**Individual Differences in Information Integration Studies of Children's Judgment/Decision-Making: Combining Group with Single-subject Design via Cluster Analysis**

Etienne Mullet, Institute of Advanced Studies (EPHE) Paris
Joeri Hofmans, Vrije Universiteit Brussels
Anne Schlottmann, University College London

9118 words

Our work uses experimental methods to test children's judgment/decision-making (JDM). Experimental work often focuses on task and process analyses at the group level, with individual differences treated as error variability. Here, we describe how to assess/interpret individual differences within experiments using single-subject design. Traditionally, single-subject design appears in single case studies, with issues of generalizability arising. Our approach, in contrast, involves groups of standard size, analyzed at group and individual subject level. We then group individuals with similar patterns, for conclusions about the existence and contributions of systematic individual differences to development. Our examples here use Information Integration Theory (IIT). Our general perspective, however, could be useful for other experimental paradigms as well.

IIT is a well-known approach to studying multi-dimensional judgment/problem-solving in adults and children (Anderson, 1981, 1982, 1991, 1996). It diagnoses how multiple informers combine into a unitary judgment, describing the process through algebraic models. These abound in JDM, but are rarely tested directly. IIT has made important basic contributions to JDM, providing, for instance, tests of how the structure of human judgment corresponds to classic models for probability (Lopes, 1976), expected value (Shanteau, 1974), and sequential decision-making (Shanteau, 1970).

Recently, similar tests were done with children (Schlottmann & Wilkening, 2011). The examples in this chapter concern the Expected Value (EV) model, under which outcome value (v) and probability (p) combine multiplicatively, while component EVs combine additively, illustrated for two-outcome events below:

$$EV = p_1 \times v_1 + p_2 \times v_2 \qquad (1)$$

Children from pre-school age make such structurally appropriate EV judgments. They know that EV depends on probabilities and values, combining them multiplicatively, although there are doubts, for adults and children, about component additivity. Children also know determinants of probability, for instance, that this increases with number of targets, but decreases with number of non-targets.

A key feature of IIT is that it empirically separates how individuals evaluate probabilities and values from their strategies for integrating informers. Individuals typically do not quantify informers in an objective, precise way, but estimate subjective values and probabilities intuitively. Children's subjective values differ from adults' and may, occasionally, seem odd. While this may produce deviations from normativity, such subjective evaluation is qualitatively different, in our view, from structurally incorrect judgment. If children come closer to the mathematical standard with age, or if some children are better at this, it could reflect either or both. IIT studies, which often focus on the structural components of judgment, have established conclusively that these are operational from young ages. But individual children still differ in their judgments. How can we deal with this?

A well-known problem with conclusions from group data is that these may contain averaging artifacts, misrepresenting any individual's performance. Conceivably, an impression that children integrate probability and value may arise for the group, even if no child does so: Some may focus on probability, others on value. Group-level judgments may show additivity of component EVs, even though individuals are either risk-seeking or risk-averse, but this averages out. IIT assesses this empirically, through designs that can also be analyzed at single-subject level. We then see whether most participants show the group pattern, or whether systematic individual differences contribute.

There are several issues, however, when applying this to children. This paper illustrates these with two examples from primary-schoolers' EV judgment and discusses approaches to analysis, single-subject ANOVA, means-based classification (Schlottmann & Anderson, 1993; Schlottmann, 2001), and cluster analysis (Hofmans & Mullet, 2013).

## Single-subject ANOVA

On first glance, single-subject ANOVA may seem sufficient to assess individuals. This method is objective, requiring no specialist knowledge. To get an ANOVA error term, participants simply judge more than one replication of what is a within-subjects design at group level. However, the power of single-subject ANOVA is limited by noise in the data and the number of noise-reducing replications elicited from participants. To show that children make continuous estimates of probabilities or value, we need multiple factor levels, resulting in 3 x 3, or larger, designs; children often cannot attend for more than two replications, and multiple sessions are difficult to arrange in schools. Moreover, children's data are inherently noisier than adults'. Effect sizes similar to those at group level may thus not reach significance for individuals. Single-subject ANOVAs are useful, because significant effects are convincing evidence. When effects are missing, however, the problem may be low power, not the child. Confidence intervals do not help; these also depend on variance. Often, therefore, assessment of individual differences must rely more on comparing individual and group data patterns than statistical tests.

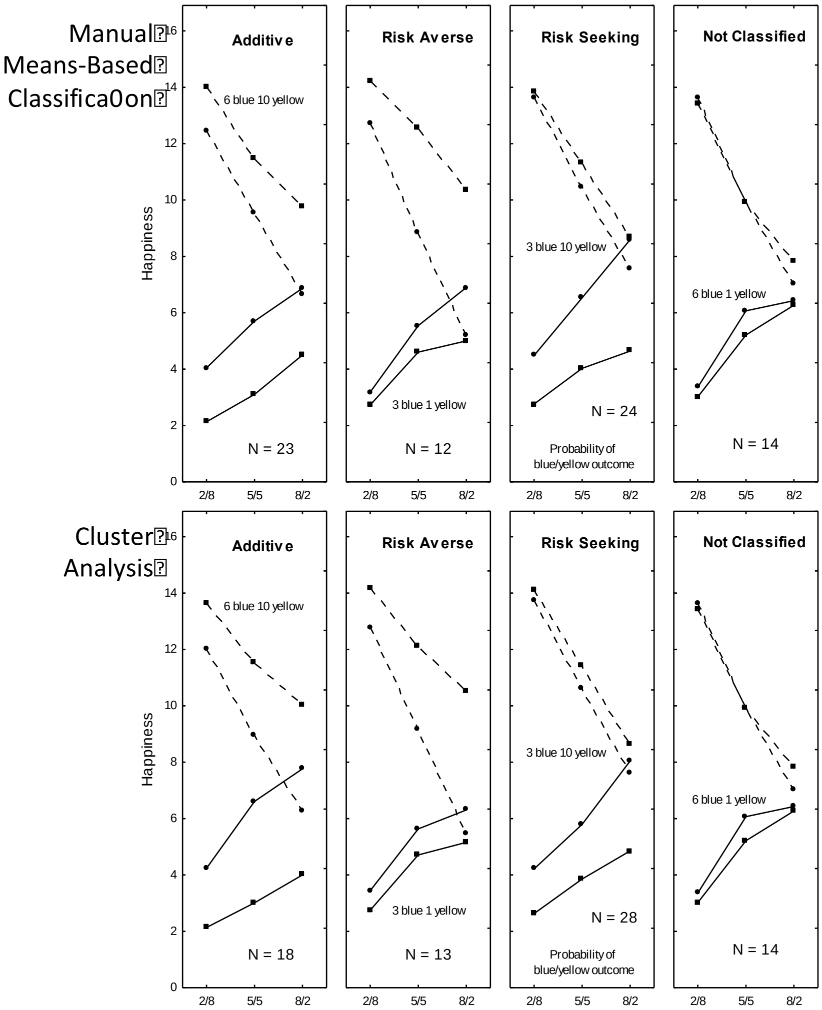
Children's individual data patterns, however, also suffer from noise, typically looking quite messy, far from the clean interaction patterns appearing at group level or for adults. Purely visual inspection then becomes unreliable and subjective. Here, we discuss more objective approaches to this problem.

Our first example is a study from Schlottmann (2001), in which 6-year-olds, 9-year-olds and adults (N=73, half girls) judged EV of roulette-type games in which a puppet would win one crayon prize if a marble landed on one colour ("blue"), and another crayon prize if it landed on the alternative colour ("yellow"), with outcome probabilities and size of crayon prizes on blue and yellow varied factorially, in a 3 x 2 x 2 design, respectively, per participant. Children judged happiness with each game on a stick scale, suitable even for pre-schoolers. Each participant judged at least two replications, allowing for group and single-subject analysis. Some (Falk & Wilkening, 1998) argue that puppet paradigms do not engage children sufficiently for good performance, but children's sophisticated judgment patterns found in this and other studies, similar to those for adult controls, are clear evidence against this. At the group level, all ages judged EV similarly: They (1) considered both outcomes, blue and yellow, (2) appreciated that EV increased with the probability of the larger outcome, even if this meant that increased probability of the blue outcome sometimes increased EV (if it carried the larger prize), sometimes decreased it (if it carried the smaller prize), (3) integrated values and probabilities multiplicatively and (4) appeared to add component EVs, all as normatively prescribed.

Single-subject level statistics confirmed only (1) and (2), however. The interaction reflecting (3) was theoretically small, rarely reaching significance at individual level, especially for younger children with large error. As for (4), given the low power single-subject design, the absence of deviations from parallelism cannot support additivity convincingly. In fact, prior work with adults (Shanteau, 1974) and children (Schlottmann & Anderson, 1994) suggested outcome non-additivity, which the study aimed to consider in more detail.

## Classification of Individuals' Means Patterns

For stronger evidence, the study went on to consider whether means patterns conformed to the predictions of the EV model. Regarding (3), normative multiplication implies that variations in prize/value of an outcome affect EV more if outcome probability increases, producing diverging fan patterns, as seen, at group level, in Figure 1. Each panel shows children's mean judgment of four outcome combinations (curve factor), as a function of their probabilities (horizontal). The solid bottom curve in each panel, for instance, is for games in which 3 crayons could be won if the marble landed on blue, and 1 if it landed on yellow. The curve increases towards the right, as proportion of blue marbles and likelihood of the blue outcome increases. But the second curve, differing only in a 6-crayon prize on blue, increases faster, with curve divergence predicted by the multiplicative model; such fans appear throughout. At the individual level, rather than relying on visual impressions, multiplication was diagnosed from a defined means pattern, if the simple effect of blue was larger for the highest than lowest probability of winning it. This appeared for most children, confirming the group pattern.




| | Additive | Risk Averse | Risk Seeking | Not Classified |
|---|---|---|---|---|
| Younger (n=28) | 5 (18%)<br>2 (7%) | 4 (14%)<br>6 (21%) | **10 (36%)**<br>**11 (39%)** | 9 (32%) |
| Older (n=28) | 9 (32%)<br>6 (21%) | 5 (18%)<br>5 (18%) | **9 (32%)**<br>**12 (43%)** | 5 (18%) |
| Adult (n=17) | **9 (53%)**<br>**10 (59%)** | 3 (18%)<br>2 (12%) | 5 (29%)<br>5 (29%) | |
| Total (n=73) | 23 (32%)<br>18 (25%) | 12 (16%)<br>13 (18%) | **24 (33%)**<br>**28 (38%)** | 14 (19%) |

*Figure 1: Clusters identified through the means pattern (top panels) or cluster analysis (bottom panels); number of children (and %) are listed underneath (top row in each cell is for the top panels). Each panel shows EV judgments as function of blue and yellow outcome combination (curve factor, listed in the figure), and probability of each outcome (horizontal, given as number of blue/yellow marbles).*

In contrast, (4), additivity of component EVs, appeared only rarely for children. Additivity means that the blue prize effect is independent of the yellow prize. This appears, in Figure 1, when comparing the dashed and solid curve pairs, for games differing only in the yellow prize, 1 (bottom) or

10 crayons (top). Outcome additivity implies equal divergence in these curve pairs, seen in the left panel. Additivity increased with age.

Two non-additive strategies appeared more often for children, linked to risk attitudes discussed in the literature: One group of "Risk-Seekers" included participants with larger effects of the prize on blue if this was the larger prize, so the risky option. Another group of "Risk-Avoiders" included participants with larger effects of the blue prize if this was the smaller prize, so the riskless option. Children may come to these strategies as a compromise between outcomes, with risk-seekers or risk-avoiders starting from the risky/safe outcome, respectively, with insufficient adjustment for the alternative (Schlottmann, 2001). There were also some not-classifiable children, disordinal or without blue prize effect. This group decreased with age, as additive participants increased. Reyna and Ellis (1994) first reported that risk attitudes change developmentally, while our study first showed individual differences within age groups, with risk-seeking most frequent for children. Levin and Hart (2003) later confirmed risk-seeking in children in a more standard choice paradigm.

In this study, means-based, single-subject classification could establish the presence/absence of individual differences in children's judgment, relating these to experimental task analysis. The objective, means-based classification helped overcome the low power of single-subject statistics. Nevertheless, the approach faces difficulties, one being the practical issue of requiring laborious, by-hand classification of individual data patterns. To make this reliable, and this is the second issue, the patterns considered should be simple. Here, we had a reasonably complex 3-way design, and individual 3-way patterns were messy. Instead, the classification took a "divide and conquer" approach, considering separately evidence for components (1) to (4) of children's understanding. This was done for conceptual simplicity, helped stabilize the data by averaging over disregarded factors, but also meant we did not consider how these components relate.

A third, more principled, problem is that this approach is restricted to looking for pre-defined patterns, suggested by theory. If children use unexpected strategies, these might be missed, especially if hidden in noise. This is not a problem specific to our method, of course, but ails all confirmatory approaches. It has been discussed, for example, for Siegler's (1981) rule assessment approach (Anderson & Wilkening, 1982; Wilkening & Anderson, 1991).

Finally, while ideal, theoretical data patterns may be perfectly clear, in practice one has to distinguish between them within noisy data. For instance, under normative outcome additivity, the simple effect of the blue prize is the same for both levels of the yellow prize, but in practice the simple effects will almost never be exactly equal. The approach therefore requires a fairly arbitrary criterion value for how close simple effects should be to be considered equal.

Below we explore how cluster analysis overcomes some of these limitations. We re-analyze the data considered above in different ways, firstly, to see how the approaches correspond when classifying in the same theory-driven, confirmatory way; secondly, to evaluate whether an exploratory, data-driven cluster analysis provides additional insights. We then apply cluster analysis to a new data set that the other approaches cannot handle.

**Cluster Analysis in IIT**

Hofmans & Mullet (2013) first described the use of K-means cluster analysis to study individual differences in IIT studies. Cluster analysis automates pattern classification. All cluster procedures aim to find groups/clusters of participants with similar profile or pattern across multiple variables. This fits naturally with IIT's main interest in the integration pattern across dimensional variables. Specific cluster techniques vary in type of clustering and similarity/distance measures. K-means clustering is non-hierarchical and centroid-based, and it is popular because it is simple and less susceptible to outliers or specifics of the distance measure than other methods (Hair et al. 2008). IIT applications have been with adults, secondary students (e.g., Liegeois et al., 2003), or elderly people (Leoni et al., 2002), typically to situations involving distinct, even polarized attitudes, often ethical judgment (e.g. Muñoz Sastre et al.

2007, Kpanake et al., 2013). Here, we apply the approach to younger children, varying more continuously, with overlapping patterns, in judgment.

The K-means method partitions a data matrix $\boldsymbol{X}_{I\times J}$ ($I$ individuals measured on $J$ variables) into a binary partitioning matrix $\boldsymbol{P}_{I\times K}$, and a centroids matrix $\boldsymbol{C}_{K\times J}$. The partitioning matrix contains the memberships of $I$ objects to $K$ clusters, while the centroids matrix contains the centroid for each cluster; this being the vector of means on all $J$ variables across all objects belonging to that cluster. The K-means method chooses $\boldsymbol{P}_{I\times K}$ and $\boldsymbol{C}_{K\times J}$ such that the following least squares loss function (representing the error sum of squares (SSE)) is minimized:

$$\mathrm{SSE} \underset{P,C}{=} \min \|\mathbf{X} - \mathbf{PC}\|^2 \tag{2}$$

Thus, the K-means method searches for partitions so that the squared Euclidean distance between individual observations and the centroid vector for the cluster is at least as small as the distances to the centroids of the other clusters. This search for $\boldsymbol{P}_{I\times K}$ and $\boldsymbol{C}_{K\times J}$ is iterative, typically following the procedure below (Hofmans et al., 2015; Steinley, 2003):

1. *Initialization stage*: Individuals are randomly assigned to one of $K$ clusters, keeping the distribution of individuals across clusters as even as possible. Once all individuals are allocated, initial cluster centroids are calculated.
2. *Update of* $\boldsymbol{P}$: The fit between each individual $i$ and each cluster $k$ is calculated by
$$d^2(i,k) = \textstyle\sum_{j=1}^{J}\left(x_{ij} - \bar{x}_j^k\right)^2 \tag{3}$$
after which each individual is assigned to the cluster for which $d^2$, the squared Euclidean distance to the centroid, is minimal.
3. *Update of* $\boldsymbol{C}$: Using the updated cluster memberships, new centroids are calculated.
4. *Loop*: Steps 2 and 3 are repeated until individuals cannot be reallocated anymore.

The number of clusters, K is an input parameter set by the researcher. Algorithmic, graphical and formulaic methods can help decide on K (Steinley, 2006), but probably the easiest method is graphical: One plots SSE of each cluster solution against values of K. The curve decreases, as added parameters increase fit; a steep slope from K to K+1 clusters signals that cluster K+1 is helpful (as SSE substantially decreases), a flat slope suggests it adds little explanatory power, so the model with K clusters is preferable. This approach is similar to the well-known scree test in factor analysis. Determination of changes in slope can be somewhat subjective, however, so the most important criterion for choosing K is that the cluster solution is meaningful. If there are no a priori reasons for a particular K, one may gradually increase K and compare solutions. If, for a given K, further increases maintain the original cluster structure, merely splitting out very small, outlier clusters, then the K solution is probably good. If an increase from, say, 3 to 4 clusters produces a novel solution with distinct cluster profiles, then the better solution depends on interpretation and external criteria, e.g., replicability.

The advantage of using cluster analysis with IIT is that IIT constrains cluster interpretation, which can be difficult with the multivariate data to which cluster analysis is typically applied. It is easy to visualize distinct cluster profiles in scatterplots of two-dimensional data, the standard example, but realistic cases have high-dimensional, noisy data, here 18 variables from the 2 x 3 x 3 design. In IIT studies, the cluster profiles, fortunately, are integration patterns, no different from those found at group or single-subject level, and interpreted accordingly. IIT vastly increases the level of resolution at which clusters can be interpreted, while cluster analysis provides IIT with an efficient method to classify individuals; the approach is synergistic.

We combine cluster analysis with ANOVA, because the interpretation of factorial interaction patterns is crucial to IIT and natural in an ANOVA framework. K-means techniques are simple, available in all major statistical packages, and cluster interpretation complements the IIT perspective. However, there are other ways to test for individual differences in experimental data in general and IIT in particular, e.g., multilevel regression. “The repeated-measures ANOVA model is merely a restricted version of the multilevel or general linear mixed model” (Hoffman & Rovine, 2007; pp.102-103), so one might code the factors and interactions as sets of dummy variables, serving as model predictors. The multilevel regression can be expanded to a latent class multilevel regression by replacing the assumption that the relationship between predictors (the dummies) and outcome is the same for all participants with the assumption that participants can come from different unobserved subpopulations, each characterized by class-specific predictor-outcome relationships (or class-specific integration patterns). The probabilistic model of latent class analysis allows for statistical comparison of competing models, which cluster analysis cannot do. However, while latent class analysis relaxes some assumptions of K-means analysis (Meyer & Morin, 2016), it has other assumptions, e.g., regarding data distribution, and their violation can greatly affect cluster recovery (Steinley & Brusco, 2011a), leading to lively debate about the relative advantages and disadvantages of the two approaches (e.g., Steinley & Brusco, 2011b). We focus on K-means here, because it is less complex and more accessible.

K-means analysis applied to IIT results in clusters of people with similar patterns of ratings, however, these ratings reflect a number of different processes: How people subjectively value the stimuli, how they then integrate these values, and how they finally translate this integrated internal response into an outward rating (Anderson, 1981, 1982, 1991, 1996). Hofmans & Mullet (2013) showed how one may pre-process the data in various ways before clustering, to filter out some of these functions and target specific sources of individual differences. Some such approaches are discussed later.

**‘Top Down’ Cluster Analysis: Confirmatory Search for Pre-Defined Patterns**

Our first question was how well the automatic cluster classification approximates the above solution found through inspection of individual means patterns. Thus, clusters were sought only in the blue x yellow interaction pattern, which might be additive, risk-seeking or risk-averse, as defined earlier. The by-hand classification used the raw data, considering whether each pattern was parallel or not. Absolute values were disregarded, because the main interest was children’s ability to integrate component EVs additively, not individual differences in assessment of probabilities and values. Cluster analysis of raw data, in contrast, finds clusters in absolute co-ordinate space, so for instance, might distinguish a cluster with generally high values from another cluster with low values.

To get around this difference between approaches, individuals’ data were standardized before cluster analysis. After standardization, clusters tend to differ mainly in pattern, means being equated. Still, while both approaches disregard differences in absolute values, standardization additionally equates variability. When raw judgments have small range, their patterns tend to disappear in the by-hand classification of means differences, but are amplified in standardized data. It is therefore not clear how well the approaches can correspond.

After standardisation, to search for theoretical, pre-defined patterns, initial cluster centres were set to three ideal patterns. To simulate as closely as possible the by-hand approach, these ideal patterns were the empirical yellow x blue patterns computed from the data, -.92, .31, .63, 1.24 (risk-averse), -1.20, .71, -.42, .91 (risk-seeking), and -1,10, .52, -.51, 1.09 (additive). The 14 patterns unclassifiable by hand were not included in the automatic classification either.

Figure 1 above compares the three-cluster solutions from the by-hand (top panels) and automatic classifications (bottom). The data patterns identified by the two approaches are similar. The distribution of participants across clusters is comparable. Age trends are also similar, additive participants increase with age under both classifications.

The two classifications agree for 68% of individuals (Table 1; chance agreement 33%; corrected contingency coefficient C/Cmax = .75). No risk-averse participant under one approach was risk-seeking under the other, rather all disagreements concerned the risk-averse/additive and risk-seeking/additive boundaries. These boundaries in the by-hand approach were set arbitrarily, and fairly conservatively, to limit allowable deviations from normative parallelism to no more than 1 point on the 17-point scale. The cluster analysis, in contrast, adjusts cluster membership automatically, minimizing distance within and maximizing distance between clusters; this seems a more principled approach. In 7 cases, automatic classification as additive was more lenient, in 12 cases, it was more conservative than the by-hand classification.

*Table 1: Agreement between by-hand classification (rows) and cluster analysis (columns).*

| | Confirmatory Cluster Analysis | | | |
|---|---|---|---|---|
| Means-based Classification | Additive | Risk Averse | Risk Seeking | Total |
| Additive | **11** | 3 | 9 | 23 |
| Risk Averse | 2 | **10** | 0 | 12 |
| Risk Seeking | 5 | 0 | **19** | 24 |
| Total | 18 | 13 | 28 | 59 |

The automatic cluster classification and the by-hand classification generally agree well. If the goal is to find groups of participants conforming to particular, pre-defined patterns, then cluster analysis with initial cluster centres set to theoretical values provides an efficient approach that can yield solutions similar to by-hand classification of means patterns. Note that the agreement found here might be the upper limit: For maximum comparability, we treated the same children as unclassifiable under both approaches, and set initial cluster centres to match the means-based classification. However, further simulations with different initial cluster centres did not reduce agreement.

The role of children excluded from the classification requires discussion: As mentioned, confirmatory procedures have the disadvantage of only finding patterns they are looking for. If many children defy classification, however, it signals that something may have been missed. The by-hand classification excluded 14 children, 9 younger ones, typically with very irregular patterns, which does not seem excessive. Cluster analysis, however, normally includes all participants. A way around this is to set K to find more clusters than the three theoretically defined ones, allowing unpredicted patterns/outliers to emerge as additional “break out” clusters. This approach is taken below.

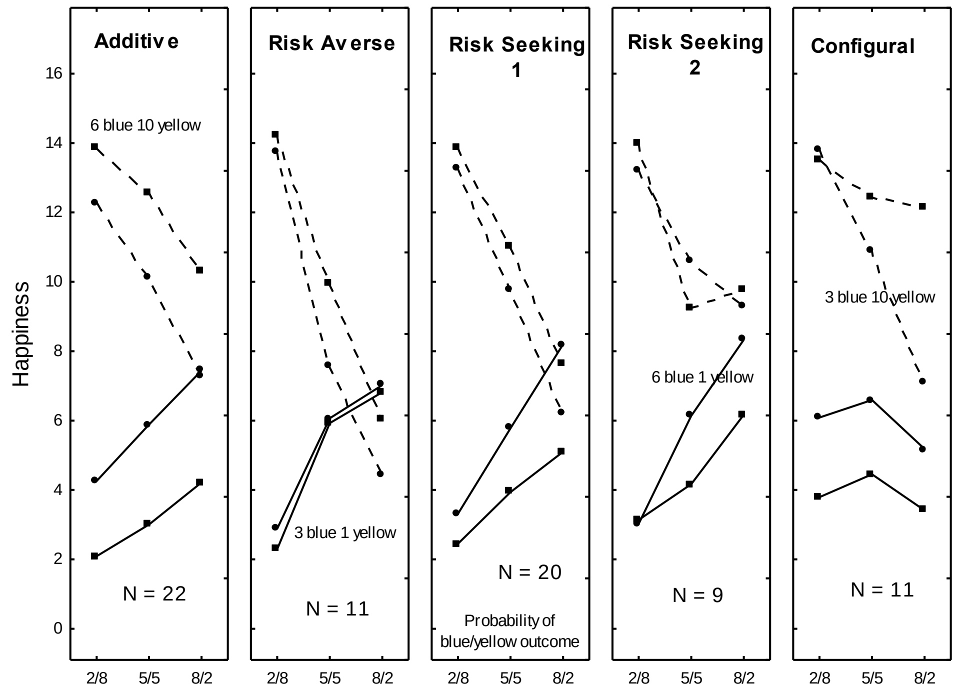
A different approach to finding participants cannot be classified is the 2-step procedure advocated by Hofmans and Mullet (2013), which assumes that individual differences in scale values relate to individual differences in integration. Under this approach, in a first step, clusters are sought for probability or value main effects separately, with integration pattern considered only in a second step. Disordinal participants at the first stage are eliminated, assuming that children who cannot rank order probabilities or values lack task understanding.

This assumption may not always hold: Children may have good understanding of EV, yet idiosyncratic ways of assessing prizes. We attempted the 2-step procedure here, and while all but two younger children’s judgments consistently increased with probability of the larger valued outcome, two disordinal value clusters emerged: Seven children judged 6/1 lower than 3/1 crayon combinations and six children judged 6/10 lower than 3/10 combinations. While these disordinalities seem striking from the adult perspective, children may simply prefer the colours of the 3-crayon bundle over those of the 6-crayon bundle, or prefer the excitement of the riskier 3/10 crayon games to the relative security of the 6/10 crayon games; some children said as much. Thus, unorthodox valuation need not indicate lack of EV understanding; in line with this view, these “odd” clusters did not show disordered integration per se.

**‘Bottom-up’ Cluster Analysis: Exploratory Search**

We contrasted the above, largely theory-driven cluster analysis with another pursuing a data-driven approach: No-one was excluded, data were clustered on the complete probability x blue x yellow pattern, more than three clusters were considered, no theoretical profile was set to search for particular patterns. The goal was to explore how the data partition naturally, independent of theoretical preconception/restriction. Again, we used standardized data, to focus on differences in integration and weighting.

A solution with 5 clusters emerged (Figure 2), including additive, risk-averse and two risk-seeking clusters, with Risk-Seeking 2 lacking any regular effect of blue, when it is the riskless option. In the unpredicted, configural cluster, children show a sizable, appropriate probability effect only for the largest difference between alternative prizes (3/10 crayons), with fairly flat curves, without clear probability effect elsewhere. This approach, at best a functional approximation of EV, appears at the youngest age, replaced by risk-seeking as modal pattern among older children, and additivity for adults.

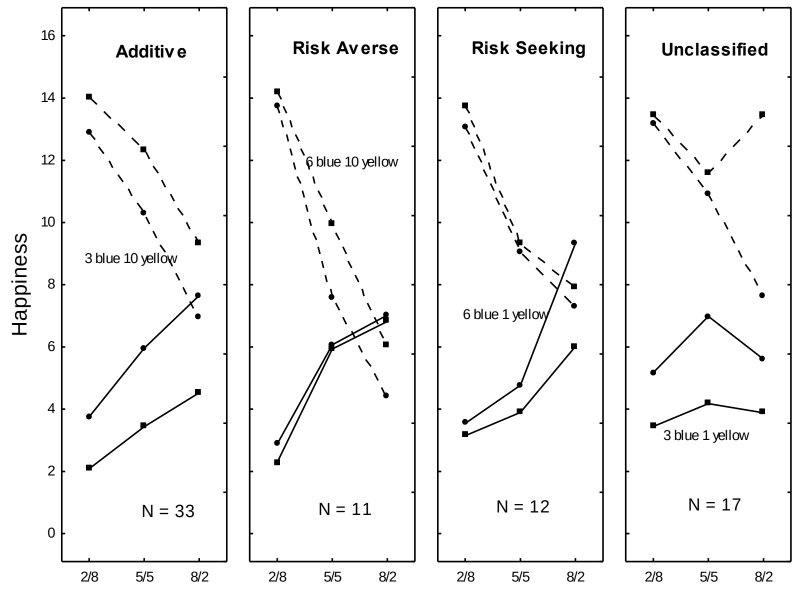



| | Additive | Risk Averse | Risk Seeking1 | Risk Seeking2 | Configural |
|---|---|---|---|---|---|
| Younger (n=28) | 4 (14%) | 4 (14%) | 6 (21%) | 4 (14%) | **10 (36%)** |
| Older (n=28) | 4 (14%) | 6 (21%) | **12 (43%)** | **5 (18%)** | 1 (4%) |
| Adults (n=17) | **14 (82%)** | 1 (6%) | 2 (12%) | 0 | 0 |
| Total (n=73) | **22 (30%)** | 11 (15%) | 20 (27%) | 9(12%) | 11 (15%) |

*Figure 2: Five clusters from the exploratory analysis including all participants.*

K was then increased to check for potential outliers at the cluster margins, i.e., children with disordered integration pattern, either reflecting very noisy judgments or minimal understanding. This approach identified 17 outliers in a solution with 5 large and 11 very small clusters. These were removed, and the analysis redone. This time a 3-cluster solution emerged (Figure 3). Pattern and trends were largely unchanged, with additive, risk averse and risk seeking patterns, and development towards additivity. The mean outlier pattern was close to the configural pattern identified before.

The exploratory analyses thus provide novel information on some very young children with unpredicted, irregular patterns. These patterns could not be read individually, but contained information hidden in noise: As a group, these children judged games by their outcome levels, with probability effect only for the largest contrast in outcome values. The EV model predicts the steepest probability effect in this case, but objectively the 3/10 contrast is not much beyond 1/6, and the latter produced almost flat ratings. Subjective need not equal numerical values, however, and children may see the 10-crayon bundle as far superior, also suggested by upward displacement of the top curves. This may make only the 3/10 games truly exciting, capable of producing happiness or disappointment. These children are, in a sense, thrill seekers. That finding a thrill improves attention to probability, of course, requires more research.




| Age | Additive | Risk Averse | Risk Seeking | Unclassified | Total |
|---|---|---|---|---|---|
| Younger | 8 (29%) | 4 (14%) | 4 (14%) | **12 (43%)** | 28 (100%) |
| Older | **12 (43%)** | 6 (21%) | 7 (25%) | 3 (11%) | 28 (100%) |
| Adults | **13 (76%)** | 1 (6%) | 1 (6%) | 2 (12%) | 17 (100%) |
| Total | **33 (45%)** | 11(15%) | 12(16%) | 17(23%) | 73(100%) |

*Figure 3: Three clusters after 17 outliers were removed.*

The exploratory analysis created uncertainty over the prevalence of risk-seeking, which reduced after outlier removal, while additivity increased (compare frequencies in Figures 2 and 3). That outlier removal highlights early competence is expected. However, the additive cluster of Figure 3 shows less separation of curves in the top than bottom pair and than Figure 2, so is in itself shifted towards risk seeking, making the cluster distinction less clear-cut. For that reason, we consider the whole-sample solution preferable. It also fits better with Levin and Hart (2003) whose child-adapted choice measure of risk attitude found high levels of risk-seeking, with children generally preferring gambles to sure gains of equal EV, in a game context similar to ours. However, following Tversky and Kahneman’s (1981) work on Asian disease problems, the adult literature often assumes that gain/save frames, as studied here, should not elicit pronounced risk-seeking, in line with our solution without outliers. Theory does not always support a clear decision; we should look for replicability next.

Children’s risk attitudes, like adults, vary with person and situation, differing within children as a function of frame  (Levin & Hart, 2003; Schlottmann & Tring, 2005), or whether children make judgments or choices (Schlottmann & Tring, 2005). Levin and Hart’s (2003) risk attitudes were stable over 3 years (Levin et al., 2007) correlating with personality variables and parental risk attitudes, so the usefulness of their choice measure is beyond doubt. Nevertheless, the processes underlying choice between options may well differ from those underlying judgment of a single option, as suggested by preference reversals (Hsee et al., 1999; Slovic, 1995).

In sum, our analyses confirm the suitability and practicality of cluster analysis to study individual differences in JDM: Results agreed with previous approaches that children differ in risk attitude, with developmental trend towards additivity; the result is reliable, re-appearing under different procedures, and regardless of how outliers are treated. Besides testing theoretical views, the cluster approach can also explore unpredicted patterns of individual differences, highlighting here that some young children consider probability only in some circumstances.

**Judgment and Choice in Save and Loss Frames**

Our second example applies cluster analysis to study framing effects in children's judgment and choice (Schlottmann et al., 2010). Cluster analysis has explored individual differences in framing effects on adults' ethical judgments (Muñoz Sastre et al., 2010), but the approach is novel to the domain and age here. The judgment task was similar to the gamble task already considered, except that jellybeans were the prizes. Children again judged, on a stick scale, happiness with each gamble, but here the puppet never gambled to win, but to keep/lose a stake. Thus, we considered children's loss processing, and whether this is as sophisticated structurally as their gains processing. Many argue that loss processing is more complex, being emotionally more taxing and inducing more conflict (e.g., Connolly & Zeelenberg, 2002; Kermer et al., 2006; Mellers et al., 1999; Lopes, 1987, Tom et al., 2007).

The same gambles were either framed positively, in terms of what could be saved, or negatively, in terms of what could be lost. Both frames were set in the context of puppet/child having to share a windfall of sweeties with the experimenter, with the gambles introduced to make this more fun. If the marble landed on yellow, the puppet/child would keep a whole bag of jellybeans, but if it landed on blue, they would have to share and lose some. The design factorially combined 4 levels of risk (0.25 to 1, so risky and sure losses were involved) with 4 levels of how many jellybeans were at risk (from 5 to all 20 jelly beans in each bag). Four- to 5-year-olds (Reception class), 5- to 6-year-olds (Year 1), 7- to 8-year-olds (Year 3) and adults, 96 in total, approximately half girls, judged two frames in two sessions about a week apart, following a procedure similar to Schlottmann and Tring (2005).

Besides judgments, each session also contained choices in the corresponding frame. In particular, children could give away some jellybeans for sure, or they could gamble to keep all 20 or none at all, with specified risk, in a 3 (level of risk in the gamble) x 3 (number given away in the sure option) factorial. Some choices involved equal EV options, as in standard framing tests. Whether the positive or negative frame session came first, and whether choice or judgment came first within session was counterbalanced; this had only minor effects, so is ignored.

At the group level, judgments and choices at all ages varied appropriately with risks and values involved; framing effects were minor. For judgment, loss processing was mostly as sophisticated as gain processing: All ages, in both frames, showed fanning, as expected under the normative EV model, although the youngest children discriminated less than older participants between different values, especially for losses. Also, all ages judged games described positively slightly higher than the same games described negatively, consistent with loss aversion (Kahneman et al., 1991). When judgments of gambles were compared to those for sure things of equal EV, gambles were typically judged slightly higher than corresponding sure things in loss, but not save trials, a pattern consistent with traditional measures of risk attitudes, however, frame differences were very small.

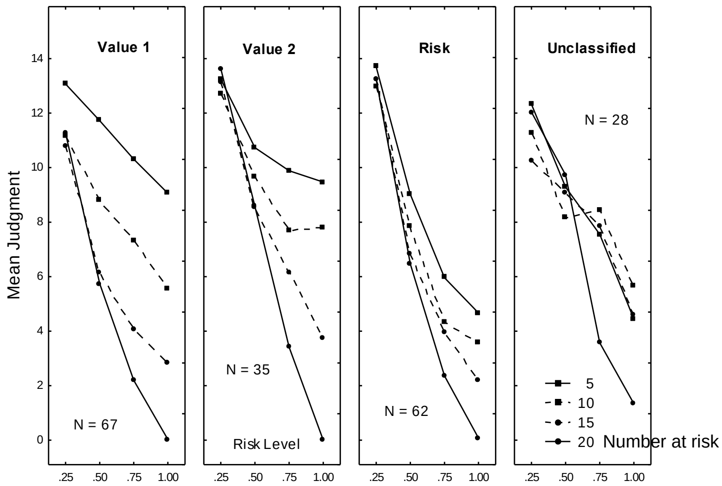
In the choice task, sure thing preference increased as risk in the gamble increased, and as the number lost for sure decreased. In theory, categorical choices maximize EV: Participants should always choose the option with smaller loss, with 50% choices for equal EV options. As in other studies, choices were more graded (e.g., Shanks et al., 2002). Overall, there were slightly more sure thing choices in save than loss frame, again consistent with the traditional view. This pattern appeared at 3 of 4 ages, but, as in judgment, frame differences were small; only reception children preferred the gamble in the loss frame. Frame differences were no larger for equal EV choice trials, as in standard framing tests.

We believe that frame differences are small here, because dimensional variation of risks and values, in judgment or choice, highlights the structure of the problem space for children, constraining responses, helping them construct appropriate strategies and minimizing the influence of non-normative person or decision factors, so often reported with isolated or ambiguous stimulus presentations. Person and situational factors may interact, of course, so again we need to consider individuals. We cannot, however, use single-subject ANOVA here: To get data for two frames and two tasks, each child saw only one replication; lengthier sessions would have been difficult to sustain. We could attempt means-based classification, but single replication data are often not stable enough for this, a difficulty amplified here because frame effects should be smaller than the main effects of probabilities or values.

**Cluster Analysis of Children's EV Judgment about Losses**

Cluster analysis can be done on single replication data, and because it considers the whole data set, it is less affected than a pure single-subject approach. It thus provides a unique tool here. As we did not know how individual differences might manifest, our approach was exploratory. Since frame differences should affect how participants evaluate risks and values, the raw data were analyzed.

Preliminary analyses of the judgments found three sizable clusters in save and loss frame, plus 12 outlier clusters involving 14 highly disordinal participants, 11 of which were reception children. These were eliminated. As the main clusters for each frame were similar, the analysis was re-done conjointly, with each participant providing two units of observations, from save and loss frame, to simplify and stabilize the solution. The same clusters re-appeared; Figure 4 shows the EV judgments as a function of risk level (horizontal) and number at risk if the blue outcome occurs (curve factor).




| | Value 1 | Value 2 | Risk | Unclassified | Total Observations (2 per child) |
|---|---|---|---|---|---|
| Reception (4-5 years) | 4 (8%) | 6 (13%) | **16 (33%)** | **22 (46%)** | 48 (100%) |
| Year 1 (5-6 years) | 16 (33%) | 16 (33%) | 16 (33%) | 0 | 48 (100%) |
| Year 3 (7-8 years) | 11 (23%) | 8 (17%) | **25 (52%)** | **4 (8%)** | 48 (100%) |
| Adults | **36 (75%)** | 5 (10%) | 5 (10%) | 2 (4%) | 48 (100%) |
| Total | 67(35%) | 35(18%) | 62(32%) | 28(15%) | 192(100%) |

*Figure 4: Clusters for the judgment data of the loss study, with save and loss frame observations analyzed conjointly, i.e., each child provided two observations; probability of the blue outcome (risk level) is on the horizontal, and the number at risk on blue is the curve factor. The pattern of not classified children is shown on the right.*

All three clusters show normative curve divergence indicative of EV multiplication, with strong risk effect (slope). They differ, however, in weighting of the value dimension: The "Risk" cluster shows only limited curve separation, while participants in the other two clusters emphasized the number at risk more. The Value1 cluster has categorically reduced ratings if more than 5 jellybeans are at risk, whereas EVs converge at low risk in the Value2 pattern. Both patterns contrast with low ratings for high risk games even if only a few are at risk in the Risk cluster. Eliminated children have a pattern similar to the risk cluster, but with less distinction of different values, mainly separating all from some at risk.

The age distribution in Figure 4 shows a strong developmental trend: Children were more often in the risk cluster (and the youngest children were often disordinal), while 75% of adults were in the Value1 cluster. Tthis is not the most normative approach: it has a configural element, with the smallest number at risk separated beyond what a pure EV strategy would predict. Adults were also twice as likely as children to be consistent across frames (Table 2).

*Table 2: Number (and %) of participants changing cluster affiliation from save to loss frame; children changed more often, the typical pattern is a change to the risk cluster in the loss frame.*

| Age Group | Value to Risk Cluster | Change between Value Clusters | Risk to Value Cluster | No Change «resistant to framing» | Unclassified Children | Total |
|---|---|---|---|---|---|---|
| Reception | 4 (17%) | 3 (13%) | 0 | 6 (25%) | **11 (46%)** | 24 (100%) |
| Year 1 | **13 (54%)** | 2 (8%) | 1 (4%) | 8 (33%) | 0 | 24 (100%) |
| Year 3 | **10 (42%)** | 2 (8%) | 3 (13%) | 7 (29%) | 2 (8%) | 24 (100%) |
| Adults | 4 (17%) | 4 (17%) | 1 (4%) | **14 (58%)** | 1 (4%) | 24 (100%) |
| Total | 31 (38%) | 11(13%) | 5 (6%) | 35 (43%) | 24 (25%) | 96 (100%) |
| Frame Effect (Save – Loss) | +1.64 | +1.08 | -0.08 | +0.58 | +1.21 | |

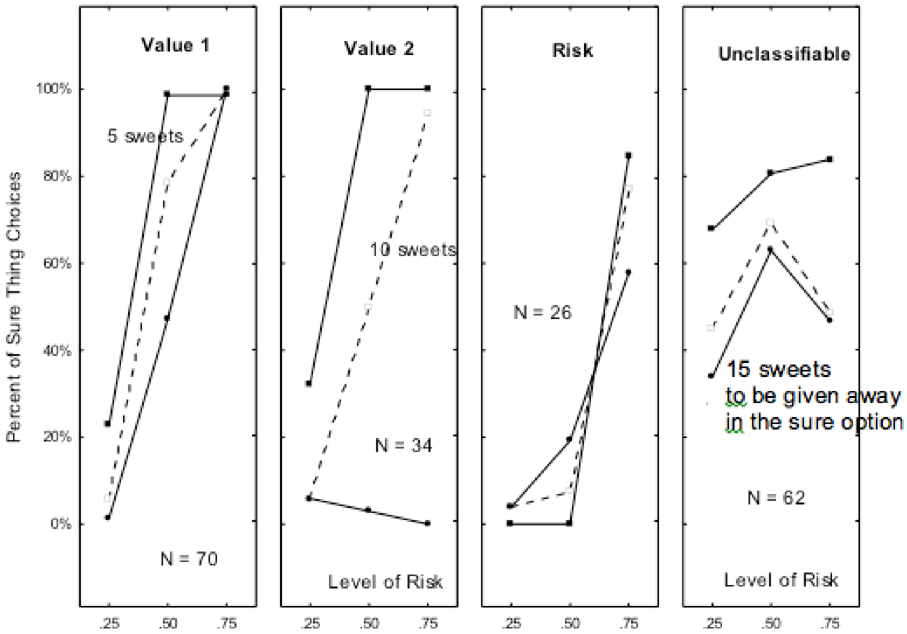
Importantly, changes from save to loss frame in Table 2 were systematic: Many participants were "resistant to framing" (Parker & Fischhoff, 2005), including a quarter or more of the children, but if participants changed between clusters, they were far more likely to change from a value-oriented cluster for saves to the risk cluster for losses, rather than the reverse, or than changing between value clusters. Participants thus attended more to risk in the loss than the save frame. This fits with standard views on framing. People do not just prefer risk, but attend more to it when losses are described.

The change in cluster affiliation also predicted the overall frame difference: Participants changing from a value to the risk cluster showed the strongest frame effect, with save frame judgments about 1.64 higher than loss frame judgments. Thus, individual differences in cluster membership were associated with clear and meaningful age and frame differences difficult to see at group level.

**Cluster Analysis of Children's Choices**

The choice data were less stable than children's judgments. We used the same approach as for judgments, classifying all 192 data patterns conjointly, with 2 observations per participant. Three main clusters appeared again, plus 62 unclassifiable profiles, distributed over 12 clusters. Fifteen participants had irregular patterns in both frames, 17 only in the save frame, 15 in the loss frame. Most were from the younger two age groups, suggesting that children had more difficulty in the choice than judgment task, consistent with past research (Schlottmann & Wilkening, 2011). This limits the usefulness of the cluster analysis, amplifying the low resolution inherent in choice tasks.

Figure 5 shows the three main clusters. As in judgment, all three show strong risk effect (slope), but in the "Risk" cluster (third panel) participants attended almost exclusively to risk, while the other two clusters were more value-oriented, with participants also choosing the sure thing more if the number to be given away for sure decreased (curve factor). In the largest Value1 cluster, the risk effect remains far stronger than the value effect. In the Value2 cluster, the value effect is larger, participants are closer to the indifference line for equal EV choices, with a closer-to-normative strategy, that higher EV options should be preferred categorically, regardless of degree of difference between options. While this is the "best" pattern, it is less clear than the EV pattern seen for judgment in far more participants. The mean unclassifiable pattern on the right is the only pattern showing no risk effect.




| | Value 1 | Value 2 | Risk | Unclassified | Total |
|---|---|---|---|---|---|
| Reception | 6 (12%) | 10 (21%) | 9 (19%) | **23 (48%)** | 48 (100%) |
| Year 1 | 18 (37%) | 5 (10%) | 7 (15%) | 18 (38%) | 48 (100%) |
| Year 3 | **22 (46%)** | 5 (10%) | 8 (17%) | 13 (27%) | 48 (100%) |
| Adults | **24 (50%)** | 14 (29%) | 2 (4%) | 8 (17%) | 48 (100%) |
| Total | 70(36%) | 34(18%) | 26(14%) | 62(32%) | 192(100%) |

*Figure 5: Three clusters for the choice data, plus the mean unclassifiable pattern.*

Judgment and choice thus share two features: Individual differences appear in children's consideration of value, not risk. Also, younger children have more variable approach than older children and adults. Both findings echo our earlier study on gains processing. Early focus on probability may seem surprising at first, but fits with work highlighting attention to probability in young children (Schlottmann & Anderson, 1991) and even infants (Denison & Xu. 2014; Teglas et al., 2007).

In contrast to judgment, no systematic frame differences appeared in choice cluster distribution (Table 3): The distribution is similar for saves and losses. However, individuals were consistent across frames in only 57% of cases. Agreement between judgment and choice was also low, appearing for 18% of observations. This may reflect the low stability of the choice data.

*Table 3: Number (and %) of participants in three clusters, in two frames, for judgment and choice. Clear frame differences appear for judgment, but not choice.*

| | | Value 1 | Value 2 | Risk | Total Observations |
|---|---|---|---|---|---|
| | Save | 37 (45) | 27 (33) | 18 (22) | 82 (100) |
| | Loss | 30 (37) | 8 (10) | 44 (54) | 82 (100) |
| Judgment | Total | 67 | 35 | 62 | 164 |
| | | | | | |
| | Save | 34(52) | 18(27) | 14(21) | 66(100) |
| | Loss | 36(67) | 16(30) | 12(22) | 54(100) |
| Choice | Total | 70 | 26 | 34 | 130 |

A better approach was to apply the classification of frame effects from the more stable judgment task to the choice task as well. Thus we considered frame differences in choice as a function of changes in cluster affiliation across frames in judgment (from Table 2). This turns out to also predict frame differences in the choice task, as seen in Figure 6.

The top left pair of panels show mean choices in the save and loss frame for 31 participants changing, in the judgment task, from a Value cluster in the save frame to the Risk cluster in the loss frame. These children's choices show a major shift, of 11%, towards the gamble in the loss frame as well. Children shifted mainly, however, when the alternative sure loss was sizable, with only the bottom two lines displaced downwards.

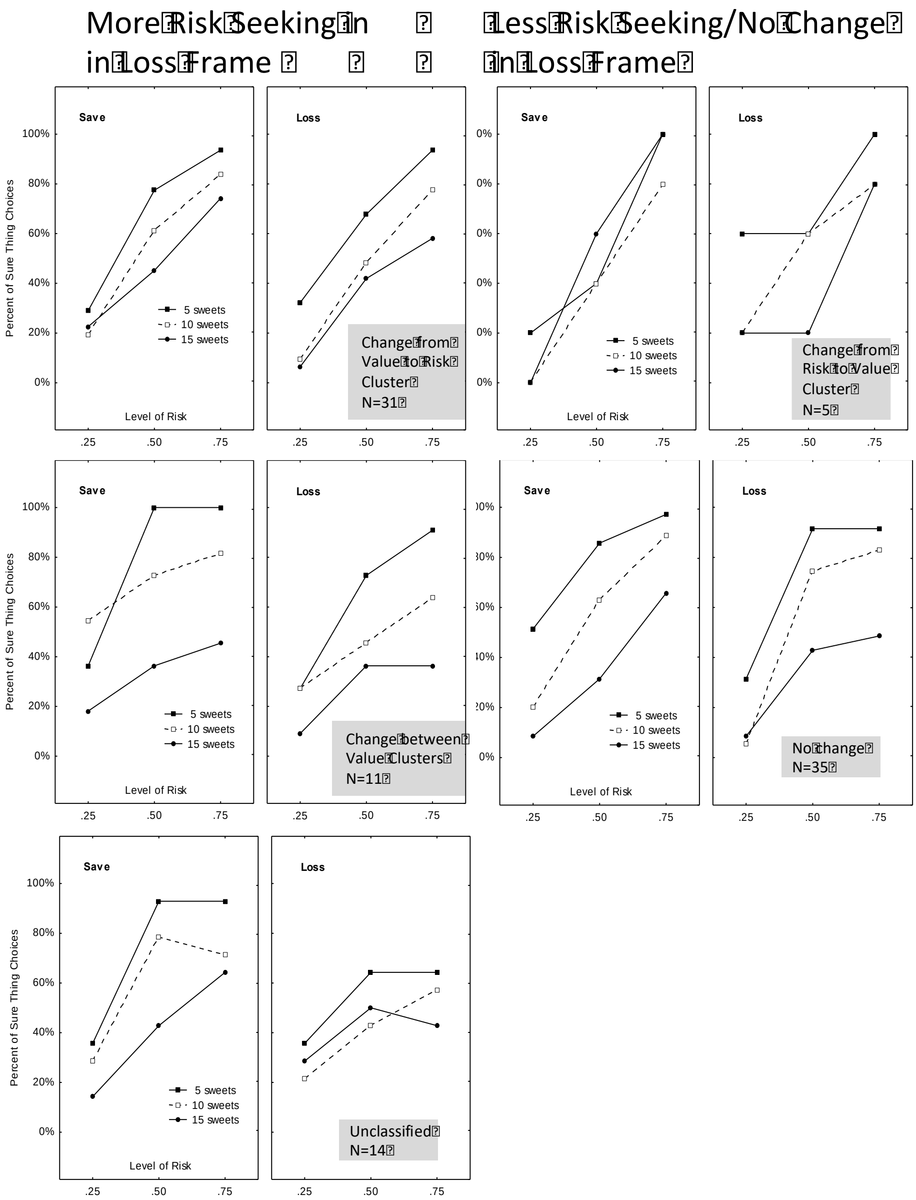



*Figure 6: The choice data in save and loss frame with participants grouped as a function of how they changed cluster affiliation from save to loss frame in the judgment task (grey boxes). Change in the judgment task predicts change in the choice task.*

For 5 participants changing in the opposite direction in the judgment task, from the Risk to a Value cluster (top right), choices shift in the opposite direction as well, with 9% more sure thing choices in the loss frame. The frame-resistant group in the judgment task (N=35) also shows little shift (1%) in their choices (second row, right).

This leaves 11 participants who moved between value clusters in judgment, and 14 unclassifiable children (bottom rows, left). Both groups show a downward shift towards risk-seeking choices for losses, 10 to 11%. While it is not clear why the first group shifts, it is perhaps no surprise that unclassifiable children, least able to consider the quantitative information, were most influenced by the qualitative frame information.

All in all 74% of participants showed consistent framing patterns in judgment and choice, either focusing more on risk, or less on risk, or showing no frame difference in both, with the remainder not classifiable or showing a frame difference that did not relate to risk in judgment. Thus, frame differences in judgment, defined as change in cluster affiliation, predicted frame differences in choice. These frame differences reflect an interaction of decision features and individual differences not apparent without the cluster analysis. Further work on this is necessary.

**Conclusions**

This chapter has illustrated how single-subject ANOVA, means-based classifications, and cluster analysis can be used to study the existence and pattern of systematic individual differences within IIT designs, in a way that aligns individual differences and processes established experimentally. In two data sets, one on judgment of gambles for gains, one on judgment and choice of gambles for losses, primary-aged children showed more pronounced individual differences than adults. These individual differences appeared as differential weight given to outcomes. In the gains task, children differed in the weight placed on risky and riskless outcome configurations, in the loss task they differed in the weight placed on the value dimension. Adults tended to weight probabilities and values more uniformly in both tasks. Individual differences thus interacted with age. They also interacted with frame: More children focused on risk in the loss frame. Moreover, change in cluster affiliation from save to loss frame in the judgment task predicted change in the choice task. The individual difference patterns found here were systematic, consistent between studies, consistent with the literature, and our findings are promising.

The present approach illustrates that it is possible, within an experimental approach, to separate systematic individual differences from error variability. Having done this, one might still find that individuals use strategies mostly corresponding to the group pattern, i.e., that there are few individual differences. This did not appear here, but in Muñoz Sastre et al (2010), for instance, one cluster captured 95% of participants. Or one might find that individuals appear unsystematic, as in the outliers eliminated here. When we grouped the data of eliminated participants, the group pattern showed clear effects, so these participants contribute information, not just noise. At the individual level, however, the signal cannot be read and standard single-subject analysis fails. Or one might find that participants are systematic, using a variety of strategies. These strategies might be distinct –the remaining 5% in Muñoz Sastre et al. (2010) fell into two categorically different groups. Alternatively, different strategies may seem graded variations on a continuum. This seems more the case in the present tasks, with clusters mainly differing in weighting of value. These are very different ways in which individuals may contribute to group patterns.

Assessment of individual differences in task performance is only a first step towards explaining them, but even this first step is often ignored in experiments, because of the difficulty of separating error from individual differences. That our clusters x frame interaction in the judgment task predicted frame differences in the choice task here suggests that our approach can potentially go to the second step, of relating individual differences in one JDM task to those in another.

Baron (2010) wrote cogently about the need to describe individual differences within experimental approaches, but worried about false positives in multiple single-subject analyses. In developmental studies, the danger is slim because single-subject tests are inherently low-powered with children. Developmentalists are often more worried about missing effects at individual level. Cluster analysis is a useful tool helping us circumvent the low power issue.

Nevertheless, the individual difference groupings identified by cluster analysis are not automatically real, but must, in the end prove their lasting value in validation studies linking the clusters to variables outside the classification data. Throughout this chapter we emphasized that the clusters were internally coherent, with clear process interpretations and developmental trends as also found in other domains (Gamelin et al., 2006, Zounon et al., 2015; Muñoz Sastre et al., 1998). Studies in other domains found meaningful learning effects, such that participants moved from less to more advanced clusters after feedback (Liegeois et al., 2003). Cluster solutions were reliable across sessions (Muñoz Sastre et al. 2014), replicable across different groups of participants (Morales et al., 2011; Nann et al., 2012) and situations (Mas et al., 2010), with clusters meeting expert (Muñoz Sastre et al., 2012) or cultural expectations (Mukashema et al., 2015; Mullet et al., in press).

The present approach has much potential as a tool for experimentalists to study person x situation and person x decision interactions in JDM, recommended by Appelt et al. (2011). The IIT approach, specifically, might be useful beyond JDM to study individual differences in any domain in which differences in valuation (the psychophysics of the stimuli) are conceptually distinct from differences in integration (combinatorial thinking), for instance, in children's intuitive physics (Wilkening & Huber, 2011). Cross-domain comparison of the extent of individual differences, especially in development, could inform debates on the domain-specificity of human cognition. Individual differences are important not just for assessment of individuals, but also for our understanding of processes, and we need experimental, not just psychometric approaches to study them. Here we have illustrated some methods that can help with this, by aligning group and individual-subject data.

## References


Anderson, N.H. (1981). *Foundations of Information Integration Theory*. New York: Academic Press.

Anderson, N.H. (1982). *Methods of Information Integration Theory*. New York: Academic Press.

Anderson, N.H. (1991). *Contributions To Information Integration Theory: Volumes 1, 2, and 3*. Mahwah, NJ: Erlbaum.

Anderson, N.H. (1996). *A Functional Theory of Cognition*. Mahwah, NJ: Erlbaum.

Anderson, N.H., & Schlottmann, A. (1991). Developmental study of personal probability. In N. H. Anderson (Ed.), *Contributions to Information Integration Theory*: *Vol. 3. Developmental* (pp. 111-134). Hillsdale, NJ: Erlbaum.

Appelt, K.C., Milch, K.F., Handgraaf, M.J., & Weber, E.U. (2011). The Decision Making Individual Differences Inventory and guidelines for the study of individual differences in judgment and decision-making research. *Judgment and Decision Making*, *6*(3), 252.

Baron, J. (2010). Looking at Individual Subjects in Research on Judgment and Decision Making (or anything). *Acta Psychologica Sinica*, *1*, 012.

Connolly, T., & Zeelenberg, M. (2002). Regret in decision making. *Current Directions in Psychological Science*, *11*(6), 212-216.

Denison, S. and Xu, F., 2014. The origins of probabilistic inference in human infants. *Cognition*, *130*(3), pp.335-347.

Falk, R., & Wilkening, F. (1998). Children's construction of fair chances: adjusting probabilities. *Developmental Psychology*, *34*(6), 1340.

Gamelin, A., Sastre, M.T. M., Sorum, P.C., & Mullet, E. (2006). Eliciting utilities using functional methodology: people’s disutilities for the adverse outcomes of cardiopulmonary resuscitation. *Quality of Life Research*, *15*(3), 429-439.

Hair, J.F., Black, W.C., Babin, B.J., & Anderson, R.E. (2008). *Multivariate Data Analysis (7th Edition).* Upper Saddle River, NJ: Prentice Hall.

Hoffman, L., & Rovine, M. J. (2007). Multilevel models for experimental psychologists: Foundations and illustrative examples. *Behavior Research Methods*, *39*, 107-117.

Hofmans, J., Ceulemans, E., Steinley, D., & Van Mechelen, I. (2015). On the added value of bootstrap analysis for K-means clustering. *Journal of Classification*, *32*, 268-284.

Hofmans, J., & Mullet, E. (2013). Towards unveiling individual differences in different stages of information processing: A clustering-based approach. *Quality & Quantity*, *47*(1), 455-464.

Hsee, C.K., Loewenstein, G.F., Blount, S., & Bazerman, M.H. (1999). Preference reversals between joint and separate evaluations of options: A review and theoretical analysis. *Psychological Bulletin*, *125*(5), 576

Kahneman, D., Knetsch, J.L., & Thaler, R.H. (1991). Anomalies: The endowment effect, loss aversion, and status quo bias. *The Journal of Economic Perspectives*, *5*(1), 193-206.

Kermer, D.A., Driver-Linn, E., Wilson, T.D., & Gilbert, D.T. (2006). Loss aversion is an affective forecasting error. *Psychological Science*, *17*(8), 649-653.

Kpanake, L., Patassi, A., & Mullet, E. (2013). Western African people’s views on the acceptability of criminal prosecution for sexual transmission of infectious diseases. *Sexually Transmitted Infections*, *89*, 290-294.

Léoni, V., Mullet, E. & Chasseigne, G. (2002). Aging and Intuitive Physics. *Acta Psychologica*, *111*, 29-43.

Levin, I.P., Hart, S.S., Weller, J.A., & Harshman, L.A. (2007). Stability of choices in a risky decision-making task: a 3-year longitudinal study with children and adults. *Journal of Behavioral Decision Making*, *20*(3), 241-252.

Levin, I.P., & Hart, S.S. (2003). Risk preferences in young children: Early evidence of individual differences in reaction to potential gains and losses. *Journal of Behavioral Decision Making*, *16*(5), 397-413.

Liégeois, L., Chasseigne, G.E., Papin, S., & Mullet, E. (2003). Improving high school students' understanding of potential difference in simple electric circuits. *International Journal of Science Education*, *25*(9), 1129-1145.

Lopes, L.L. (1976). Model-based decision and inference in stud poker. *Journal of Experimental Psychology: General*, *105*(3), 217.

Lopes, L.L. (1987). Between hope and fear: The psychology of risk. *Advances in Experimental Social Psychology*, *20*(3), 255-295.

Mas, C., Albaret, M.C., Sorum, P., & Mullet, E. (2010). French general practitioners vary in their attitudes toward treating terminally ill patients. *Palliative Medicine, 24, 60-67*.

Mellers, B., Schwartz, A., & Ritov, I. (1999). Emotion-based choice. *Journal of Experimental Psychology: General*, *128*(3), 332-345.

Meyer, J.P., & Morin, A. J.S. (2016). A person-centered approach to commitment research: Theory, research, and methodology. *Journal of Organizational Behavior*. DOI: 10.1002/job.2085

Morales, G., Lopez Ramirez, E., & Mullet, E. (2011). Acceptability of sexual relationships among people with learning disabilities: Family and professional caregivers' views in Mexico. *Sexuality and Disability, 29*, 165-174.

Mukashema, I., & Mullet, E. (2015). Attribution of guilt to offspring of perpetrators of the genocide: Rwandan people's perspectives. *Conflict Resolution Quarterly, 33,* 75-98.

Mullet, E., Kpanake, L., Kamble, S., Ahmed, R.A., Bugay, A., Muñoz Sastre, M.T., & Sorum, P.C. (2016). Mapping people's views regarding physician assisted suicide: A five-country study. *International Psychology Bulletin,* in press.

Muñoz Sastre, M.T. & Mullet, E. (1998). Evolution of the intuitive mastery of the relationship between base, exponent, and number magnitude in high school students. *Mathematical Cognition, 4*, 67-77

Muñoz Sastre, M.T., Sorum, P. C., & Mullet, E. (2012). Lay people's and health professionals' views about breaking bad news to children. *Child: Care, Health & Development, 40*, 106-114.

Muñoz Sastre, M.T., Doudou, K., Espitalier, A. L., Guyomard, E., Iturbide, M., Lampre, A., Tarroux, M., Sorum, P.C., & Mullet, E. (2014). The impact of current legislation on French people's views regarding physician-assisted-suicide. *Journal of Medical Law and Ethics, 2,* 5-13.

Muñoz Sastre, M.T., González, C., Lhermitte, A., Sorum, P.C., & Mullet, E. (2010). Do Ethical Judgments Depend on the Type of Response Scale? Comparing Acceptability versus Unacceptability Judgments in the Case of Life-Ending Procedures. *Psicológica* (2010), 31, 529-539.

Muñoz Sastre, M.T., Pecarisi, C., Legrain, E., Mullet, E., & Sorum, P.C. (2007). Acceptability in France of induced abortion for adolescents. *American Journal of Bioethics, 7*, 26-32.

Nann, S., Dousset, J.-P., Sok, C., Khim, P.Y.S., Sorum, P.C., & Mullet, E. (2012). Cambodian patients' and health professionals' views regarding the allocation of antiretroviral drugs. *Developing World Bioethics, 12,* 96-103.

Parker, A.M., & Fischhoff, B. (2005). Decision-making competence: External validation through an individual-differences approach. *Journal of Behavioral Decision Making*, *18*(1), 1-27.

Reyna, V.F., & Ellis, S.C. (1994). Fuzzy-trace theory and framing effects in children's risky decision making. *Psychological Science*, *5*(5), 275-279.

Siegler, R.S., Strauss, S., & Levin, I. (1981). Developmental sequences within and between concepts. *Monographs of the Society for Research in Child Development*, 1-84.

Schlottmann, A. (2001). Children's probability intuitions: Understanding the expected value of complex gambles. *Child Development*, *72*(1), 103-122.

Schlottmann, A., & Anderson, N.H. (1994). Children's judgments of expected value. *Developmental Psychology*, *30*(1), 56.

Schlottmann, A., & Anderson, N.H. (1993). An information integration approach to phenomenal causality. *Memory & Cognition*, *21*(6), 785-801.

Schlottmann, A., Reynolds, K., & Keane, L. (2010). A tale of two tasks: Judgment and choice in children's decision-making. British Psychological Society, Developmental Section Conference, Goldsmiths College, London, UK.

Schlottmann, A., & Tring, J. (2005). How children reason about gains and losses: Framing effects in judgement and choice. *Swiss Journal of Psychology*, *64*(3), 153-171.

Schlottmann, A., & Wilkening, F. (2011). Judgment and decision making in young children. In M.K. Dhami, A.Schlottmann & M.R. Waldmann (Eds.) *Judgment and Decision Making as a Skill: Learning, Development and Evolution*. Cambridge: Cambridge University Press.

Shanks, D.R., Tunney, R.J., & McCarthy, J.D. (2002). A re-examination of probability matching and rational choice. *Journal of Behavioral Decision Making*, *15*(3), 233-250.

Shanteau, J. C. (1970). An additive model for sequential decision making. *Journal of Experimental Psychology*, *85*(2), 181.
Shanteau, J. (1974). Component processes in risky decision making. *Journal of Experimental Psychology*, 103(4), 680-691.
Slovic, P. (1995). The construction of preference. *American Psychologist*, *50*(5), 364.
Steinley, D. (2003). Local optima in K-means clustering: what you don't know may hurt you. *Psychological Methods*, *8*(3), 294.
Steinley, D. (2006). K-means clustering: A half-century synthesis. *British Journal of Mathematical and Statistical Psychology*, *59*, 1-34.
Steinley, D., & Brusco, M.J. (2011a). Evaluating mixture modelling for clustering: Recommendations and cautions. *Psychological Methods*, *16(3)*, 63-79.
Steinley, D., & Brusco, M.J. (2011b). K-means clustering and mixture model clustering: Reply to McLachlan (2011) and Vermunt (2011). *Psychological Methods*, *16(3)*, 89-92.
Teglas, E., Girotto, V., Gonzalez, M. & Bonatti, L. (2007) Intuitions of probabilities shape expectations about the future at 12 months and beyond. *Proceedings of the National Academy of Sciences*, 104(48), 19156-19159
Tom, S.M., Fox, C.R., Trepel, C., & Poldrack, R.A. (2007). The neural basis of loss aversion in decision-making under risk. *Science*, *315*(5811), 515-518.
Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, *211*(4481), 453-458.
Wilkening, F., & Huber, S. (2011). Children's Intuitive Physics. In U.Goswami (Ed.) *Blackwell Handbook of Childhood Cognitive Development (2nd edition)*, 473-496. Chichester: Wiley.
Wilkening, F., & Anderson, N.H. (1982). Comparison of two rule-assessment methodologies for studying cognitive development and knowledge structure. *Psychological Bulletin*, *92*(1), 215.
Wilkening, F., & Anderson, N.H. (1991). Representation and diagnosis of knowledge structures in developmental psychology. In N.H.Anderson (Ed) *Contributions to Information Integration Theory* (Vol 3: Developmental) 45-80.
Zounon, O., Sorum, P.C., & Mullet, E. (2015). How people in Benin assess a couple's risk of having a baby with Sickle Cell Disease. *Journal of Community Genetics, 6*, 77-82.