However, we think that the same analysis may also hold for judgments along non-physical dimensions, such as beauty or morality, and not necessarily for matching experiments. The aim of this section is to provide an abstract framework that can be adapted to specific experiments through an appropriate specification of parameters. To guarantee a sufficient degree of abstraction and versatility of the setting, we introduce the following nomenclature:

- $(\Omega, \|\ \|)$: normed space describing the context. The norm $\|\ \| : \Omega \to \mathbb{R}^+$ is needed to measure distances between elements of Ω, and thus to handle local phenomena. Typically Ω is a subset of $\mathbb{Z}^n$ in the discrete case, and of $\mathbb{R}^n$, $n \geq 1$, for continuous contexts, and $\|\ \|$ is the discrete or Euclidean norm, respectively, but nothing prevents Ω and $\|\ \|$ from being a more complicated space and norm. We denote by $x, y, \ldots$ the elements of Ω and with

$|\Omega|$ its cardinality, area, volume, or hyper-volume, depending on its discrete or continuous structure and mathematical dimension, respectively;

- $\mathcal{S} : \Omega \to \mathbb{R}$: stimulus function, such that $\mathcal{S}(x)$ represents the stimulus coming from the element $x \in \Omega$;

- $\mathcal{F}(\Omega)$: functional space where $\mathcal{S}$ belongs. It can be, e.g., $C(\Omega)$, the space of scalar-valued continuous functions on Ω, $L^2(\Omega)$, the square integrable (or finite energy) of scalar-valued functions on Ω, and so on, depending on the problem under analysis;

- $w : \Omega \times \Omega \to \mathbb{R}^+$: induction weight function, $w(x, y)$ and $w(y, x)$ represent the strength of the influence of x on y and of y on x. If these strengths are equal, then we also require w to be a symmetrical function, i.e., $w(x, y) = w(y, x)$ for all $x, y \in \Omega$. w is typically a decreasing function of the distance $\|x - y\|$, to keep into account the fact that closer stimuli (with respect to the norm $\| \ \|$) influence each other more strongly than distant ones.

Table 1 provides the identification of the objects of the variational framework just defined in the concrete case of two psychological experiments concerning judgments along physical and non-physical dimensions.

	Physical judgment (Brightness matching)	**Non-physical judgment (Seriousness of an action)**
x	Pixel position on the screen	Time at which a question is formulated
Ω	Screen where stimuli are presented	Real line $\mathbb{R}$ of instants of time
$\| \ \|$	Euclidean norm between pixels	Absolute value of the difference between two temporal moments
$\mathcal{S}(x)$	Luminous intensity of the stimulus in x	Question about seriousness of an action asked at the time x
$\mathcal{F}(\Omega)$	Space of all possible luminous stimuli	Space of all possible questions about moral seriousness

Table 1: Identification of the objects defined in the variational framework in the case of a psychological experiment along a physical (left) and a non-physical (right) dimension.

As discussed in the previous section, the general functional associated with a context-driven effect must be characterized by the balance between two opposing mechanisms: dispersion and contrast. In the case studied in the previous section, the dispersion is *global*, i.e., the weight function w is constant for every couple of points $x, y \in \Omega$ but, for the sake of a more general framework, we will also allow the dispersion term to be *local*, i.e., to have a weight function w which decreases with the distance between x and y. As shown in previous works (Palma-Amestoy et al., 2009), one must also add an *attachment-to-original-data constraint* that consists of an adjustment to the original stimuli values. This helps to guarantee that the context-driven

perceptual or cognitive values of the stimuli do not excessively depart from the original ones, and it also helps the convergence of iterative numerical schemes for the approximation of the argmin of the functional. To better understand the exigence of an attachment-to-original-data constraint, we can think about the case of color perception of a surface under different light conditions, e.g., neon and tungsten lamp. While the human visual system is able to partially discount the noticeable spectral difference between these two illuminants, we still perceive a dominant bluish color under neon light and a reddish color under the tungsten light, which is symptomatic of an attachment to the original stimuli.

We thus introduce in our framework the three following functionals:

- $B : \mathcal{F}(\Omega) \to \mathbb{R}$: *attachment-to-original-stimuli functional*. The effect of its minimization is to preserve the original values of the stimuli;

- $C_{w_c} : \mathcal{F}(\Omega) \to \mathbb{R}$: *contrast amplification functional*. The effect of its minimization is to enhance the differences between any two stimuli $\mathcal{S}(x)$, $\mathcal{S}(y)$, with $x \neq y$. The presence of the weight function w_c allows contrast amplification to be context-dependent;

- $D_{w_d} : \mathcal{F}(\Omega) \to \mathbb{R}$: *dispersion functional*. The effect of its minimization is to adjust the values of the stimuli in the direction of the average value of the context, which can be global μ (and in this case w_d is constant), or local $\mu(x)$ (and in this case the weight function w_d is not constant);

- Moreover, we *require the minimization of these functionals to give rise to dimensionally coherent equations*, where 'dimension' here means the unity of measurement associated with stimuli.

The simultaneous presence of the three effects discussed above can be represented by the following functional, given by a linear combination of the three previous ones:

$$E_{w_c,w_d,b,c,d}(\mathcal{S}) = bB(\mathcal{S}) + cC_{w_c}(\mathcal{S}) + dD_{w_d}(\mathcal{S}) \qquad b, c, d \in \mathbb{R}, \tag{23}$$

b is the attachment-to-original-data strength, c is the contrast strength, and d is the dispersion strength. It is important to stress that the argmin of $E_{w_c,w_d,b,c,d}(\mathcal{S})$ can directly represent the perceived stimulus but it can also be used for brightness matching experiments, as discussed in section 4. In this last case, once the stimulus to be matched is identified, the Euler-Lagrange equations corresponding to the argmin must be matched and solved with respect to the perceived quantity that one wants to measure. We also stress that the simultaneous presence of dispersion and contrast is essential: without the dispersion term, the contrast process would increase without boundaries and vice-versa, if only the dispersion term were allowed to work without the compensation of the contrast one, then the stimulus would be absorbed in the context and the final percept would become trivial.

In the perception and social cognition literature there is no consensus on circumstances that lead to either one of the two effects (Gilchrist, 2015). Many variables can affect the resulting

percept, as described previously in section 1. By changing parameters and the form of the functional, our approach is able to incorporate different assumptions about the variables involved in a particular experiment. Among the various theoretical ideas within perception, gestalt laws may provide (at least) some clues about how and when an observer will experience dispersion or contrast (Agostini & Galmonte, 2002; Koffka, 1935; Wertheimer, 1923). According to Agostini and Galmonte (2002), when two items belong to different perceptual groups, a contrast effect will be observed. On the contrary, when an item is organized into one or the other of two perceptual groups, a dispersion effect will be observed. Such a mechanism is (at least partially) in line with social cognition studies about variables that affect context effects.

There are reasons to believe that the equilibrium between dispersion and contrast is also involved in non-physical judgments. One example is given by the judgment of attractiveness: an observer is exposed to a face that is considered average-looking in isolated conditions, then he/she has to judge the face again in the presence of either an attractive or an unattractive face that plays the role of the context effect. Both dispersion and contrast effects have been reported (Kenrick & Gutierres, 1980). The same happens for a single attractive or unattractive face acting as context. For example, in the appearance of a face there can be some attractive traits (e.g., the eyes, the form of the face and so on), some average-looking traits (e.g., the lips), and some unattractive characteristics (e.g., a scar, or other imperfections). The dispersion term indicates the tendency to create a holistic and uniform perception of face attractiveness, reducing the differences that may be present. At the same time, contrast maximization implies that the differences in attractiveness of the different features of a face are enhanced. Social cognition literature about contrast and dispersion effects (Biernat, 2005; Stapel & Koomen, 1998) underlines the role of the context as the backdrop against which the stimuli is experienced. More specifically, the importance of reducing the within-stimulus and the within-context differences is often stressed in order to make the comparison. Stereotypes (Biernat & Kobrynowicz, 1997; Stapel & Koomen, 1998) as well as social standards (Higgins, 1987, 1990) are unidimensional and uniform representations of attitudes and traits of a given person or a group. Famous effects such as the halo effect (Nisbett & Wilson, 1977), the focusing effect (Schkade & Kahneman, 1998), and anchoring biases (Tversky & Kahneman, 1974) are based on the tendency to rely very strongly on a single piece of information at the expense of other elements. Again, such effects rely on the idea that the within-differences of the single piece of information used as a criterion are minimized. Such considerations testify to the importance of the dispersion term in our variational model in order to perceive stimuli as uniform as possible to make a comparison; concurrently, this comparison is enhanced by the contrast effect.

To summarize, the proposed setting provides a view from above that helps put into evidence and formalize the two conflicting actions of dispersion and contrast in each judgment (either psychophysical measures in physical domains or non-physical judgments) whose balance determines the final percept. In the following subsections, we will provide significant examples of functionals of the kind just described.

5.1. Examples of functionals $E_{w_c,w_d,b,c,d}(\mathcal{S})$ already present in the literature

We start the description of a functional of the type (23) already present in the literature with an important example of a functional that mixes dispersion and contrast. Even though it does not represent a context-driven effect, we have chosen to begin with it because it is the starting point for the development of the context-driven functionals that we will discuss afterwards.

The so-called Caselles-Sapiro functional, proposed by Sapiro and Caselles (1997), provides a highly non-intuitive variational interpretation of histogram equalization. Whenever we have the histogram of a variable, its equalization forces the histogram to be flat, so that all the realizations of that variable have the same probability of occurrence. The very remarkable finding of these authors was that, using the notations introduced above, and considering normalized stimuli $\mathcal{S}(x) \in [0, 1]$, the argmin of the following functional

$$E_{\text{Hist. Eq.}} = 2 \sum_{x\in\Omega} \left(\mathcal{S}(x) - \frac{1}{2} \right)^2 - \frac{1}{|\Omega|} \sum_{x\in\Omega} \sum_{y\in\Omega} |\mathcal{S}(x) - \mathcal{S}(y)| \tag{24}$$

in the discrete setting, and

$$E_{\text{Hist. Eq.}} = 2 \int_{\Omega} \left(\mathcal{S}(x) - \frac{1}{2} \right)^2 dx - \frac{1}{|\Omega|} \iint_{\Omega^2} |\mathcal{S}(x) - \mathcal{S}(y)| \, dxdy \tag{25}$$

in the continuous setting, has an equalized histogram. To understand this result, notice that the first term of $E_{\text{Hist. Eq.}}$ can be interpreted as a dispersion functional with respect to the average value between the extremes 0 and 1 taken by $\mathcal{S}(x)$, while the second term can be interpreted as a (non context-dependent) contrast functional. In fact, the first term is clearly minimized by a constant stimulus equal to 1/2, while the second term, due to the presence of the minus sign in front of it, is minimized when the absolute value of the differences between two stimuli are maximized. In light of this result, an equalized stimulus is characterized by the equilibrium between being uniform and having as much diversity as possible, which is a highly non-trivial result.[3]

This profound finding has been used in previous works (Bertalmío et al., 2007, 2009; Palma-Amestoy et al., 2009) to model context-driven effects in color vision and it has been applied to the enhancement of color digital images. In this case, $\mathcal{S}$ is the intensity function of an image in each separated chromatic channel RGB, $I : \Omega \to [0, 1]$, where Ω is the spatial support of the image and $[0, 1]$ is the normalized dynamic range of pixel intensities. The first functional inspired by the results of Caselles and Sapiro appeared in Bertalmío et al. (2007) and can be written as follows:

$$E_1(I) = d \sum_{x\in\Omega} \left(I(x) - \frac{1}{2} \right)^2 + \sum_{x\in\Omega} (I(x) - I_0(x))^2 - c \sum_{x\in\Omega} \sum_{y\in\Omega} w(x, y) S\left(I(x) - I(y)\right), \tag{26}$$

[3]We stress that in Sapiro and Caselles (1997) the authors show that the factor 2 in front of the dispersion term is the only choice of the weight parameter that gives rise to histogram equalization. When we deal with functionals relative to perceptual or cognitive experiments, the parameters setting must be performed by a suitable tuning procedure that may vary from case to case in order to find the optimal solution to our problem.

in the discrete setting, and

$$E_1(I) = d \int_\Omega \left(I(x) - \frac{1}{2}\right)^2 dx + \int_\Omega (I(x) - I_0(x))^2 \, dx - c \iint_{\Omega^2} w(x,y) S\left(I(x) - I(y)\right) dxdy. \quad (27)$$

in the continuous setting, where the function S an antiderivative for a sigmoid. It can be seen that all the functional terms mentioned above appear in $E_1(I)$: The first term is the dispersion to the average intensity level between 0 and 1, the second is the adjustment to the original intensity values $I_0(x)$, and the third is a nonlinear and context-driven contrast term. The nonlinearity is due to the presence of S, which is introduced to mimic some peculiar properties of the human visual system (see Bertalmío et al., 2007 for more detail).

One of the most important characteristics of human vision is its robustness with respect to changes of illumination[4] (Fairchild, 2005; Land & McCann, 1971); this property is not reflected by the analytical expression of the functional $E_1(I)$, which is not invariant with respect to transformations such as $I \mapsto \lambda I$. To remedy this problem, in Palma-Amestoy et al. (2009), another kind of functional was considered, namely:

$$\begin{aligned} E_2(I) = d \sum_{x\in\Omega} \left(\mu \log \frac{\mu}{I(x)} - (\mu - I(x))\right) + \sum_{x\in\Omega} \left(I_0(x) \log \frac{I_0(x)}{I(x)} - (I_0(x) - I(x))\right) \\ + c \sum_{x\in\Omega} \sum_{y\in\Omega} w(x,y)\, \varphi\left(\frac{\min(I(x), I(y))}{\max(I(x), I(y))}\right) \end{aligned} \quad (28)$$

in the discrete setting, and

$$\begin{aligned} E_2(I) = d \int_\Omega \left(\mu \log \frac{\mu}{I(x)} - (\mu - I(x))\right) dx + \int_\Omega \left(I_0(x) \log \frac{I_0(x)}{I(x)} - (I_0(x) - I(x))\right) dx \\ + c \iint_{\Omega^2} w(x,y)\, \varphi\left(\frac{\min(I(x), I(y))}{\max(I(x), I(y))}\right) dxdy \end{aligned} \quad (29)$$

in the continuous setting, where the functional parameter φ is a strictly increasing positive function.

[4]Taking into account the cognitive domain, it is possible to draw a parallel between such robustness to changes of illumination with the well-known hedonic treadmill theory (Brickman & Campbell, 1971; Eysenck, 1990). This theory is based on the observed tendency of a quick restoration to a relatively stable level of happiness after the experience of major positive or negative life events (Brickman & Campbell, 1971; Diener, Lucas, & Scollon, 2006; Fujita & Diener, 2005; Lykken & Tellegen,1996; Silver, 1982; Wildeman, Turney, & Schnittker 2014). For example, interviewing lottery winners and paraplegics in order to assess their happiness levels due to the positive (winning the lottery) or negative (becoming paralyzed) event, it was observed that after a few years, both groups returned to their baseline level of happiness (Brickman & Campbell, 1971). As a matter of fact, the hedonic treadmill metaphor explicitly refers to adaptation processes in sensory domains. Although much evidence supports this theory, there is an ongoing debate about how to measure happiness, the role of individual differences, and the differential impact of particular events on happiness (Frederick, 2007).

Let us discuss the meaning of the terms in E_2, starting with the last one. First of all, notice that the ratio defined by $\min(I(x), I(y))/\max(I(x), I(y))$ is minimized when the minimum decreases and the maximum increases, which of course corresponds to a contrast amplification. Moreover, the fraction is invariant with respect to transformations like $I \mapsto \lambda I$. In order to be dimensionally coherent, the quadratic dispersion and attachment can no longer be used here, thus the authors selected the so-called *entropic* dispersion and attachment terms; the first and the second appearing in (29), respectively. Given the statistical interpretation of entropy, minimizing the first terms amounts to minimizing the variability of intensity levels around the average μ and around the original data $I_0(x)$, which is what is expected from the dispersion and attachment terms. The advantage, with respect to the quadratic terms, is that the first derivatives of the entropic functionals are dimensionally coherent with those coming from the contrast term. We refer again to Palma-Amestoy et al. (2009) for more details. We hope that these examples may help the reader understand how the contrast and dispersion terms can vary according to the particular features exhibited by a problem.

5.2. Variational interpretation of linear relationships in the linear and logarithmic scale for physical and non-physical dimensions

The argmin of functionals of the type (23), in the majority of cases, is expressed via an implicit equation that cannot be solved analytically. However, for some *particularly simple, yet already significant*, expressions of the functional terms, the argmin can be expressed via a linear equation that can be directly compared with actual observations through a linear fit. It is particularly important to analyze these situations, because they are, by far, the most common in psychology. In particular, let us distinguish between linear behavior in the linear and logarithmic scale. A linear behavior in the linear scale is simply expressed by the law: $u = \alpha + \beta v$, where u and v are abstract variables and α and β are coefficients. Linear relationships like the previous one are in general associated with quadratic functionals. To avoid repeating functional expressions, we will consider only the discrete formulation, as the continuous version can be easily obtained by replacing sums with integrals.

Let us start by considering the following functional:

$$E_{a,b,c}(\mathcal{S}) = \frac{d}{2}\sum_{x\in\Omega}(\mathcal{S}(x)-\mu)^2 + \frac{b}{2}\sum_{x\in\Omega}(\mathcal{S}(x)-\mathcal{S}_0(x))^2 - \frac{c}{2}\sum_{x\in\Omega}(\mathcal{S}(x)-\mu)^2\,, \tag{30}$$

where $\mathcal{S}(x)$ is an abstract stimulus, $\mathcal{S}_0(x)$ is the actual stimulus and

$$\mu = \frac{1}{|\Omega|}\sum_{x\in\Omega} S(x), \tag{31}$$

is the constant value of an average background in which $\mathcal{S}_0(x)$ is embedded. A practical example of such a configuration is when $\mathcal{S}$ is a visual stimulus and $x \in \Omega$ is a time variable, if the intervals x_n, x_{n+1} between the n-th and the $(n+1)$-th presentations of the visual stimuli are long enough, then no interaction between $\mathcal{S}(x_n)$ and $\mathcal{S}(x_{n+1})$ is expected. Thus, only the interaction

between $\mathcal{S}(x)$, $x \in \Omega$ and the background μ is expected. Notice that in this particular case, the difference between dispersion and contrast is represented simply by the opposite sign in front of the corresponding terms and by the two strengths $a \neq c$, which must be different to avoid the trivial functional represented only by the attachment-to-data term.

The basic variable in this configuration is the difference $\mathcal{S}(x) - \mu$, which can express either dispersion or contrast. Thus, as we have done in section 4, if we compute the Euler-Lagrange equations of the functional (30) and express the argmin $\underline{\mathcal{S}}(x)$ in terms of $\mathcal{S}(x) - \mu$ as follows:

$$\underline{\mathcal{S}}(x) = \mathcal{S}_0(x) + \frac{d}{b}\left(\mathcal{S}(x) - \mu\right) - \frac{c}{b}\left(\mathcal{S}(x) - \mu\right), \tag{32}$$

then we can interpret $\underline{\mathcal{S}}(x)$ as the stimulus corresponding to the optimal balance between contrast, dispersion, and attachment to the original stimulus $\mathcal{S}_0(x)$. $\underline{\mathcal{S}}(x)$ can be rearranged as follows:

$$\underline{\mathcal{S}}(x) = \mathcal{S}_0(x) + \frac{c-d}{b}\mu - \frac{c-d}{b}\mathcal{S}(x). \tag{33}$$

Eq. (33) expresses a linear relationship, as much as eq. (15). However, in eq. (33) it is clear that the slope of the linear relationship, i.e., $-(c-d)/b$ is negative if the contrast between $\mathcal{S}(x)$ and background prevails over the adjustment, i.e., $c > d$, whereas the slope is positive when $d > c$, i.e., if adjustment between $\mathcal{S}(x)$ and the background overcomes the contrast effect.

The great generality of variational principles allows us to easily extend the previous configuration to a new one, where we allow an interaction, global or local, among the $\mathcal{S}(x)$s. Let us start by adding a *global quadratic interaction*:

$$\begin{aligned}\tilde{E}_{b,c_1,c_2,d}(\mathcal{S}) = &\frac{d}{2}\sum_{x\in\Omega}\left(\mathcal{S}(x)-\mu\right)^2 + \frac{b}{2}\sum_{x\in\Omega}\left(\mathcal{S}(x)-\mathcal{S}_0(x)\right)^2 \\ &- \frac{c_1}{2}\sum_{x\in\Omega}\left(\mathcal{S}(x)-\mu\right)^2 - \frac{c_2}{4}\sum_{x\in\Omega}\sum_{y\in\Omega}\left(\mathcal{S}(x)-\mathcal{S}(y)\right)^2,\end{aligned} \tag{34}$$

where the contrast strengths c_1, c_2 can be equal or different. The argmin of this functional is given by the following equation:

$$\underline{\mathcal{S}}(x) = \mathcal{S}_0(x) + \frac{d}{b}\left(\mathcal{S}(x)-\mu\right) - \frac{c_1}{b}\left(\mathcal{S}(x)-\mu\right) - \frac{c_2}{b}\sum_{y\in\Omega}\left(\mathcal{S}(x)-\mathcal{S}(y)\right) \tag{35}$$

which, taking into consideration eq. (31), can be rearranged in this way:

$$\underline{\mathcal{S}}(x) = \mathcal{S}_0(x) + \frac{c_1 - d + c_2|\Omega|}{b}\mu - \frac{c_1 - d + c_2|\Omega|}{b}\mathcal{S}(x). \tag{36}$$

The generalization, which takes into account a *local quadratic interaction*, is provided by these functionals:

$$\begin{aligned}\tilde{E}_{b,c_1,c_2,d,w}(\mathcal{S}) = &\frac{d}{2}\sum_{x\in\Omega}\left(\mathcal{S}(x)-\mu\right)^2 + \frac{b}{2}\sum_{x\in\Omega}\left(\mathcal{S}(x)-\mathcal{S}_0(x)\right)^2 \\ &- \frac{c_1}{2}\sum_{x\in\Omega}\left(\mathcal{S}(x)-\mu\right)^2 - \frac{c_2}{4}\sum_{x\in\Omega}\sum_{y\in\Omega} w(\|x-y\|)\left(\mathcal{S}(x)-\mathcal{S}(y)\right)^2.\end{aligned} \tag{37}$$

The argmin equation in this case is:

$$\underline{\mathcal{S}}(x) = \mathcal{S}_0(x) + \frac{d}{b}(\mathcal{S}(x) - \mu) - \frac{c_1}{b}(\mathcal{S}(x) - \mu) - \frac{c_2}{b}\sum_{y\in\Omega} w(\|x-y\|)(\mathcal{S}(x) - \mathcal{S}(y)) \tag{38}$$

which, again, can be re-arranged as follows:

$$\underline{\mathcal{S}}(x) = \mathcal{S}_0(x) + \frac{c_1 - d + c_2}{b}\mu + \frac{c_2}{b}\sum_{y\in\Omega}\mathcal{S}(y) - \frac{c_1 - d + c_2\sum_{y\in\Omega} w(\|x-y\|)}{b}\mathcal{S}(x). \tag{39}$$

For the sake of completeness, let us observe that interaction among stimuli can also be more complicated than the quadratic one: A common example is given by the so-called *logistic* or *sigmoid-like* interaction, in which the basic contrast variable is represented by $\sigma(\mathcal{S}(x) - \mathcal{S}(y))$, where σ is a sigmoid. If $\mathbb{S}$ is an antiderivative of σ, then the functional corresponding to a sigmoid-like interaction is the following:

$$\begin{aligned}\tilde{E}_{b,c_1,c_2,d,\mathbb{S}}(\mathcal{S}) = {} & \frac{d}{2}\sum_{x\in\Omega}(\mathcal{S}(x) - \mu)^2 + \frac{b}{2}\sum_{x\in\Omega}(\mathcal{S}(x) - \mathcal{S}_0(x))^2 \\ & - \frac{c_1}{2}\sum_{x\in\Omega}(\mathcal{S}(x) - \mu)^2 - \frac{c_2}{4}\sum_{x\in\Omega}\sum_{y\in\Omega}\mathbb{S}(\mathcal{S}(x) - \mathcal{S}(y)),\end{aligned} \tag{40}$$

in the case of global interaction, and

$$\begin{aligned}\tilde{E}_{b,c_1,c_2,d,w,\mathbb{S}}(\mathcal{S}) = {} & \frac{d}{2}\sum_{x\in\Omega}(\mathcal{S}(x) - \mu)^2 + \frac{b}{2}\sum_{x\in\Omega}(\mathcal{S}(x) - \mathcal{S}_0(x))^2 \\ & - \frac{c_1}{2}\sum_{x\in\Omega}(\mathcal{S}(x) - \mu)^2 - \frac{c_2}{4}\sum_{x\in\Omega}\sum_{y\in\Omega} w(\|x-y\|)\mathbb{S}(\mathcal{S}(x) - \mathcal{S}(y)),\end{aligned} \tag{41}$$

in the case of a local interaction. The analysis of the Euler-Lagrange equations in this case is more involved than in the previous case.

The second type of linear behavior that we want to model with variational principles is the one in the logarithmic scale, expressed by the law: $\ln u = \alpha + \beta \log v$. This relationship expresses a *power law in the linear scale*; in fact, the exponentiation of both sides of the previous equation gives $u = e^{\alpha}e^{\beta\log v} \equiv \overline{\alpha}e^{\log v^{\beta}} = \overline{\alpha}v^{\beta}$. To write the functionals from (30) to (37) in the logarithmic domain, one simply has to perform the change of variable $\mathcal{S} \mapsto \tilde{\mathcal{S}} \equiv \log\mathcal{S}$ and solve the corresponding Euler-Lagrange equations with respect to the logarithmic variables. In particular, it is interesting to re-write the functional (37) in the logarithmic domain:

$$\begin{aligned}\tilde{E}_{b,c_1,c_2,d,w}(\tilde{\mathcal{S}}) = {} & \frac{d}{2}\sum_{x\in\Omega}\left(\log\frac{\mathcal{S}(x)}{\mu}\right)^2 + \frac{b}{2}\sum_{x\in\Omega}\left(\log\frac{\mathcal{S}(x)}{\mathcal{S}_0(x)}\right)^2 \\ & - \frac{c_1}{2}\sum_{x\in\Omega}\left(\log\frac{\mathcal{S}(x)}{\mu}\right)^2 - \frac{c_2}{4}\sum_{x\in\Omega}\sum_{y\in\Omega} w(\|x-y\|)\left(\log\frac{\mathcal{S}(x)}{\mathcal{S}(y)}\right)^2,\end{aligned} \tag{42}$$

if we set $d = c_2 \equiv 1$, $b = c_1 \equiv 0$ and we use the functional for a matching experiment, then we obtain exactly the same variational framework used to interpret Rudd-Zemach's model of brightness induction.

We conclude by noting that the ability to interpret linear equations in the linear and logarithmic domain via variational principles is of paramount importance; in fact, as we have seen in this section, the linear coefficients can be directly interpreted as the difference between the strength of dispersion and contrast in the particular phenomenon under analysis. Moreover, possible deviations from linear behavior can easily be integrated in the variational model with the introduction of nonlinear perturbation terms in the analytical expression of the functional. Finally, let us notice that, if we deal with a nonphysical dimension, all the previous considerations still work by replacing the stimulus $\underline{\mathcal{S}}$ with a judgment function and the average value of the background μ with a context K. Thus, in this case, a linear relationship between the judgment function J and the context K can be written as follows:

$$J(x) = \alpha + \beta K. \tag{43}$$

As we have seen above, depending on the analytical expression of the functional terms, we can predict the sign and the bound the value of the slope of the linear relationship. In fact, if the empirical measurements of a direct judgment experiment fit with eq. (33) with μ replaced by K, then a measured negative slope means that contrast enhancement prevails with respect to dispersion control, while a measured positive slope means the opposite. Furthermore, if the empirical measurements of a matching judgment experiment can be described by the argmin of functionals of the type (37) or (42), with $d = c_2 \equiv 1$, $b = c_1 \equiv 0$, then we can predict that the absolute value of the slope will be bounded between 0 and 1, as discussed in section 4. Notice that a null value for the coefficient in front of a functional term means that the term does not have an influence on the global functional, and hence on the final percept; instead, a unitary value for the coefficient means that the term is accounted for its intrinsic strength, without being weighted. The great versatility of variational principles allows, in any case, the possibility of changing these coefficients to account for different theoretical proposals.

6. The empirical case for the contextual variational framework

In this section, we provide a preliminary validation of the predictions of our variational framework with respect to the balance between dispersion and contrast. We will particularly focus on eq. (33) and we will assume that functionals of the type (37), with $d = c_2 \equiv 1$, $b = c_1 \equiv 0$, can be used for a variational description of the experiment under analysis. Two issues that need to be addressed are: (i) contextual phenomena can be modeled by a linear relationship even for non-physical judgments; (ii) the slope of the linear relationship is bounded between 0 and 1 when dispersion control prevails and between -1 and 0 when contrast enhancement is stronger.

6.1. Constraints about the slope of the linear relationship between context and judgment

The literature on context effects related to non-physical dimensions judgments is vast (Biernat, 2005; Kenrick & Gutierres, 1980; Matthews & Stewart, 2009b; Moskowitz, 2005; Parducci, 1968; Parker et al., 2008; Pepitone & DiNubile, 1976; Stapel et al., 1997). A review of the literature on context effects in social cognition or cognitive psychology is beyond the scope of this article. Moreover, an actual check of the predictions is not always feasible because in many papers there are not enough data to verify such predictions. Indeed, only a limited number of papers, which deal with a linear model, reported the slope of the contextual effect or, alternatively, at least two points related to the same judgments in different contexts (usually mean values measured on the same scale). In the last case, of course, the presence of an actual linear relationship is questionable. Despite these difficulties, an unsystematic and limited check of the literature tends to support the variational framework. Papers that allow us to verify wheter the constraints previously discussed related to eq. (43) hold have usually measured a particular dimension on a Likert scale of a stimulus $\mathcal{S}$.

Disclaimer: the application of statistical or mathematical models to the responses collected on a Likert scale is a common but flawed procedure in the social sciences. For the sake of completeness, in this section we will provide several quantitative examples in line with this procedure. However, we stress that the comparison between the *ordinal values* that constitute the Likert scale and the *numerical* discrete or continuous values of a mathematical model is still an open problem in the social sciences (Camparo, 2013; Carifio & Perla, 2007; Casacci & Pareto, 2015; Wegener, 1982; Young, 1975). Future works should address explicitly the problem of how a monotone transformation of the non-physical judgments could affect the comparison with the variational model predictions; among the different strategies, a possible solution can be to employ the optimal scaling approach to change the distances between the categories until a particular optimum is reached (Takane, 2005). In order to provide a validation that is free from the Likert scale issue, we will describe the study of Matthews and Stewart (2009b) which deals with the psychophysics of an object's prices (White & Vilmain, 1986). In this case, even if the participants are called to produce a non-physical judgment, the responses are not collected by means of a Likert scale.

Let us start our analysis by considering a hypothetical example: The stimulus $\mathcal{S}$ could be the picture of a person and the dimension could be the attractiveness of the person. Taking the mean (or median) value of the observed responses in a sample, it is possible to obtain the judgment J (in isolation or in different contexts). Contextual information K consists of stimuli belonging to the same class of $\mathcal{S}$, but characterized by larger or smaller values of the same dimension (e.g., a very beautiful person or a very ugly person). So, in a non-physical judgment experiment, participants are asked to judge a stimulus $\mathcal{S}$ *and* a corresponding context K.[5] Usually, there are only two

[5]It is important to note that, in non-physical judgments, the context K must also be judged in order to be defined. This is not the case for psychophysical experiments, where the context can be measured through a physical device.

different context configurations, K_1, K_2, which are judged and are considered stable with respect to the introduction of a further stimulus. With J_1 and J_2, we will indicate the judgment of the stimulus $\mathcal{S}$ in the contexts K_1 and K_2, respectively. For the sake of simplicity, let us keep the same symbols J_1, J_2, K_1, K_2 for their mean or median values in the Likert scale rating. If a linear relationship is found between the stimulus judgments J_1 and J_2 and the context judgments K_1 and K_2, then the slope of the linear relationship will be:

$$\beta = \frac{J_2 - J_1}{K_2 - K_1}. \tag{44}$$

Table 2 reports K and J values found in three studies; two about contrast effect (Leding, Horton, & Wootan, 2015; Parducci, 1968), and one about dispersion effect (Tversky & Kahneman, 1974), along with the slope β computed by using the formula (44). With regard to the contrast effect, Parducci (1968) reported the rating of the seriousness (on a 5-point Likert scale) of the same set of acts (the stimulus S) when they were considered together with trivial or very bad actions (the context K). The mean values of the contextual sentences were, respectively, 1.97 for K_1, mild context, and 4.06 for K_2, bad context. The mean values of the stimulus judgments in the two previous contexts were 3.21 for J_1 and 2.69 for J_2, respectively, thereby highlighting a contrast effect. In line with our framework, the slope β of equation (43) is equal to -0.25. A more recent study (Leding et al., 2015) has investigated contrast effects in judging attractiveness. There is a substantial amount of psychological literature on attractiveness ratings, as well. For example, Kenrick and Gutierres (1980) found that exposure to attractive women caused a contrast effect in attractiveness judgments about photographs of average-looking women. Moreover, given the increasing access to the Internet and the importance of visual representations of individuals in virtual environments, Leding et al. (2015) have investigated whether exposure to avatars with different levels of attractiveness (the context K) may influence judgments of the attractiveness of real people (J). Judgments were made on a 5-point Likert scale (from -2 to 2). In their first study (Leding et al., 2015, experiment 1), the authors found that, when the attractiveness of the avatars was high (K_1 = 0.88), observers rated the photo of a person as less attractive (J_1 = -0.87). By contrast, when the attractiveness of the avatars was low (K_2 = -1.24), the same person was judged as being more attractive (J_2 = -.30). Furthermore, in experiment 2, they also investigated the order of presentation (the rating of the three avatars and then the evaluation of the photo and vice versa). When the avatars were presented first, the contrast effect was replicated; in the case of high attractiveness avatars (K_1 = 0.80), observers rated the photo of the person as less attractive (J_1 = -0.68), whereas in the case of low attractiveness (K_2 = -1.65), the photo was rated as more attractive (J_2 = 0.007). In both cases, the slope β was comprised between 0 and -1 (-0.27 and -0.28, respectively).

Switching to the dispersion effect, anchoring[6] provides a well-known example of such a cognitive mechanism. For example, Tversky and Kahneman (1974) described an experiment

[6]Anchoring occurs when individuals use a piece of information to make subsequent judgments. Given an anchor value, subsequent judgments are biased toward such value (Tversky & Kahneman, 1974).

Study	Context K_i	Judgment J_i	β
Parducci (1968)	1.97 4.06	3.21 2.69	-.25
Leding et al. (2015) Exp 1	-1.24 0.88	-0.30 -0.87	-.27
Leding et al. (2015) Exp 2	-1.65 0.80	0.007 -0.68	-.28
Tversky & Kahneman (1974)	10 65	25 45	.36

Table 2: Contextual and stimulus values for three studies (Leding et al., 2015; Parducci, 1968; Tversky & Kahneman, 1974) and corresponding slopes. It could be noted that slopes are always constrained between -1 and 1.

in which participants were asked to estimate the percentage of African countries in the United Nations. Before answering, they were requested to state whether the estimate was higher than the low anchor $K_1 = 10\%$ or lower than the high anchor $K_2 = 65\%$. Results indicated a strong dispersion effect; in fact, they found a value of $J_1 = 25\%$ for the low anchor and $J_2 = 45\%$ for the high anchor. Taking into account these points, in line with our predictions, the slope β is equal to .36. We also evaluated a more recent study by Mussweiler and Strack (2000) based on the same kind of task (not reported in Table 2). This study investigated the different magnitudes of the dispersion due to anchoring as a function of previous knowledge and perceived plausibility of the anchors. The slopes β of the reported data are always bounded between 0 and 1 (minimum 0.04 - maximum 0.73).

All the previous cases are based on the unverified assumption of the presence of a linear relationship because the slopes are computed by using only two points. In order to overcome such a limitation, we performed an experiment aimed at evaluating (i) a contrast effect in a manner similar to that used by Parducci (1968), and (ii) a dispersion effect based on anchoring, following Tversky and Kahneman (1974). However, unlike the previous examples, we took into account several values of the context in order to check the presence of an actual linear relationship. Given the validity of such an assumption, it is possible to check if the slope of the linear relation is constrained according to the variational framework. We will also describe in more detail Matthews and Stewart's (2009b) paper. This is a relatively recent paper that reported both dispersion and contrast effects for non-physical judgments taking into account a high number of different values for contextual information. The authors reported the slopes of such effects for the whole sample but also for each participant. Importantly, in this work contrast and dispersion effects are measured without relying on a Likert scale. For these reasons, it represents a good example of a case study for the variational framework predictions.

6.1.1. Experiment

In order to verify the prediction of our variational framework, we assumed a linear relationship between context and judgment. The studies on context effects taken into account in the previous section evaluated such effects for only two different levels of contextual informa-

tion. Thus, we performed an experiment aimed at verifying the linearity assumption taking into account several contextual values. We employed two classical experimental paradigms previously described: The Parducci (1968) paradigm about the moral evaluation of sentences, and the anchoring effect of Tversky and Kahneman (1974), inducing contrast and dispersion effects, respectively. Therefore, we considered two scenarios: the moral evaluation scenario in which people had to rate the moral seriousness of different actions, and the Kant's date of birth scenario in which people had to guess when Immanuel Kant was born with different anchors. In both cases, we expected to find a linear relationship. If the linearity assumption is met, according to the constraints of the proposed framework, we predict a slope between -1 and 0 for the contrast effect and a slope between 0 and 1 for the dispersion effect.

Participants. Three hundred and ninety subjects (285 female) participated voluntarily. Their mean age was 20.19 (sd = 3.91), with a range of 18 to 54 years. Participants were recruited in classroom settings in the University of Florence. The experimental task was part of other tasks connected to unrelated experiments. In each scenario (moral evaluation and Kant's date of birth), participants were randomly assigned to a single context condition (out of seven, including no context).

Materials. In the moral evaluation scenario, participants were asked to evaluate (on a Likert scale) the morality of one target sentence. Such a task could be preceded by the evaluation of a contextual sentence or it could be performed in isolation. The target sentence was *Borrowing a small amount of money from friends without repaying*. The six possible contextual sentences were: *Exceeding the speed limit by about 10 km/h in safe conditions when driving on a highway*; *Stealing a towel from a hotel*; *Bringing a dog on the beach where it is forbidden*; *Selling spoiled milk to a hospital*; *Giving false testimony at a criminal trial in exchange for a sum of money*; and *Murdering a relative*. Every sentence shown had to be evaluated on a scale from 1 (absolutely trivial) to ten (absolutely serious). In the Kant's date of birth scenario, participants were asked to estimate the year of birth of the philosopher Immanuel Kant. Such a task could be preceded by an anchor question or it could be performed in isolation. The anchor question was 'Was Immanuel Kant born before or after *anchor year*?' Anchor years were: 1574, 1624, 1674, 1774, 1824, and 1874.

Results. With regard to the moral evaluation scenario, the mean judgment given for the target sentence in isolation was 8.05 (sd = 1.99). In Table 3 the mean judgments given to each contextual sentences with the corresponding judgments given to the target sentence are reported. The scatterplot (Figure 4) shows a rather weak correlation between context and judgments ($r = -.17$, $p = .002$) with possibly six outlier points corresponding to judgments of 4 or less for the target sentence. Excluding such points, the Pearson's correlation becomes weaker but it is still statistically significant ($r = -.14$, $p = .014$); in this case, the regression equation is $y = 8.66 - 0.07x$.

The residual plot for the moral evaluation scenario (Figure 5) was patternless and it showed no clear trend, suggesting a linear relationship between the context and moral judgments. By us-

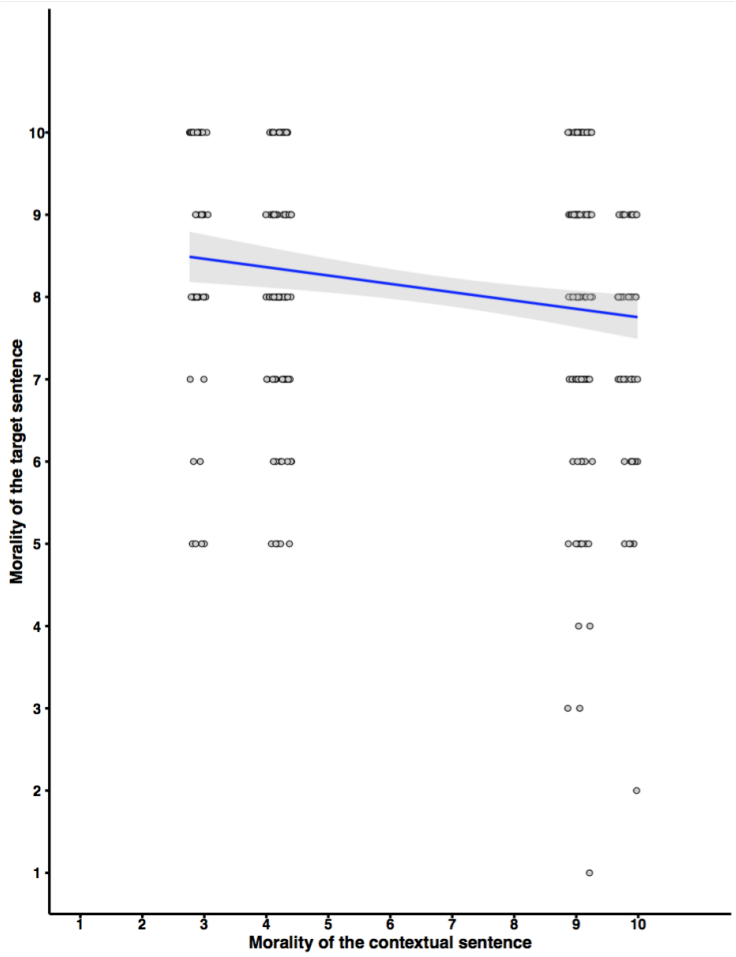

Figure 4: Morality of the target sentence as a function of the morality of the contextual sentence. Noise was added for better visibility. Regression equation is $y = 8.77 - 0.1x$ (blue line) where the shadowed area corresponds to the confidence interval.

ing an ANOVA to verify the linearity between the context and the judgments, the linear component was statistically significant ($\mathcal{F}_{(1,325)} = 9.67$, $p = .002$), whereas the deviation from linearity component was not ($\mathcal{F}_{(4,325)} = 1.02$, $p = .365$). The linear relationship between the context and the moral judgments was confirmed by means of a Ramsey RESET test (Ramsey, 1969) against a quadratic ($\text{RESET}_{(1,328)} = 0.62$, $p = .431$), cubic ($\text{RESET}_{(1,328)} = 1.11$, $p = .292$) and quartic ($\text{RESET}_{(1,328)} = 1.67$, $p = .198$) power effect as well as the square root ($\text{RESET}_{(1,328)} = 0.23$, $p = .626$).

With regard to the Kant scenario, the mean estimated year in isolation was 1768 (sd = 81). Table 4 shows the relationship between anchor years and the mean estimated year of Kant's birth. The scatterplot (Figure 6) shows a positive and statistically significant relationship ($r = .46$, $p < .001$) between the anchor and the estimated values. Also for the Kant scenario, the residual plot (Figure 7) was patternless, suggesting a linear relationship between the anchor and the estimated year of birth. Indeed, the linear component of the ANOVA was statistically significant ($\mathcal{F}_{(1,294)}$

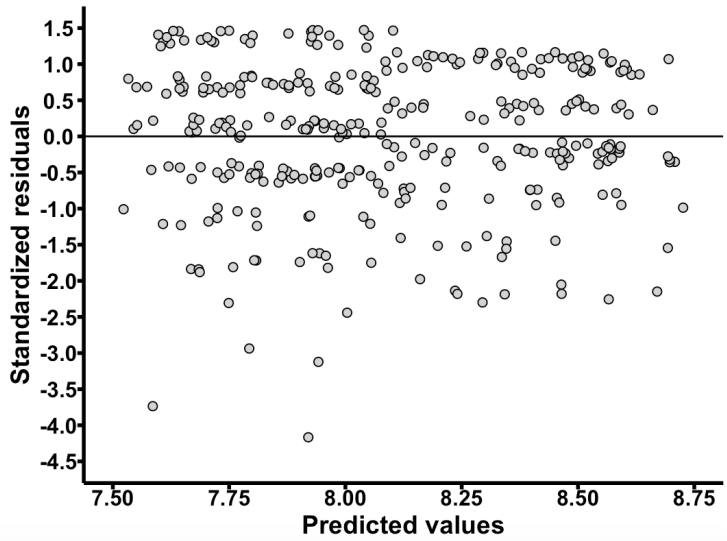

Figure 5: The residual plot shows the relation between standardized residuals and predicted values for the moral evaluation scenario. Noise was added for better visibility. Other than a few outliers, residuals are randomly distributed around the horizontal line corresponding to zero mean.

= 77.48, $p < .001$) whereas the deviation from linearity was not ($\mathcal{F}_{(4,294)} = 0.33$, $p = .857$). The Ramsey RESET test was not statistically significant for the quadratic ($\text{RESET}_{(1,297)} = 0.82$, $p = .365$), cubic ($\text{RESET}_{(1,297)} = 0.80$, $p = .371$) and quartic ($\text{RESET}_{(1,297)} = 0.78$, $p = .377$) power effect or as a square root function ($\text{RESET}_{(1,297)} = 0.85$, $p = .357$).

For both the moral and the Kant scenarios, the relationship between context values and judgments was linear and constrained between 0 and 1 or -1 and 0 for contrast and dispersion effects, respectively, in line with the constraints predicted by the proposed variational framework. It must be underlined that only in the moral scenario were data collected by means of a Likert scale (and thus with the consequent limitations associated with the use of an ordinal scale).

6.1.2. The Matthews & Stewart (2009b) study

The Matthews and Stewart (2009b) paper is focused on how the context influences the judgment of non-physical dimensions. In particular, they investigated judgments of objects' prices in a sequential decision making paradigm. Within the study of basic perceptual properties, as we have seen before, it has been observed that psychophysical judgments are strongly influenced by the local context: In particular, in sequential decision making tasks, the stimuli and the responses given in previous trials may influence the response of the current trial. Matthews and Stewart (2009b) cite Jesteadt, Luce, and Green (1977) who propose a linear regression model to account for context effects that arise in sequential decisional making experiments, based on the following equation:

$$J_n = \gamma + \alpha_0 P_n + \alpha_1 P_{n-1} + \beta_1 J_{n-1} + \epsilon, \tag{45}$$

where J_n is the current judgment at trial n, which is a function of P_n, P_{n-1}, the objective magnitude of the stimuli presented in the current trial n and in the previous trial $n-1$, respectively,

Context			Target Sentence	
Sentence	**Mean**	**sd**	**Mean**	**sd**
Exceeding the speed limit by about 10 km/h in safe conditions when driving on a highway	2.92	1.66	8.49	1.52
Stealing a towel from a hotel	4.15	2.02	8.38	1.40
Bringing a dog on the beach where it is forbidden	4.25	2.26	8.24	1.52
Giving false testimony at a criminal trial in exchange for a sum of money	9.02	1.23	8.07	1.73
Selling spoiled milk to a hospital	9.10	1.55	7.97	1.82
Murdering a relative	9.83	0.68	7.41	1.45

Table 3: Results of the moral evaluation experiment. There is a contrast effect between the judgments given to the contextual sentences and the corresponding judgments related to the target sentence.

Anchor	Estimated year	
	Mean	**sd**
1574	1704	92
1624	1713	90
1674	1720	76
1774	1763	46
1824	1784	69
1874	1802	69

Table 4: Results of the Kant's date of birth experiment. There is a dispersion effect between the years that play the role of the anchor and the estimation of the year in which Kant was born.

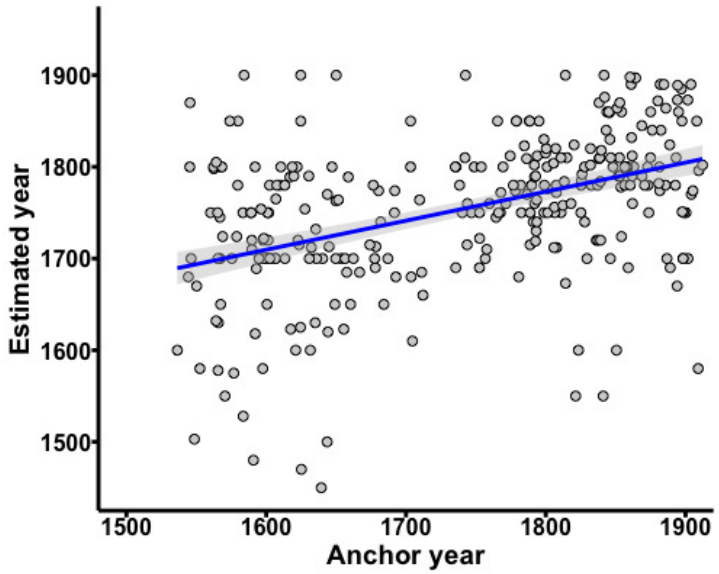

Figure 6: Estimated year of Kant's birth as a function of the anchor year. Noise was added for better visibility. Regression equation is $y = 1164 + 0.33x$ (blue line) where the shadowed area corresponds to the confidence interval.

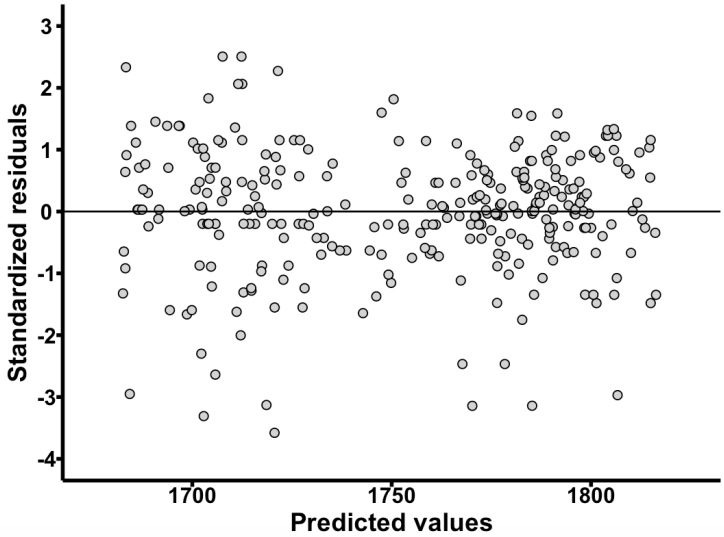

Figure 7: The residual plot shows the relation between standardized residuals and predicted values for the Kant scenario. Noise was added for better visibility. Other than a few outliers, residuals are randomly distributed around the horizontal line corresponding to zero mean.

and of J_{n-1}, which is the judgment relative to the trial $n-1$. Apart from the error ϵ, the other constants appearing in the previous formula are the regression coefficients relative to the current and previous objective stimulus (α_0 and α_1), and to the previous judgment (β_1). The key observation is that there is a contrast effect between J_n and P_{n-1} expressed by α_1 and, *at the same time*, there is a dispersion effect between J_n and J_{n-1} expressed by β_1 (DeCarlo & Cross, 1990;

	Experiment 1	**Experiment 2a**	**Experiment 2b**
Total participants	25	28	28
Positive values	23	26	14
Significant positive values	9	12	5
Maximum and minimum of positive values	0.032, 0.475	0.027, 0.465	0.052, 0.311
Mean	0.163	0.162	0.046

Table 5: Dispersion effects in the Matthews & Stewart (2009b) study. For each experiment, the table reports the total number of participants, how many unstandardized β_1 were positive, positive and significant, the maximum and minimum values, and the mean across the whole sample. Experiment 1 did not include particular manipulations, whereas experiment 2 provided two conditions: (a) no feedback and fixed time of presentation (b) feedback after each judgment and a fixed presentation time of each item.

	Experiment 1	**Experiment 2a**	**Experiment 2b**
Total participants	25	28	28
Negative values	19	17	11
Significant negative values	2	4	2
Maximum and minimum of negative values	-0.017, -0.188	-0.014, -0.278	-0.008, -0.226
Mean	-0.034	-0.049	0.058

Table 6: Contrast effects in the Matthews & Stewart (2009b) study. For each experiment, the table reports the total number of participants, how many unstandardized α_1 were negative, negative and significant, the maximum and minimum values, and the mean across the whole sample. Experiment 1 did not include particular manipulations, whereas experiment 2 provided two conditions: (a) no feedback and fixed time of presentation (b) feedback after each judgment and a fixed presentation time of each item.

Jesteadt et al., 1977; Matthews & Stewart, 2009a; Mori, 1998).

Matthews and Stewart (2009b) investigated such sequential effects in the domain of non-physical dimensions asking for judgments related to the prices of various items (such as chairs and shoes). In the first experiment, participants rated the price of a set of 100 chairs where each item was shown until a response was given. They then employed the regression model of Jesteadt et al. (1977) to evaluate the influence of context on the judgment of price (after a log transformation of each variable). The authors performed three experiments: A preliminary study with free presentation time and no particular manipulation, and a second experiment with two conditions. The first one was without feedback and with the presentation of each item for a fixed amount of time, whereas in the second condition, there was feedback given after each judgment and a fixed presentation time of each item. In Table 5, we report some statistics related to the estimation of the unstandardized coefficient β_1 obtained in the three experiments. Crucially for our framework, all coefficients related to the dispersion effect were constrained between 0 and 1. With regard to the contrast effect, Table 6 summarized the findings related to the unstandardized coefficients α_1. Although in experiment 2b contrast effect was not observed, in line with our predictions, no coefficient lower than -1 was observed. Therefore, the empirical cases reported testify to the overall accuracy of the variational framework predictions. Naturally, future studies are needed to further empirically validate the predictions of the variational framework proposed in this paper.

7. Discussion and conclusions

Starting with the variational interpretation of brightness matching, we have proposed a general variational framework for context effects that may hold both for physical and non-physical judgments. Our variational interpretation indicates that contrast and dispersion effects may occur simultaneously in every judgment, the final percept being the balance between the opposing actions of contrast (the tendency to maximize the differences among the dimensions that characterize the object to be judged) and dispersion (the tendency to minimize the differences among the dimensions).

In the first part of the paper, we showed how the variational interpretation of brightness matching represents an alternative version of Rudd and Zemach's model (Rudd & Zemach, 2004) that makes the same predictions but has the added advantage of being able to explain brightness matching in terms of a balance between contrast and dispersion, and also able to be directly generalizable to physical configurations that are much more complex than those considered in Rudd and Zemach's experiments. Within perception studies, there are many different quantitative models and theoretical approaches to context effects (Gilchrist, 2006). We have shown that our framework can describe existing theories through the specification of suitable functional terms and parameters. We expect that future works will apply the proposed variational techniques to derive new constraints and, hopefully, disambiguate different theoretical proposals.

In the second part of the paper, we advocated the plausibility of the proposed framework for non-physical judgments as well, such as those observed in social cognition or cognitive psychology. Several case studies drawn from the existing literature, along with new empirical data, are in line with the constraints set by our variational approach. In particular, we performed an experiment aimed at observing classic dispersion and contrast effects. Crucially, the relationship between context and judgments was in line with the predictions of the variational model. Although this evidence tend to support our framework, it is necessary to stress some of its limitations. First, since non-physical judgments are usually collected by means of a Likert scale, the applicability of our variational approach to ordinal data it is not straightforward (Camparo, 2013; Carifio & Perla, 2007; Wegener, 1982). Thus, this issue requires a systematic investigation in order to understand the extent to which a variational solution may be applied to Likert scale data. In any case, we deem it important to stress that the study of Matthews and Stewart (2009b) on context effects in non-physical judgments supports our proposal without using a Likert scale. Second, in order to further confirm our view, a complete meta-analysis is needed, along with a direct empirical investigation into the predictions related to the constraints of the proposed model. If our variational approach will prove its validity in non-physical judgments domains also, then this work may represent a first step toward a formal analysis of psychological mechanisms able to account for perceptual and cognitive phenomena with similar characteristics. The Kahneman and Tversky research program began as the study of the statistical intuitions of experts (Tversky & Kahneman, 1971) where intuitive judgments were considered as extensions of perception to judgments about objects that are not currently present (Kahneman & Frederick, 2005). Since then, psychological literature often draws a parallel between intuitive non-physical judgments

and the mechanisms of perception (Kahneman, 2011; Thaler & Sunstein, 2008). However, this parallelism is sometimes superficially stated and not fully investigated, especially from a formal point of view. Our work tries to fill this gap, representing an attempt to use a common framework to account for context-driven effects both in perception and cognition. Moreover, we hope that our proposal will stimulate the mathematical modeling of context-driven effects in fields such as social cognition where these kinds of approaches are seldom employed (Biernat, 2005). Given that our analysis emphasizes the concurrent presence of contrast and dispersion effects in every judgment, the proposed variational framework supports the theoretical accounts that hypothesize the simultaneous occurrence of the two effects at the expense of sequential models (Biernat, 2005). Future research about context-driven effects in social cognition and cognitive psychology could also use our framework to devise new experiments and disambiguate different theoretical positions.

In general terms, the proposed variational framework does not address the question about under which conditions we observe either a prevalence of contrast or dispersion (both within perception or cognition fields). The main aim of our approach is not to solve these kinds of open problems, rather to help understand in a more profound way the interpretation of observations in psychology. We consider this to be as important as making new predictions. In fact, being able to predict a phenomenon does not necessarily mean understanding it, as the renowned example of Ptolemaic orbital system shows: planetary orbits, as predicted by Ptolemy, were accurate, but his geocentric interpretation of planetary movement was totally wrong, showing that description and comprehension can be separated in a scientific theory. One of the powerful advantages of a variational approach is that it allows for conjugating description and comprehension simultaneously. Consequently, we hope that this work (and especially the included tutorial) will also stimulate further application of variational methods in mathematical psychology beyond the issue of context-driven effects.

8. Acknowledgements

We wish to thank the two anonymous reviewers and the editor for their precious suggestions that greatly improved the manuscript. We are also grateful to Daniele Lari, Claudia Solfanelli and Giorgio Pupi for helping us with the data collection.

References

Agostini, T. & Galmonte, A. (2000). Contrast and assimilation: the belongingness paradox. *Review of Psychology*, *7*(1-2), 3–7.

Agostini, T. & Galmonte, A. (2002). Perceptual organization overcomes the effect of local sorrund in determining simultaneous lightness contrast. *Psychological Science*, *13*(1), 88–92.

Agostini, T. & Profitt, D. R. (1993). Perceptual organization evokes simultaneous lightness contrast. *Perception*, *22*(3), 263–272.

Bertalmío, M., Caselles, V., & Provenzi, E. (2009). Issues about the retinex theory and contrast enhancement. *International Journal of Computer Vision*, *83*, 101–119.

Bertalmío, M., Caselles, V., Provenzi, E., & Rizzi, A. (2007). Perceptual color correction through variational techniques. *IEEE Trans. on Image Processing*, *16*, 1058–1072.

Biernat, M. (2005). *Standard and expectancies: Contrast and assimilation in judgments of self and others.* NY: Psychology Press.

Biernat, M. & Kobrynowicz, D. (1997). Gender- and race-based standards of competence: Lower minimum standards but higher ability standards for devalued groups. *Journal of Personality and Social Psychology*, *72*(3), 544–557.

Blakeslee, B. & McCourt, M. E. (2004). A unified theory of brightness contrast and assimilation incorporating oriented multiscale spatial filtering and contrast normalization. *Vision Research*, *44*(21), 2483–2503.

Boyd, S. & Vandenberghe, L. (2004). *Convex optimization*. Cambridge University Press.

Bressan, P. (2001). Explaining lightness illusions. *Perception*, *30*, 1031–1046.

Brickman, P. & Campbell, D. T. (1971). Hedonic relativism and planning the good society. In M. H. Appley (Ed.), *Adaptation-level theory* (pp. 287–302). New York: Academic Press.

Camparo, J. (2013). A geometrical approach to the ordinal data of likert scaling and attitude measurements: The density matrix in psychology. *Journal of Mathematical Psychology*, *57*(1-2), 29–42.

Carifio, J. & Perla, R. J. (2007). Ten common misunderstandings, misconceptions, persistent myths and urban legends about likert scales and likert response formats and their antidotes. *Journal of Social Sciences*, *3*(3), 143–159.

Casacci, S. & Pareto, A. (2015). Methods for quantifying ordinal variables: a comparative study. *Quality & Quantity*, *49*(5), 1859–1872.

Chevreul, M. E. (1855/2007). *The Principles of Harmony and Contrast of Colours, and Their Applications to the Arts* (2 ed.). London: Longman, Brown, Green, and Longmans.

De Weert, C. M. M. & Spillman, L. (1995). Assimilation: Asymmetry between brightness and darkness. *Vision Research*, *35*, 1413–1420.

DeCarlo, L. T. & Cross, D. V. (1990). Sequential effects in magnitude scaling: Models and theory. *Journal of Experimental Psychology: General*, *119*(4), 375–396.

Diener, E., Lucas, R. E., & Scollon, C. N. (2006). Beyond the hedonic treadmill: revising the adaptation theory of well-being. *American Psychologist*, *61*(4), 305–314.

Ehm, W. & Wackermann, J. (2012). Modeling geometric-optical illusions: A variational approach. *Journal of Mathematical Psychology*, *56*(6), 404–416.

Evans, R. (1948). *An introduction to color*. Wiley.

Eysenck, M. W. (1990). *Happiness*. Hove, UK: Laurence Erlbaum Associates Ltd.

Fairchild, M. (2005). *Color appearance models*. Hoboken, NJ.: Wiley.

Fechner, G. (1860). *Elemente der Psychophysik I u. II*. Leipzig: Breitkopf und Hartel.

Festinger, L. (1954). A theory of social comparison processes. *Human Relations*, *7*, 117–140.

Festinger, L., Coren, S., & Rivers, G. (1970). The effect of attention on brightness contrast and assimilation. *American Journal of Psychology*, *83*, 189–207.

Frederick, S. (2007). Hedonic treadmill. In R. Baumesteir & K. D. Vohs (Eds.), *Encyclopedia of Social Psychology* (pp. 419–420). London: Sage.

Fujita, F. & Diener, E. (2005). Life satisfaction set point: Stability and change. *Journal of Personality and Social Psychology*, *88*(1), 158–164.

Gelfand, I. & Fomin, S. (1963). *Calculus of variations*. Prentice Hall: Englewood Cliffs, NJ.

Gilchrist, A. (2006). *Seeing black and white*. Oxford University Press.

Gilchrist, A. (2015). Perceptual organization in lightness. In J. Wagemans (Ed.), *Oxford handbook of perceptual organization* (pp. 391–412). Oxford: Oxford University Press.

Heintz, R. K. (1950). The effect of remote anchoring upon the judgment of lifted weights. *Journal of Experimental Psychology*, *40*(5), 584–591.

Helson, H. (1963). Studies of anomalous contrast and assimilation. *Journal of the Optical Society of America*, *53*, 179–184.

Higgins, E. T. (1987). Self-discrepancy: A theory relating self and affect. *Psychological Review*, *94*(3), 319–340.

Higgins, E. T. (1990). Personality, social psychology, and person-situation relations: Standards and knowledge activation as a common language. In L. A. Pervin (Ed.), *Handbook of Personality: Theory and Research* (pp. 301–338). NY: The Guilford Press.

Higgins, E. T. & Lurie, L. (1983). Context, categorization, and memory: The 'change of standard' effect. *Cognitive Psychology*, *15*, 525–547.

Jesteadt, W., Luce, R. D., & Green, D. M. (1977). Sequential effects in judgments of loudness. *Journal of Experimental Psychology: Human Perception and Performance*, *3*(1), 92–104.

Kahneman, D. (2011). *Thinking, fast and slow*. NY: Macmillan.

Kahneman, D. & Frederick, S. (2005). A model of heuristic judgement. In *The cambridge handbook of thinking and reasoning* (pp. 267–293). Cambridge: Cambridge University Press.

Kenrick, D. T. & Gutierres, S. A. (1980). Contrast effects and judgments of physical attractiveness: When beauty becomes a social problem. *Journal of Personality and Social Psychology*, *38*(1), 131–140.

Kingdom, F. & Moulden, B. (1991). White's effect and assimilation. *Vision Research*, *31*(1), 151–159.

Koffka, K. (1935). *Principles of Gestalt Psychology*. NY: Harcourt, Brace, and World.

Krantz, D. L. & Campbell, D. T. (1961). Separating perceptual and linguistic effects of context shifts upon absolute judgments. *Journal of Experimental Psychology*, *62*(1), 35–42.

Land, E. & McCann, J. (1971). Lightness and Retinex theory. *Journal of the Optical Society of America*, *61*(1), 1–11.

Leding, J. K., Horton, J. C., & Wootan, S. S. (2015). The contrast effect with avatars. *Computers in Human Behavior*, *44*, 118–123.

Locke, J. (1690/1979). *An essay concerning human understanding*. Oxford: Oxford University Press.

Lykken, D. & Tellegen, A. (1996). Happiness is a stochastic phenomenon. *Psychological Science*, *7*(3), 186–189.

Martin, L. L. & Achee, J. W. (1992). Beyond accessibility: The role of processing objectives in judgment. In L. L. Martin & A. Tesser (Eds.), *The Construction of Social Judgments* (pp. 195–216). Hillsdale, NJ: Lawrence Erlbaum Associates.

Matthews, W. J. & Stewart, N. (2009a). The effect of interstimulus interval on sequential effects in absolute identification. *The Quarterly Journal of Experimental Psychology*, *62*(10), 2014–2029.

Matthews, W. J. & Stewart, N. (2009b). Psychophysics and the judgment of price: Judging complex objects on a non-physical dimension elicits sequential effects like those in perceptual tasks. *Judgment and Decision Making*, *4*(1), 64–81.

Melamed, M. E. (1971). Response requirements and the contrast effect in loudness judgments. *Perception & Psychophysics*, *10*(4), 309–312.

Mori, S. (1998). Effects of stimulus information and number of stimuli on sequential dependencies in absolute identification. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *52*(2), 72–83.

Moskowitz, G. B. (2005). *Social Cognition: Understanding Self and Others*. Guildford Press.

Moskowitz, G. B. & Roman, R. J. (1992). Spontaneous trait inferences as self-generated primes: Implications for conscious social judgement. *Journal of Personality and Social Psychology*, *62*, 728–738.

Mussweiler, T. (2003). Comparison processes in social judgment: Mechanisms and consequences. *Psychological Review*, *110*, 472–489.

Mussweiler, T. & Strack, F. (2000). The use of category and exemplar knowledge in the solution of anchoring tasks. *Journal of Personality and Social Psychology*, *78*(6), 1038.

Nisbett, R. E. & Wilson, T. D. (1977). The halo effect: Evidence for unconscious alteration of judgments. *Journal of Personality and Social Psychology*, *35*(4), 250–256.

Noventa, S. & Vidotto, G. (2012). A variational approach to behavioral and neuroelectrical laws. *Biological Cybernetics*, *106*, 339–358.

Palma-Amestoy, R., Provenzi, E., Bertalmío, M., & Caselles, V. (2009). A perceptually inspired variational framework for color enhancement. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *31*(3), 458–474.

Parducci, A. (1968). The relativism of absolute judgments. *Scientific American*, *219*(6), 84–90.

Parker, S., Bascom, J., Rabinovitz, B., & Zellner, D. (2008). Positive and negative hedonic contrast with musical stimuli. *Psychology of Aesthetics, Creativiy and the Arts*, *2*(3), 171–174.

Pepitone, A. & DiNubile, M. (1976). Contrast effects in judgments of crime severity and the punishment of criminal violators. *Journal of Personality and Social Psychology*, *33*(4), 448–459.

Petty, R. E. & Wegener, D. T. (1993). Flexible correction processes in social judgment: Correcting for context-induced contrast. *Journal of Experimental Social Psychology*, *29*, 137–165.

Provenzi, E. (2013). Achromatic induction: A variational interpretation of Rudd-Zemach's Edge Integration Model. In *Signal-Image Technology & Internet-Based Systems (SITIS), 2013 International Conference*, (pp. 424–429). IEEE.

Provenzi, E. & Caselles, V. (2014). A wavelet perspective on variational perceptually-inspired color enhancement. *International Journal of Computer Vision*, *106*, 153–171.

Provenzi, E., De Carli, L., Rizzi, A., & Marini, D. (2005). Mathematical definition and analysis of the retinex algorithm. *Journal of the Optical Society of America A*, *22*(12), 2613–2621.

Provenzi, E., Fierro, M., Rizzi, A., De Carli, L., Gadia, D., & Marini, D. (2007). Random spray retinex: A new retinex implementation to investigate the local properties of the model. *IEEE Transactions on Image Processing*, *16*, 162–171.

Provenzi, E., Gatta, C., Fierro, M., & Rizzi, A. (2008). Spatially variant white patch and gray world method for color image enhancement driven by local contrast. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *30*, 1757–1770.

Ramsey, J. B. (1969). Tests for specification errors in classical linear least squares regression analysis. *Journal of the Royal Statistical Society Series B*, *31*(2), 350–371.

Rudd, M. & Zemach, I. (2004). Quantitive properties of achromatic color induction: An edge integration analysis. *Vision Research*, *44*, 971–981.

Sapiro, G. & Caselles, V. (1997). Histogram modification via differential equations. *Journal of Differential Equations*, *135*, 238–266.

Schkade, D. A. & Kahneman, D. (1998). Does living in California make people happy? A focusing illusion in judgments of life satisfaction. *Psychological Science*, *9*(5), 340–346.

Schwarz, N. & Bless, H. (1992). Constructing reality and its alternatives: An inclusion/exclusion model of assimilation and contrast effects in social judgments. In L. L. Martin & A. Tesser (Eds.), *The Construction of Social Judgments* (pp. 217–245). Mahwah, NJ: Lawrence Erlbaum Associates.

Sherif, M., Taub, D., & Hovland, C. L. (1958). Assimilation and contrast effects of anchoring stimuli on judgments. *Journal of Experimental Psychology*, *55*(2), 150–155.

Silver, R. L. (1982). *Coping with an undesirable life event: A study of early reactions to physical disability*. PhD thesis, Unpublished Doctoral Dissertation, Northwestern University.

Stapel, D. A. & Koomen, W. (1998). When stereotype activation results in (counter) stereotypical judgments: Priming stereotype-relevant traits and exemplar. *Journal of Experimental Social Psychology*, *34*, 136–163.

Stapel, D. A., Koomen, W., & van der Pligt, J. (1996). The referents of trait inferences: The impact of trait concepts versus actor-trait links on subsequent judgements. *Journal of Personality and Social Psychology*, *70*, 437–450.

Stapel, D. A., Koomen, W., & van der Pligt, J. (1997). Categories of category accessibility: The impact of trait concept versus exemplar priming on person judgments. *Journal of Experimental Social Psychology*, *33*(1), 47–76.

Strack, F., Schwarz, N., & Gschneidinger, E. (1985). Happyness and reminiscing: The role of time perspective, affect, and mode of thinking. *Journal of Personality and Social Psychology*, *49*, 1460–1469.

Takane, Y. (2005). Optimal scaling. In B. Everitt & D. Howell (Eds.), *Encyclopedia of Statistics for Behavioral Sciences* (pp. 1479–1482). Chichester: Wiley.

Thaler, R. H. & Sunstein, C. R. (2008). *Nudge: Improving Decisions about Health, Wealth, and Happiness*. New Haven and London: Yale University Press.

Tversky, A. & Kahneman, D. (1971). Belief in the law of small numbers. *Psychological Bulletin*, *76*(2), 105–110.

Tversky, A. & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*(4157), 1124–1131.

von Bezold, W. (1876/2014). *The Theory of Color in Its Relation to Art and Art-Industry*. Charleston, SC: Nabu Press.

Wallach, H. (1948). Brightness constancy and the nature of achromatic colors. *Journal of Experimental Psychology*, *38*(3), 310–324.

Wallach, H. (1963). The perception of neutral colors. *Scientific American*, *208*, 107–116.

Weber, E. (1846). Der tastsinn und das gemeingefuhl. In R. Wagner (Ed.), *Handworterbuch der Physiologie*. Braunschweig.

Wedell, D. H., Parducci, A., & Geiselman, R. E. (1987). A formal analysis of rating of physical attractiveness: Successive contrast and simultaneous assimilation. *Journal of Experimental Social Psychology*, *23*, 230–249.

Wegener, B. (Ed.). (1982). *Social attitudes and psychophysical measurement*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Wertheimer, M. (1923). Untersuchungen zur lehre von der gestalt. *Psychologische Forschung*, *4*, 301–350.

White, J. D. & Vilmain, J. A. (1986). The psychophysics of price: A critique of the weber-fechner approach in consumer behavior. In N. K. Malhotra (Ed.), *Proceedings of the 1986 Academy of Marketing Science (AMS) Annual Conference* (pp. 30–34). Springer International Publishing.

White, M. (1979). A new effect of the nature on perceived lightness. *Perception*, *8*, 413–416.

Wildeman, C., Turney, K., & Schnittker (2014). Hedonic consequences of punishment revisited. *Journal of Criminal Law and Criminology*, *104*, 133–164.

Wundt, W. M. (1896). *Grundriss der Psychologie*. Leipzig: Wilhelm engelmann.

Young, F. W. (1975). Methods for describing ordinal data with cardinal models. *Journal of Mathematical Psychology*, *12*(4), 416–436.

Appendix: Proof of Proposition 4.1

We will perform the proof only in the continuous case, since the one in the discrete scenario is analogous. Let us compute the variation of the functional $E_w(\log\ L) = A(\log\ L) - C_w(\log\ L)$ with respect to $\log\ L$. To do this, it is convenient to write $\log\ L(x) \equiv \tilde{L}(x)$. The variation of the first term of the functional is trivial to compute and gives:

$$\delta A(\tilde{L}, \tilde{J}) = \int_\Omega (\tilde{L}(x) - \tilde{\mu})\tilde{J}(x)\, dx \tag{46}$$

$\tilde{J}(x)$ being a generic functional perturbation of $\tilde{L}(x)$. The variation of the second term instead can be written as follows

$$\begin{aligned}\delta C_w(\tilde{L}, \tilde{J}) = &\frac{1}{2}\iint_{\Omega^2} w(\|x-y\|)\left[\tilde{L}(x) - \tilde{L}(y)\right]\tilde{J}(x)\, dxdy \\ &- \frac{1}{2}\iint_{\Omega^2} w(\|x-y\|)\left[\tilde{L}(y) - \tilde{L}(x)\right]\tilde{J}(y)\, dxdy.\end{aligned} \tag{47}$$

Now, interchanging the role of the 'mute' variables x and y in the second integral, and thanks to the symmetry of the induction weight, i.e., $w(\|x-y\|) = w(\|y-x\|)$, we can write

$$\delta C_w(\tilde{L}, \tilde{J}) = \iint_{\Omega^2} w(\|x-y\|)\left[\tilde{L}(x) - \tilde{L}(y)\right]\tilde{J}(x)\, dxdy. \tag{48}$$

Since $\delta E(\tilde{L}, \tilde{J}) = \delta A(\tilde{L}, \tilde{J}) - \delta C_w(\tilde{L}, \tilde{J})$, thanks to Fubini's theorem, we have

$$\delta E_w(\tilde{L}, \tilde{J}) = \int_\Omega \left(\tilde{L}(x) - \int_\Omega w(\|x-y\|)\left[\tilde{L}(x) - \tilde{L}(y)\right] dy\right)\tilde{J}(x)\, dx, \tag{49}$$

The argmin of $E_w(\tilde{L}, \tilde{J})$ satisfies the Euler-Lagrange equation $\delta E_w(\tilde{L}, \tilde{J}) = 0$ for all perturbations $\tilde{J}$. Thanks to the fundamental lemma of variational calculus (Gelfand & Fomin, 1963), this is possible if and only if $\tilde{L}(x)$ satisfies

$$\tilde{L}(x) - \tilde{\mu} - \int_\Omega w(\|x-y\|)\left[\tilde{L}(x) - \tilde{L}(y)\right] dy = 0, \tag{50}$$

i.e.,

$$\tilde{L}(x) = \tilde{\mu} + \int_\Omega w(\|x-y\|)\left[\tilde{L}(x) - \tilde{L}(y)\right] dy. \tag{51}$$

Turning back to the original variables, we have

$$\log\ L(x) = \mu + \int_\Omega w(\|x-y\|)\log\frac{L(x)}{L(y)}\, dy, \tag{52}$$

so we see that the argmin of E_w satisfies eq. (13) that defines the logarithmic brightness, and thus proposition 4.1 is proven. □