THEORIES OF CONTEXTUAL JUDGEMENT IN RELATION TO WELL-BEING AND OTHER OUTCOMES

RUFUS POLLOCK

UNIVERSITY OF CAMBRIDGE

DECEMBER 9, 2008

ABSTRACT. We present an overview of existing theories on contextual judgement/utility situating them within a general framework. We consider the extent to which they are empirically distinguishable and apply them to some well-known economic questions such as the relation of happiness and the income distribution, status races and quitting.

Keywords: Happiness, Well-Being, Status, Relative Utility

JEL Classification: A10, D01

1. INTRODUCTION

Theories of 'contextual judgement' first arose in psychology to explain evaluations of physical stimuli (weights, sizes, numbers etc). It was found that the value assigned to a particular stimulus depended, not simply on its own attributes, but also on those of the other stimuli in the 'context set'.[1]

It is is natural to extend this idea of 'contextual judgement' to economic and/or hedonic quantities such as income and job satisfaction as well as to overall well-being (utility). In these cases, 'contextual' means that an individual's evaluation of, say, their income will be determined, in part, by the incomes of those around them.

From this brief description, it should already be clear that theories of contextual judgement share similarities with those of interdependent preferences (Sobel, 2005) as well as a close connection with models of relative utility (Frank, 1985; Bordley, 1986; Frank, 1989; Clark and Oswald, 1998; Hopkins and Kornienko, 2004, 2007). We shall return to these connections in greater detail below.

Corresponding author: Rufus Pollock, Faculty of Economics and Emmanuel College, University of Cambridge. Email: rp240@cam.ac.uk or rufus@rufuspollock.org.

[1]Parducci (1965); Parducci and Perrett (1971)

The paper is organized as follows. In the first section, we introduce some standard terminology and definitions. We follow this, in the second section, with some concrete examples of situations where a contextual approach is valuable. The third section then introduces the theory itself. Here, in addition to providing a general discussion we outline the main contextual theories in use. With the theory in place, in the fourth section we discuss briefly its implications for empirical work, particularly the ways in which we can test for the existence of contextual effects and compare different theories. Last we apply the models to some specific cases such as the relationship of reported happiness and income.

2. Terminology and Definitions

There is some outcome variable x, for example income, consumption, job-satisfaction etc (more examples and details are presented in the next section – if it would help to have a concrete example in mind the reader may wish to look ahead briefly now).

For simplicity, we take x to be scalar valued and lie on the real line but the following can be adapted to more general cases without difficulty. Take a particular agent i and let x_i be the associated outcome. Consider a context set C consisting of other outcomes of the same type but achieved by 'others' ('others' may either be other entities or the same entity at a different point in time).[2]

It will often be convenient to convert the context set to a probability distribution in the following manner. First, recall that C contains members which are real-valued and that C is discrete and finite. We may therefore define the probability assigned to an outcome as the number of times it occurs divided by the total number of outcomes. This gives us a discrete probability distribution. It is useful to go beyond this to consider the case where C is non-finite (e.g. continuous) as this is mathematically more tractable and will usually represent a good approximation if C is large. In this case C may be taken as represented by some density function on the real line. With little loss of generality for the problems we are considering we may assume that this density is normalizable and corresponds to some Lesbegue measurable function. Thus, corresponding to any context set C we have a

[2]It is notationally convenient to include in the context set C the outcome being evaluated. In the case where the context set is large this will clearly have little impact but when the context set is small this might no longer be so.

cumulative distribution function F – and given any F we can interpret it as representing some context set C. Thus, in what follows we shall often use C and F interchangeably. With this equivalence in place we may now define:

Definition 1. The judgement of x_i, given context set C, is provided by the real-valued judgement function: $J = J(x, C)$. Formally a judgement function is a mapping:

$$J : \mathbf{R} \times \mathbf{F} \to \mathbf{R}$$

Where $\mathbf{F}$ is the set of probability distributions on $\mathbf{R}$.

In reality the actual judgement (J) is not available to others as it is a subjective experience. Rather the judgement is presented to the outside world in the form of some explicit category (verbal or otherwise), or, implicitly, in its role of informing some choice process. The question of how observed categorizations or choices relates to the underlying experience or judgement is clearly of some importance and we return to it below. For the time being, however, we shall assume that the relationship is known and invertible and hence we may proceed as if the judgement is itself observable.

In theory, C (and J) could vary with outcomes. However, we shall always assume that a J and C can be chosen so that they are common across outcomes.[3] In particular, for any outcome in C, the context set of this outcome is also (a subset of) C.[4] We shall refer to context sets which have this property as coherent. This permits us to consider average judgement:

Definition 2. The average or total judgement of a set of outcomes which form a coherent context set is (recall that F is the density function associated to C):

$$J(C) = \bar{J}(C) = \int J(x, F)dF(x)$$

In some circumstances the focus is on the evaluation of a set of experiences (i.e. the context set itself) and we do not know the evaluation of individual 'stimuli' within that

[3] Formally this means J may be decomposed into a function: $H : F \to G$ where G in turn is the set of functions: $\mathbf{R} \to \mathbf{R}$ so that $J(x, C) = J(x, F) = H(F)(x)$.

[4] The subset point is important. For example, in the case where the outcomes are temporally ordered, say the incomes of a given individual over time, the evaluation of any outcome may have as its (relevant) context set only those outcomes which occurred previously.

whole. In that case $J(C)$ may be defined directly as a function of C rather than being defined by some form of averaging.

As it stands the definition of a contextual judgement allows for a situation in which the context does not matter, i.e. $J(x, C) = g(x)$ for some function g. Such judgement functions will not be of great interest to us. This motivates the following definition:

Definition 3. A judgement function J is:

- 'pure self' (or self-only) if $J(x, C) = g(x)$ for some function g (and all C).
- 'contextual' if it is not 'pure self'.

Thus, in this paper, the restriction just discussed can be rephrased as the fact that we will concern ourselves only with the case of 'contextual' judgement functions. Our next step is to name some particular important restrictions on contextual judgement functions.

Definition 4. A judgement function J is:

- 'pure' contextual if a positive shift in all values has no impact on judgement: $J(x + k, C + k) = J(x, C)$ (where $C + k$ denotes the set of outcomes which have all increased by k).
- 'separable' if the judgement is linearly separable into pure self and pure contextual components: $J(x, C) = g(x) + j(x, C)$ (where j is pure contextual).
- 'rank based, or (pure) relative' if the judgement function depends only on relative (rank) position in the context set: $J(x, C) = g(R_C(x))$ for some function g and $R_C(x)$ being the rank of x within the set C (normalized to be between 0 and 1).

Remark 5. The definition of 'pure' contextual functions was based on stability under transformation. It is useful to formalize the idea a little. Suppose we have an invertible function h then hC or $h(C)$ denotes the context set under translation by h, that is $h(C)$ is the new set in which each of the elements is transformed by h. In terms of F this corresponds to a translation of the distribution function by h, that is if F is the distribution function the new distribution function is $F \cdot h^{-1}$. We shall say a judgement function, J is stable under the transformation h if, for all outcomes, x, in the context set C: $J(h(x), h(C)) = J(h(x), F \cdot h^{-1}) = J(x, C)$. With this definition, a 'pure' contextual judgement function is one that is stable under translation.

Remark 6. The 'pure' contextual case encapsulates that in which judgement is 'relative'. However, we have reserved the term relative for the case in which judgement is purely based on rank order. As we shall see below, there are 'pure' contextual judgement functions which are not 'rank based'.

Remark 7. As it will figure prominently in what follows it is useful to be precise about the rank of a variable in a set. First, supposing the set is finite, the unnormalized rank of an item is simply the position (based at zero) in which that item appears in the set when it is ordered from smallest to largest (with ties dealt with by averaging). The normalized ranking, which is more relevant for our purposes, is obtained from the representation of the context set as a distribution F. In this case, the rank $R(x)$ of a value x equals $F(x) - \frac{F(x)-F(x^-}{2}$ where $F(x^-) = lim_{x' \to_- x} F(x')$.[5] The second-term corrects for what occurs when there are mass points. Where the distribution has no mass-points (i.e. is continuous and has continuous pdf) the second term disappears and the rank equates to $F(x)$. For the sake of simplicity, in what follows, we shall almost always assume that this is the case, and therefore that rank equates exactly to position in the distribution $F(x)$.

3. Example Applications

The use of a contextual judgement approach is natural in any situation where a context/reference set influences a decision-maker. Here we discuss various concrete real-world examples where this might be so. In each cases we identify:

(1) The item about which a judgement is being formed.
(2) The context set for the formation of this judgement.

It is also worth commenting briefly on the interpretation of the 'judgement' itself. For economists, one of the most natural applications of the contextual framework judgement is when judgement equates to utility or well-being.[6]

With this interpretation, the contextual judgement framework, in its general form, is essentially identical to the case of interdependent preferences.[7] However, as illustrated by

[5]In the discrete case this particular formulation results in the rank lying between $p_{min}/2$ and $1 - p_{max}/2$ where p_{min}, p_{max} are the probabilities of the minimum and maximum values respectively.

[6]Of course, it can be objected that the judgement is distinct from the underlying experienced utility. However, such 'underlying' experience is almost never observed and for the choices we do observe it it will be the 'judgement' which is the operational quantity and not the underlying utility. We therefore lose nothing by equating the two variables. See Clark et al. (2008) for further discussion on this point.

[7]See Sobel (2005) for a detailed review.

the examples below, while utility/well-being is one obvious and natural interpretation it is not the only one. For example, the judgement could be a narrower measure of 'satisfaction', such as that related to a job or a particular environment or it can even be taken to be a direct monetary (or prestige) payoff in the case of explicit races such as those involved in sport or research. Furthermore, the focus of interdependent preference is often rather different from that found here – for example, fairness and reciprocity figure prominently.

3.1. **Income and Well-Being.**

- The outcome is income.
- The context set are the incomes of others.
- Judgement is reported utility or well-being.

Note that there is flexibility in the definition of the context set. For example, it could be one's work colleagues, one's peers, one's neighbours (Luttmer, 2005), all members of society or even members other societies.

3.2. **Status Races (and Patent Races).**

The classic status race case, as studied by e.g. Frank (1985); Hopkins and Kornienko (2004), corresponds to:

- The outcome is (visible) consumption or income.
- The context set are the consumption or incomes of others.
- Judgement corresponds to 'status' and is 'pure relative' or 'rank-based', i.e. derived from the outcome via a rank-based function.

The patent race case, studied by e.g. Loury (1979); Dasgupta and Stiglitz (1980), corresponds to:

- The outcome is research expenditure.
- The context set are the research expenditure of other participants.
- 'Judgement' corresponds to the payoff from participation which in the standard cases take the simple form of a patent worth V to the winner and zero to everyone else. Formally, take h as the 'hazard' function connecting expenditure to rank and R as the rank function. Then $J(x, C) = g(R(h(x)))$ where $g(1) = V$ and is 0 otherwise – i.e. the winner gets V and everyone else receives 0.

Note that it is easy to generalize this last case to more complex payoffs as a function of rank. For example, most sporting competitions award prizes to those in second and third places as well as to those who come first.

3.3. **Self-Comparison.**

- Outcome is any relevant value (subjective well-being/happiness, satisfaction with work, income, leisure etc).
- Context set is set of own outcomes in previous periods rather than outcomes of others.

For example, Loewenstein and Sicherman (1991) examines preferences for different wage profiles (sequences of wages). They find that increasing wage profiles were preferred to decreasing ones even though the latter delivered a higher present value – a result which obviously does not fit with traditional theory but which can be accommodated within the contextual framework (if the present outcome is compared against outcomes up until the present and increasing profile ensures that the present outcome is always at the top end of the context set).

Loewenstein and Prelec (1993) looks at a more general case of preferences for particular sequences of outcomes (for example, meals at different types of restaurant or scheduling a good and a bad experience) and again find that placing outcomes within a defined sequence (for us 'context') affects preferences for those outcomes.[8] While the interpretation of these results in the paper focus on a specific model based on reference and anticipated utility, a contextual judgement/utility approach is also a natural fit.

3.4. **Evaluation of Pain.**

- Outcome is current level of pain.
- Context set is pain experienced within the current reference frame.
- Evaluation of experience is based on total judgement of the entire context set.

This example was motivated by Redelmeier and Kahneman (1996); Redelmeier et al. (2003), which investigate patients' evaluation of colonoscopy and lithotripsy procedures.

[8] In one case (unpleasant aunt, nice colleagues example) they also show how the temporal proximity of the outcomes affects whether they are evaluated together, as part of the same 'sequence' (context), or separately.

They found that the patients overall evaluation of the experience was better fit by a 'peak-end' rule than by a simple averaging of the reported 'moment-to-moment' level of pain. In our notation $J(C)$ is function of the 'peak' outcome, $max(C)$, and the most recent outcome, x_T: $J(C) = g(max(C), x_T)$ (in the classic case g is a simple average). However, as we shall discuss further below, other well-known judgement rules may also provide a good fit of this data.

3.5. **Job Satisfaction and Quitting.**

- Outcome = Current period value.
- Context Set (C) = job satisfaction in previous periods (or alternatively other associated factors such as wage).
- Quitting then either based on total judgement of C as a whole or on judgement of current outcome in context of C.

This example was inspired by Clark and Georgellis (2004) who investigate the result of applying a 'peak-end' rule to quitting a job.

3.6. **Evaluation of a Product.**

- Outcome = Rating of a product (or a product bundle)
- Context Set (C) = Other products (or product bundles) that are available

This example would imply a generalization of the standard theory of consumer choice in that the utility (judgement) of a consumption bundle would depend, not only on the characteristics of that bundle itself, but on the other options available. In the case of a single product there is a direct connection to the large literature on 'framing' which finds that the context in which a given product is evaluated, in particular the other set of goods within the consumer's comparison set at that moment, can have a large impact on choice.

4. Theory

As set out so far the theory of contextual judgement is very general – probably too general in that it will place no testable restrictions on the data.[9] Thus, our first step should be to ask what kind of restrictions we should introduce if we are to derive empirically testable (and useful) predictions. It is also crucial to remember that the addition of almost

[9]See Kubler (2004) for the problems of testing dynamic models of intertemporal choice even in the case when preferences are more restricted than here.

any variable to a model will improve its fit with data. We must therefore always be asking whether the additional variables that a contextual theory requires deliver sufficient additional explanatory power to warrant their inclusion.

First, recall that associated with any context set we have a cumulative distribution function F. Since C defines F and F defines C in what follows we shall often use them interchangeably. It will also be convenient to assume that F is continuous and that its associated pdf is also continuous. These requirements obviously exclude the case where C is discrete and hence in what follows it is implicitly assumed that C/F can be taken as relating to a continuous distribution. It is a straightforward matter to return to the discrete case so we do not feel much is lost by this simplification of the exposition.[10]

Since, C and F are interchangeable we can, if we wish to emphasize the relationship, write $J(x, F)$ rather than $J(x, C)$. Note that this does *not* imply that the judgement function is rank based, i.e. that $J(x, F) = g(F(x))$. For example, we could have $J(x, F) = x - min(F)$ (where $min(F)$ denotes the minimum value of F/C). It should be clear that this is not a rank-based formulation (recall that J – and hence g in the rank-based model – should be defined independent of the context set).

We will limit our attention here to 'pure contextual' judgement functions.[11] We may also assume that the outcome x is 'good' in the sense that more of it is preferred:

Assumption 8. If the context set is fixed then a larger outcome is judged as greater than

$$x > x' \implies J(x, C) > J(x', C)$$

Our next restriction relates changes in the context to judgement of a given outcome. Specifically we assume that, fixing an outcome, if all others outcomes 'improve' the judgement of that outcome declines, formally:

Assumption 9. If C, C' are two contexts both including a given outcome x_i with F, F' their associated distribution functions then using $\leq$ to denote stochastic dominance for

[10]Of course, allowing for masses of probability and hence for F to be quasi-continuous we could include the discrete case within the continuous one. However, as we wish to ensure that F is continuous and its differential always exists this is not possible.

[11]When we wish to return to the more general contextual case, as we shall below, it will always be presented in a separable form. Hence we lose nothing by examining the 'pure contextual' case here.

distributions (i.e. $F' \leq F$ iff $F'(x) \geq F(x) \forall x$):

$$F \geq F' \implies J(x_i, C) \leq J(x_i, C')$$

Note that it is tempting to reduce this condition to the weaker one that a change in the context which reduces relative position (rank) negatively affects the judgement:

Assumption 10. If C, C' are two contexts both including x with F, F' their associated distribution functions then:

$$F(x) > F'(x) \implies J(x, C) > J(x, C')$$

However, while this restriction would be satisfied by rank-based models, as we shall see below, there are obvious examples of contextual models where this condition is not satisfied – at least for some changes in C/F. This example demonstrates how limited we are in finding simple but general conditions with which to restrict the form of the judgement function. Thus, rather than to continue to search for general restrictions our next step will be to consider example models in in which the judgement function is limited in very specific ways.

5. Contextual Judgement Models: Examples

5.1. **Sufficient Statistics.** Before proceeding to the exposition of the particular models we point out a basic commonality of the main contextual models discussed in detail below. This commonality is that all of them restrict the judgement function to a basic set of 'sufficient statistics' of the context set, C/F. The main 'statistics' used are:

- The mean of the context set: $\bar{x} = \mu = \mu(F)$.
- The standard deviation of the context set: $\sigma = \sigma(F)$.
- Higher order moments, in particular the skewness: $\sigma^3 = \sigma^3(F)$.
- The maximum value: $x_{max} = M = max(F)$.
- The minimum value: $x_{min} = m = min(F)$.
- The range: $r = M - m$.
- The rank of an outcome x: $R(x) = F(x)$.[12]
- (Temporal case) The last (most recent) outcome: x_T.

[12]It is important to distinguish between F itself and the rank function

Note that total/average judgement will be an averaging of the full context set. Hence, it will usually have a *simpler* functional form than the original judgement function. For example, in the case of range-frequency theory the total judgement is simply a linear function of skewness – something not true of the underlying judgement function.

Recall, that we are focusing on judgement functions which are 'pure', that is unaltered under translations by a constant: $J(x + k, C + k) = J(x, C)$. This suggests that the judgement function must be a function only of statistics which are themselves 'linear' (NB: this restriction clearly need not apply to average judgement). In particular, this excludes judgement functions based (directly) on context set moments of higher order than the mean.

Lastly, note that for many standard continuous distributions, for example the normal, statistics such as the maximum and minimum will be infinite. However, we should remember that a continuous distribution is being used simply as an approximation of some underlying finite set. Thus, it would be natural to truncate the support of the distribution at some point where the loss in 'probability' was suitably low.

5.2. **Adaptive and Target-Based Theories.** In an adaptive or target-based theory a given outcome is compared to some summary statistic of its context set – usually its average. In adaptation judgement is still increasing in outcome but is declining in the reference level.[13] In target-based theories the aim is to be as close to the reference level as possible.

In the terminology used here classic adaptation corresponds to a judgement function of the form:

$$J(x, C) = f(x, \bar{x})$$

With the function f satisfying $f_x > 0$ (subscripts indicating partial differential) and $f_{\bar{x}} < 0$. It is obviously a straightforward modification to allow other types of reference value, replacing $\bar{x}$ with a more general statistic. In particular, it, as is often the case, the context set is temporal (for example a sequence of outcomes, such as income, for the same individual) one might wish to weight outcomes, with the more recent having more prominence in the summary statistic.

[13]"The affective value of the experiential correlate of a stimulus varies conversely with the sum of the affective values of those experiences preceding this correlate which constitute with it a unitary temporal group" Paris et al. (1983).

Target-based theories differ from adaptation only in the nature of the f function. If s is the reference statistic then $f_x > 0$ for $x < s$ and $f_x < 0$ for $x > s$. This implies the 'judgement' is increasing as one approaches the reference level.

As such target-based theories have a natural affinity with models with models of conformity. Concrete examples where target-based models may play a role include evaluation of one's weight (Oswald and Powdthavee, 2007) and one's level of smoking (Christakis and Fowler, 2008).

We would point out though that these types of example do raise questions about what judgement represents. In particular 'judgement' can either represent some form of hedonic 'payoff' (be it utility, health or otherwise) or a 'judgement' in a traditional sense of an evaluation of some fact about the world (what the optimum weight is). In particular, it is unclear whether others becoming, say, thinner makes me feel different about my own weight per se or shifts my views what the healthy weight actually is (and beyond this I don't care at all what others weigh). We could label these two approaches as 'utility' focused and 'signalling' focused. It is important to distinguish them because they have very different policy implications – for example, suppose one were considering the impact of the provision of information about 'healthy' eating and weight levels, then in a utility-focused framework one would expect very little impact while in the signalling case one would expect a larger one.

5.3. **Rank Theory.** As the name suggests in these models the only 'statistic' used in assessing an outcome is rank: $R(x) = F(x)$, that is the judgement function is:

$$J(x, C) = g(F(x))$$

'Rank' models often appear under the term 'races' – whether for status or other outcomes (e.g. a patent). In almost all cases g is taken to be linear though this need not always be the case – for example in the case of an all-or-nothing patent race (discussed above) g is a step function which is 0 up to 1 (all 'runners-up') and V at 1 (the winner). One of the key attributes of race models is that total/average judgement is invariant to the context. To see this let G be the integral of the function g then total/average 'judgement':

$$J(C) = J(F) = \int J(x, F)dF(x) = \int g(F(x))dF(x) = G(F(x_{max})) - G(F(x_{min}) = G(1) - G(0)$$

Thus, the total judgement depends only on g and not on F (in the standard case where status is rank, i.e. $g(x) = x, J(x, C) = F(x)$, then $J(C) = 0.5$).

5.4. **Range-Frequency Theory (Range-Rank Theory).** The following is based on Parducci (1995) (78 ff.) and Paris et al. (1983). Range-frequency theory combines a pure rank evaluation (the frequency part) with a 'spatial' or 'range' evaluation based on the location of an outcome relative to the extremes of the distribution.

Formally, it derives from the following two principles, that, assignment of a set of 'stimuli' to categories is done so as to:

Principle 1 (Equal Range): divide the outcome range into sub-categories of equal range.

Principle 2 (Equal Frequency): ensure there are an equal number of outcomes in each category (this equates to ranking them).

Range-frequency theory combines these together. Formally, the end 'Judgement' (J):

$$J = wR + (1 - w)F$$

where R and F are the judgements from Range and Frequency principles respectively.[14] Using the notation from above, the range judgement can be expressed formally as (recall $r = x_{max} - x_{min}$ is the range):

$$R_i = \text{Range}(x_i) = \frac{x_i - x_{min}}{x_{max} - x_{min}} = \frac{x_i - x_{min}}{r}$$

While the frequency judgement is simply the rank:[15]

$$F_i = \text{Frequency}(x_i) = F(x_i)$$

Combining these together the judgement of an outcome x is given by:

$$J(x_i, C/F) = w(\frac{x_i - x_{min}}{r}) + (1 - w)F(x_i)$$

[14]According to Parducci (1995) experimental evidence in psychology suggests $w \approx 0.5$.

[15]In the finite case the normalized rank of an outcome (with no repetitions) is given by $\frac{r_i - 1}{N-1}$ where r_i is the rank in the contextual set and there are N stimuli.

Defining $mid = x_{max} + x_{min}/2$ as the midpoint of the range we now have that the average/total judgement is: (for an explicit derivation see below).

$$\bar{J} = 0.5 + w\frac{\bar{x} - mid}{r}$$

This is essentially a measure of skewness (though negatively correlated with conventional measures) varying between 0 and 1. Dropping the 0.5 this becomes a range between -0.5 and 0.5 and can be expressed as:

$$\text{Average Judgement} = w\frac{\text{Avg. Outcome} - \text{Midpoint of Range}}{\text{Range}}$$

Intuitively, what does this mean? Consider two alternative sets of experience (A,B), say the experience of two different jobs or periods of one's life. Suppose that the day-to-day experiences in both situations were almost entirely the same except that: in one set, say A, there were a few really good experiences while in the other, B, there were a few really bad ones. Let us assume that these 'extreme' experiences formed a small enough part of the overall set that they had little effect on the average level and hence this average is the same for both sets.[16]

Now, which set of experiences would be rated the better? Intuitively, we would expect A to be rated the better. After all, it was the same as B in all ways except that it had a few really good experiences while B had a few really terrible ones. However, according to range-frequency this is completely the wrong way round: B will be preferred to A. Why? Because under B you spent most of your time at a level much better than the worst (those really bad experiences) and near your best (the average), while under A you spent most of your time far below the best (those really good experiences) and near your worst (the average). In terms of the formula, by construction the average was the same for both A and B. However, under A, the midpoint was above the average (because of those really good experiences) while under B the midpoint was below the average (because of those really bad experiences). As a result A is rated as worse than B.

[16]Alternatively one could allow some small compensation in the levels of other experiences in each case to ensure they both ended up at the average level.

5.4.1. *Explicit Derivation of Average Judgement Formula.* Average frequency is given by:[17]

$$\int F(x)dF(x) = 0.5(F^2(x_{max}) - F^2(x_{min})) = 0.5$$

Average range is given by:[18]

$$\int R_i dF(x) = \int \frac{x - x_{min}}{r} = \frac{\bar{x} - x_{min}}{r}$$

Thus average judgement is:

$$\begin{aligned}
J &= wR + (1-w)F \\
&= w(\frac{\bar{x} - x_{min}}{r}) + 0.5(1-w) \\
&= 0.5 + w\frac{\bar{x} - mid}{r} + w(mid - x_{min}r - 0.5) \\
&= 0.5 + w\frac{\bar{x} - mid}{r} + w(\frac{x_{max} - x_{min}}{2r} - 0.5) \\
&= 0.5 + w\frac{\bar{x} - mid}{r} + w(\frac{r}{2r} - 0.5) \\
&= 0.5 + w\frac{\bar{x} - mid}{r}
\end{aligned}$$

6. Empirical Testability

With these various contextual models in front of us there are two major empirical questions to ask: a) how can we test for evidence of contextual judgement b) how can we distinguish one particular contextual model from another.

6.1. Evidence for Contextual Judgement.

Let us rephrase the first question as: how can we test for evidence that a judgement function is not 'pure' self? Put in this way the answer appears obvious: include any statistic based on the context set and see whether this 'improves' the fit with the data. In particular, one would regress U_i, the reported

[17] In the discrete case it is even simpler:

$$\frac{1}{N}\sum_{i=1}^{N} F_i = \frac{1}{2N(N-1)}(N(N+1) - 2N) = 0.5$$

[18] In the discrete case this would be:

$$\frac{1}{N}\sum_{i=1}^{N} R_i = \frac{\bar{x} - x_{min}}{r}$$

judgement, on X_i, the outcome variable, and S_i some context statistic:[19]

$$U_i = \alpha X_i + \beta S_i + \cdots$$

A finding of $\beta < 0$ would then be taken as evidence that judgement was not 'pure' self and hence had some ('pure') contextual component.[20] Observe also that we obviously need S_i to vary with i for otherwise no test is possible.

Of course, this basic test has problems, especially related to omitted variables. In particular, if S is positively correlated with variables which negatively impact U then β will be negatively biased. However, at least at first sight there do not seem obvious candidates for such omitted variables. In particular, higher S implies that those 'around you' (in your context set) have better outcomes, and one would expect higher outcomes for others to be negatively correlated with any omitted variables that impact on judgement.

There is, though, one other issue of some significance. This arises from the fact that a regression is a (crude) approximation to whatever the real judgement function is. If S_i is correlated in some systematic way with X_i then adding S_i may be simply helping to correct a misspecification in the regression (related to the pure-self component) rather than picking up genuine contextual effects. To illustrate, let us consider a case where judgement is in fact pure-self, $J = g(x)$, but displays diminishing returns. Concretely, $U = g(X) = X - X^2$ – U is utility, X income say – and the context statistic S_i is the $\mu(C_i)$ the mean of some group related to each X_i – for example, their neighbours. Assuming that the exact quadratic nature of U is unknown so the regression is still:

$$U_i = \alpha X_i + \beta S_i + \cdots$$

Now, it should be clear that if S_i is positively correlated with X_i^2 then β will be found to be negative (more generally if it is negatively correlated with the omitted part of $g(X)$). A positive correlation with X_i^2 isn't immediately obvious since $S_i = \mu(C_i)$ but it is possible if the distribution of values within a context set varies with the average of the values in the

[19] Alternatively, rather than include S_i independently we might use relative outcomes:

$$U_i = \alpha X_i + \beta \frac{X_i}{S_i} + \cdots$$

Of course if the regression were done in logs this would become linear again.

[20] One could peform the (weaker) test for $\beta \neq 0$. However assumption 9 above implies that β should be negative.

context set in a particular way. Specifically one requires that at low and high levels values are positively skewed (i.e. most people are poorer than the average) while at intermediate levels it is negatively skewed (most people are richer than the average). This will result in the difference between x and μ covarying systematically with x (and hence μ as x changes which results in a negative coefficient on μ in the regression.[21]

To summarize: the basic test for contextual effects involves adding a statistic based on context to the regression – most commonly the context mean or the ratio of the outcome to context mean. While this should be reasonably robust one needs to be careful both regarding omitted variables and mis-specification of the underlying functional forms.

6.2. **Distinguishing Contextual Models.** Supposing then that we have done successfully established the presence of contextual effects. At this point we come to the second question: how we can distinguish one contextual model from another.

To illustrate the difficulties consider the following example. As discussed above, Redelmeier and Kahneman (1996), based on evidence of the evaluation of a painful medical procedure, suggest that evaluations of temporal outcomes may follow a 'peak-end' rule (PE). However, at least based on the summaries of the data presented in the paper, it appears the result could be fit equally well by assuming that evaluations were based on a range-frequency average judgement (RF).[22] Specifically, in the RF case $\bar{J} \approx \frac{\bar{x}-mid}{r}$, which equates to a (negative) measure of skewness. Now, in many case a lower 'end' under 'peak-end' was also correlated with a situation in which the pain was lower than its maximum for longer (see e.g. Fig 1 in the paper). Noting that here we are dealing with pain and thus higher levels of pain equate to lower x, this sort of pattern would result in a lower average judgement under RF as well as under PE.[23] Thus, it could be that patients are basing their evaluations on a RF approach rather than a PE one.

[21]Intuitively a linear regression on a (negative) quadratic function produces a line which is too high at low and high values and too low in the middle. With the skews as described $\mu - x$ will be be positive at low and high values and negative and intermediate values. Thus, adding in μ will improve fit with a negative coefficient on μ.

[22]It would obviously be interesting here to re-analyze the full original dataset.

[23]The possible conflation of PE and RF explanations is even more true of the follow-up paper, Redelmeier et al. (2003). There the treatment was intentionally prolonged a little at a level of minimal pain. While this would obviously reduce the 'end' pain value it also makes the outcome distribution more negatively skewed (you now spend more time far above the worst).

Is it possible then to distinguish the two cases? The answer is clearly yes. For example, RF theory, contrary to PE, would suggest that:[24]

- Increasing the period of low pain would improve evaluations (more time time is now spent in the high-end of the experience range).
- Holding all else constant, an increase in the peak level of pain would actually improve evaluation (the distribution is even more skewed and the midpoint will have dropped more than the mean hence $\bar{J}$ will have increased under RF).

Furthermore, the direct empirical test is simple: rerun the original regression using the range-frequency statistics (mean, midpoint) in place of (or in addition to) the peak-end statistics (peak and end). This point holds more generally. Many of the models discussed above predict that judgement (whether total or individual) should be based on a relatively limited set of context statistics (which differ across the models). It should therefore be feasible to compare the performance of different context statistics in a given dataset and therefore obtain some indication of which models are better than others.

7. Happiness, Status and Inequality

Contextual theories led themselves naturally to analysing overall happiness in groups since we make take the group as the context set in making judgements. Let us consider then the following simple model:

There are N entities (individuals, households etc). Each entity i has a single 'outcome' variable, x_i, which could be thought of as being income, consumption or the like. Where useful the variable x shall stand for the vector of outcomes, or (almost) equivalently, a continuous distribution representing outcomes.

Individual i has a contextual utility (or well-being) function, U, which is 'separable':

$$U = U(x_i, x_{-i}) = u(x_i) + \alpha j(x_i, x_{-i})$$

Where u is standard utility function dependent only on own-outcome and displaying diminishing returns. j is a 'pure' contextual judgement function which depends both on

[24] In both cases, we should note that the predictions hold only 'up to a point' for it is unlikely that patients evaluations are *purely* contextual. In particular, it is likely that the total overall level of pain matters in and of itself. In this case, increasing the level of pain, or the time for which it is experienced, has some direct negative impact. While for small increases this effect may be dominated by a 'pure' contextual effect for larger increases this is unlikely to be so.

own-outcome and on the outcomes of others and which we often refer to as *comparison utility* in order to distinguish from *standard utility* u.

Remark: Here, it is natural to interpret u as representing the benefit derived directly from one's own activities and resources, independent of comparison with others, while j represents 'status', or more widely one's satisfaction with one's position and outcomes in comparison with others.

For comparison utility it will be valuable to focus on a specific case derived from our analysis above. In particular, take j as the judgement function arising from range-frequency theory:

$$j_i = j(x_i, x_{-i}) = wR_i + (1-w)F_i$$

where R and F are the judgements from Range and Frequency principles respectively. Our next step is to aggregate to get average well-being. We do this giving equal weight to each individual outcome:

$$\bar{U} = \frac{1}{N}\sum_{i=1}^{N} u_i + \alpha j_i = \bar{u} + \alpha\bar{j}$$

Where average standard utility is $\bar{u} = E_x(u(x))$ and average judgement is:

$$\bar{j} = E_x j(x) = w\frac{\bar{x} - 0.5(x_{max} + x_{min})}{x_{max} - x_{min}}$$

7.1. **Simulation.** At this point, our next step is to do some numerical experiments to flesh out this theoretical model. For this purpose we consider applying the above analysis to the case of the income distribution for a country. It is generally accepted that the income distribution, at least for developed countries, can be well approximated by a log-normal distribution (perhaps with a Pareto distribution for the tail).[25]

We start with the standard log-normal with parameter 1.0. There is no clear way (without empirical data which is currently lacking) to calibrate the respective levels of standard and comparison utility via α. Thus, our approach is set α equal to the mean of the standard utility for the distribution under consideration (0.787 in this case). For the standard utility function we take the frequently used natural log form: $u(x) = ln(1+x)$ (the 1 is added simply for analytical convenience so that the distribution can start at 0

[25] See e.g. Souma, *Physics of Personal Income*, http://arxiv.org/abs/cond-mat/0202388.

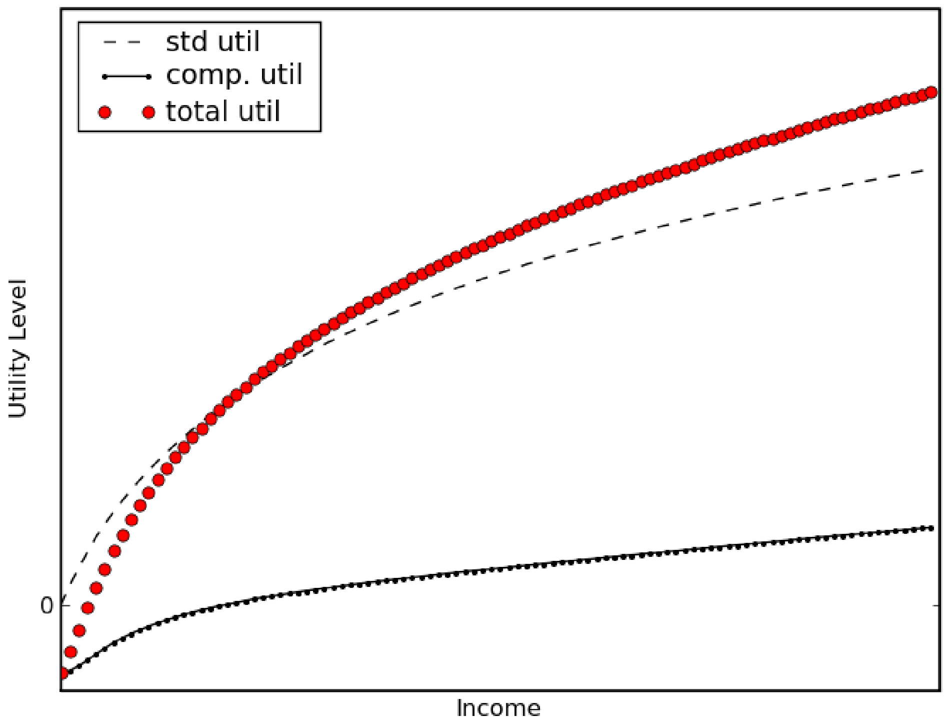

FIGURE 1. Total, standard and comparison utility for Log-Normal

rather than 1). Individual utility as a function of their income is then shown in Figure 1 (the graph is shown up to the top 1% point on the income distribution).

The form and interaction of the two different sources of 'happiness' or 'well-being' in total utility can be clearly seen: both standard and comparison utility grow in a concave fashion from their initial values though comparison utility flattens out more rapidly and becomes approximately linear quite quickly.[26]. Because standard utility and comparison utility both grow together total utility grows more sharply than either of them and hence the marginal utility of extra income in terms of total utility will be higher than the marginal utility of extra income in terms of standard or comparison utility.

This has an interesting implication if we consider for a moment the situation where everyone gets a little bit better off (equally in amount). The judgement function for comparison utility (in commmon with most) has the property that a constant linear shift

[26]Here, as is expected comparison utility starts out negative (and hence so does total utility). For those who find this problematic it is worthwhile to recall that the utility functions are only defined up to a constant and so it would be trivial to add a constant and make all value non-negative.

in the 'context set' (here the incomes) has no effect on the average judgment. Thus the benefit of a one unit increase in income for all in terms of total utility is simply the benefit in terms of standard utility (as there is no effect on comparison utility).

Thus, at the societal level the marginal benefit of extra income (equally distributed) is just that arising from standard utility. As just discussed this marginal benefit is below the marginal benefit seen at the individual level (which equates to total utility). Thus, when comparison utility is present in total utility individual incentives and societal-wide incentives will diverge with individuals (in this model, though also more generally). This is a concrete demonstration of how 'status races' and 'keeping up with the Jones' can result in everything 'running harder but staying in the same place'. Potentially this could be inefficient and make for worse societal well-being since the expenditure of effort involves real costs (be that in terms of people's leisure time or the emission of carbon dioxide).[27]

Returning to our example, the graph above only shows utility as a function of income. Our next step will be to plot the density function of utility – i.e. to 'interact' the standalone function with the distribution of income. This can be seen in Figure 2 which shows the utility density (i.e. utility times pdf) as a function of income. The integral under this graph will then be average value for that utility (total, standard, comparison) over the whole population.

As can be seen the results are quite striking. In particular while on the first figure the regions in which total utility is above and below standard utility seem fairly even matched. However, once the utility density is plotted it is clear that the bulk of the population lies in the region where comparison utility is negative and hence total utility is less than standard utility. This is born out by the figures for this distribution shown in the table below (NB: these are unnormalized so only relative values, not absolute ones have meaning). As can be seen average total utility is always below average standard utility, often well below (average comparison utility is negative). Also interesting is to consider what happens when the distribution gets more unequal (corresponding here to a higher scale parameter). In the examples shown we have not controlled the mean and

[27]Caution is warranted here as it is possible that there may offsetting effects that make this 'over-expenditure' of effort beneficial (for example a gap between the social and private value of whatever is produced by this effort – be this from externalities or the standard gap between producer and total surplus). Any overall welfare evaulation would need to take these effects into account along with any incentives to over-exertion engendered by the 'status-race'.

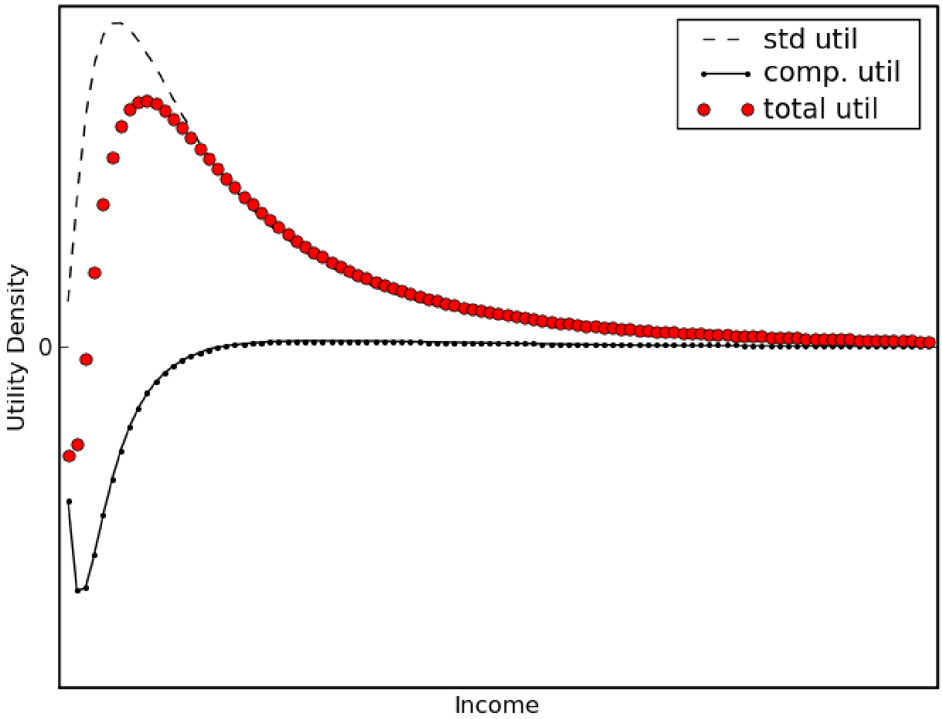

FIGURE 2. PDF of utility over the total population

Scale	Mean/SD (Income)	Total Util	Std. Util	Comp. Util	Ratio
1	1.6/2.1	0.65	0.78	-0.13	0.39
1.5	3.1/9.0	0.71	0.88	-0.18	0.22
2.0	7.4/54.1	0.79	1.0	-0.22	0.11

TABLE 1. Mean values for main parameters of interest for various lognormal distributions (scale is the lognormal scaling parameter). Ratio is the ratio of total utility to mean income.

hence an increase in scale increases both the mean and the dispersion (as measured by the variance) of the distribution. The effects are dramatic. Increasing the scale parameter from 1.0 to 1.5 increases the mean income by almost 2 (1.6 to 3.1) but total utility by only 9%. Increasing the scale to 2.0 from 1.0 results in an approximately 400% increase in the mean income but only increases average total utility by 17%.

8. Dynamics, and the Real Effect of Races

One of the most natural applications of contextual models is to 'dynamic' situations in which the outcome is not exogenously given but endogenously determined by the choices of agents. For example, the very term 'status race' implies that agents exert effort in 'racing'. Note that, as discussed in the previous section with ('pure') contextual effects we immediately have the existence of an externality – the positional externality, arising from the fact that a change in my outcome affects not just my evaluation (utility) but that of others. From Assumption 9 (higher context outcomes negatively affect own judgement) it follows that this externality is a negative one.

The immediate implication of this, one might think, is that too much effort will be expended. That is, in any dynamic game in which agents make private choices about effort, and the resulting outcomes are judged (partially) contextually, the equilibrium level of effort will be too high.

This result is certainly true whenever there are no other externalities. However, in many cases this is not so. To illustrate consider the classic example of a patent race. Two facts are well known. First, when the private value of the innovation (the value of the patent) equals the social value – so no additional 'externalities' – there is too much 'racing' by firms – that is they spend too much on R&D from a societal perspective. This confirms the basic point that with 'positional externalities' effort will be inefficiently high.

However, suppose also that the private value is less than the social value – as is, in fact, likely. In this situation, at least if private value falls sufficiently below social value, effort is likely to be inefficiently low, and the race, by increasing that effort is likely to move effort levels closer to the optimum. Thus, in this second case, though individual firms will feel the race makes them work too hard, from society's point of view this is all to the good.

We would note here that this same point can also be made about the classic 'status races' (see e.g. Frank (1985, 2005)). In the absence of any other effects the existence of positional (e.g. conspicuous consumption) and non-positional (e.g. leisure) goods will result in an inefficiently high use of the former and under-user of the latter. However, allowing any kind of positive externality from the effort expended on the positional goods may change this substantially.[28]

Of course, the additional externality could be negative rather than positive in which case the positional externality effect and its associated over-exertion of effort, are reinforced. For example, assuming carbon emissions are a negative externality, if positional goods result in more carbon emissions than non-positional ones the negative effects of the 'positional externality' are increased by the negative effects of increased carbon output.

References

Bordley, R. (1986). Satiation and habit persistence (or the dieter's dilemma). *Journal of Economic Theory*, 38(1).

Brown, G., Gardner, J., Oswald, A., and Qian, J. (2008). Does Wage Rank Affect Employees Well-being? *Industrial Relations*, 47:355–389.

Christakis, N. and Fowler, J. (2008). The collective dynamics of smoking in a large social network. *New England Journal of Medicine*, 358(21):2249.

Clark, A. and Georgellis, Y. (2004). Kahneman meets the Quitters: Peak-End Behaviour in the Labour Market.

Clark, A. E., Frijters, P., and Shields, M. A. (2008). Relative Income, Happiness, and Utility: An Explanation for the Easterlin Paradox and Other Puzzles. *Journal of Economic Literature*, 46(1):95–144.

Clark, A. E. and Oswald, A. J. (1996). Satisfaction and Comparison Income. *Journal of Public Economics*, 61(3):359–381.

Clark, A. E. and Oswald, A. J. (1998). Comparison-concave utility and following behaviour in social and economic settings. *Journal of Public Economics*, 70(1):133–155.

Dasgupta, P. and Stiglitz, J. (1980). Uncertainty, Industrial Structure, and the Speed of R&D. *The Bell Journal of Economics*, 11(1):1–28.

Frank, R. (1989). Frames of Reference and the Quality of Life. *The American economic review*, pages 80–85.

[28]To give a concrete example. Suppose my work involves developing new software products. If I do not receive the full surplus that my work produces then my working longer at the office, whether in order to keep up with the Jones or simply because everyone else is doing so, has real (positive) effects on society – that software is available sooner.

Frank, R. H. (1985). The Demand for Unobservable and Other Nonpositional Goods. *American Economic Review*, 75(1):101–16.

Frank, R. H. (2005). Positional Externalities Cause Large and Preventable Welfare Losses. *American Economic Review*, 95(2):137–141.

Hopkins, E. and Kornienko, T. (2004). Running to Keep in the Same Place: Consumer Choice as a Game of Status. *American Economic Review*, 94(4):1085–1107.

Hopkins, E. and Kornienko, T. (2007). Status, Affluence, and Inequality: Rank-Based Comparisons in Games of Status. Levine's Bibliography 122247000000001442, UCLA Department of Economics.

Kubler, F. (2004). Is intertemporal choice theory testable? *Journal of Mathematical Economics*, 40(1-2):177–189.

Loewenstein, G. and Prelec, D. (1993). Preferences for Sequences of Outcomes. *Psychological Review*, 100(1):91–108.

Loewenstein, G. and Sicherman, N. (1991). Do Workers Prefer Increasing Wage Profiles? *Journal of Labor Economics*, pages 67–84.

Loury, G. C. (1979). Market Structure and Innovation. *The Quarterly Journal of Economics*, 93(3):395–410.

Luttmer, E. F. P. (2005). Neighbors as Negatives: Relative Earnings and Well-Being. *The Quarterly Journal of Economics*, 120(3):963–1002.

Niedrich, R. W., Sharma, S., and Wedell, D. H. (2001). Reference Price and Price Perceptions: A Comparison of Alternative Models. *Journal of Consumer Research: An Interdisciplinary Quarterly*, 28(3):339–54.

Oswald, A. J. and Powdthavee, N. (2007). Obesity, Unhappiness, and The Challenge of Affluence : Theory and Evidence. The Warwick Economics Research Paper Series (TWERPS) 793, University of Warwick, Department of Economics.

Parducci, A. (1965). Category Judgment: A Range-frequency Theory. *Psychological Review*, 72:407–418.

Parducci, A. (1995). *Happiness, Pleasure, and Judgment: The Contextual Theory and Its Applications*. Lawrence Erlbaum Associates.

Parducci, A. and Perrett, L. (1971). Category Rating Scales: Effects of Relative Spacing and Frequency of Stimulus Values. *Journal of Experimental Psychology*, 89(2):427–452.

Paris, S., Sarris, V., Olson, G., Parducci, A., Lienert, G., and Stevenson, H. (1983). *Learning and Motivation in the Classroom*. Lawrence Erlbaum Associates.

Redelmeier, D. and Kahneman, D. (1996). Patients' memories of painful medical treatments: Real-time and retrospective evaluations of two minimally invasive procedures. *Pain*, 66:3–8.

Redelmeier, D., Katz, J., and Kahneman, D. (2003). Memories of colonoscopy: a randomized trial. *Pain*, (104):187–194.

Sobel, J. (2005). Interdependent Preferences and Reciprocity. *Journal of Economic Literature*, 43(2):392–436.