Improving behaviour classification consistency: a technique from biological taxonomy

Serene Hyun-Jin Choi[a], Timo A. Nieminen[b], Mark Bahr[a] and Nan Bahr[a]

[a] *Graduate School of Education, The University of Queensland, QLD 4072, Australia*
serenechoi@bigpond.com, *phone +617 3365 6234, fax +617 3365 7199*
[b] *Centre for Biophotonics and Laser Science, Department of Physics, The University of Queensland, QLD 4072, Australia*
timo@physics.uq.edu.au, *phone +617 3365 2422, fax +617 3365 1242*

Abstract

Quantitative behaviour analysis requires the classification of behaviour to produce the basic data. In practice, much of this work will be performed by multiple observers, and maximising inter-observer consistency is of particular importance.

Another discipline where consistency in classification is vital is biological taxonomy. A classification tool of great utility, the binary key, is designed to simplify the classification decision process and ensure consistent identification of proper categories.

We show how this same decision-making tool – the binary key – can be used to promote consistency in the classification of behaviour. The construction of a binary key also ensures that the categories in which behaviour is classified are complete and non-overlapping. We discuss the general principles of design of binary keys, and illustrate their construction and use with a practical example from education research.

1 Introduction

Quantitative data analysis is a research tool that it would be difficult to overstate the usefulness of. Its power and generality as a method for testing the validity of hypotheses can be seen by the breadth and depth of its application in social, biological, and physical sciences. However, by its very nature, quantitative data analysis requires quantitative data. This, then, is the challenge often faced in education research – the reduction of observations to numerical data.

To analyse behaviour – for example, the frequency of, or duration of particular behaviours, or correlations between behaviours – the observed behaviour must be identified. This is fundamentally a classification process. The development of the categories into which observed behaviours are classified is widely discussed in the literature (Barlow & Hersen, 1984; Gittleman & Decker, 1994; Herbert & Attridge, 1975; Slater, 1978; Whitley, 2002). The categories should be:

1. Mutually exclusive. There should be no overlap between the categories – no behaviour should be classifiable into two separate categories.

2. Complete. The categories should form a complete or exhaustive set of the possible behaviours. It must be possible to classify every observed behaviour into a category. This does not mean that a large number of categories is required – a small number of sufficiently broad categories can be complete.

3. Usable. The categories must be understandable – terms used must be clearly understandable and well-defined. Definitions should be concisely and clearly stated. The names

given to the categories must be appropriate. The number of categories must also be appropriate for the intended research.

It is clearly important that the classification of behaviour is performed as accurately as possible. It is especially vital when different sets of behaviour are to be compared quantitatively, such as, for example, when comparing two groups of subjects, or analysing the change in behaviour of a single group of subjects over time. If there is a significant variation in the classification of behaviour, the quantitative measures of behaviour will vary, even if the observed behaviour remains the same; this is more likely to be a problem when different workers perform the actual classification for different sets of data. Interobserver consistency is not always simply achieved – a great deal of time and effort can be expended on training the observers in order to maximise consistency in classification by them (Meltzoff, 1998). Even if a single observer performs all of the classification in question, consistency over time is still vital. The importance of consistency is widely recognised, and inter-observer agreement (or inter-observer reliability, although strictly not a measure of reliability) is generally measured (Barlow & Hersen, 1984; Meltzoff, 1998; Mitchell & Jolley, 2001; Whitley, 2002).

The consistency achieved in classification, whether by a single observer or multiple observers is likely to depend on the method used for the act of classification. The literature on how the researcher can decide into which category an observed behaviour falls is virtually nonexistent. Nevertheless, this is obviously an issue of no small importance – the reduction of raw observation to quantitative data, and the analysis thereof, cannot proceed without it. The most common method in use appears to be for the researcher to refer to a list of definitions of the categories. Observation of this method in practical use shows that it is far from ideal. If the proper category is not immediately obvious, then the definitions of all the plausible categories need to be re-read, the behaviour re-observed, and so on, until a choice can be made. A great deal of difficulty results from ambiguous behaviours that appear to fit multiple categories. How can such behaviours be consistently classified? While these problems are usually minimised if the same researcher who devised the classification scheme is the observer who quantifies ("codes") the observed behaviours, in practice, much of the coding will be performed by multiple research assistants. Given the importance of inter-observer consistency, the need for a simple and reliable method for classification that will maximise consistency is obvious.

We note that the problem of easy, accurate, and consistent classification is general and multidisciplinary – classification decisions are important in many fields (Payne & Preece, 1980). One field where the problem of classification is critical is biological taxonomy. Organisms must be able to be classified correctly, even by workers with little training in classification or experience with organisms of the type in question. One of the standard tools designed to make this possible is the *binary key* (or *dichotomous key*), an identification key where decisions are made one at a time, and each question asked of the user of the key has only two possible answers. A simple illustrative example of such a key is shown in figure 1.

The most important feature of identification keys such as the one shown in figure 1 is that decisions are made one at a time. Each decision is much simpler than if all of the required decisions were grouped together, and had to be made at once (for example, as occurs when classifying by referring to a list of categories). Therefore, each simple decision is faster and more accurate, and as long as the number of decisions to be made is not too large, an identification key can be faster to use than a list of categories. The simplification is especially important in ambiguous cases – the classifier can concentrate on the single feature that divides the decision path, rather than having to simultaneously consider all observable characteristics.

These benefits of using binary keys for identification are not restricted to biological classifi-

Key to classes of living vertebrates (Subphylum *Vertebrata*)

1. Both fins and gills present in adults
   (a) Yes – go to 2
   (b) No – go to 4
2. Jaws present
   (a) Yes – go to 3
   (b) No – Class *Agnatha* (lampreys and hagfish)
3. Skeleton has bone
   (a) Yes – Class *Osteichthyes* (bony fish)
   (b) No (cartilage only) – Class *Chondrichthyes* (cartilaginous fish – sharks and rays)
4. Skin is naked, larvae have gills
   (a) Yes – Class *Amphibia* (amphibians – frogs, toads, salamanders, etc)
   (b) No (hair, feathers, or scales present) – go to 5
5. Skin has feathers, front limbs are wings
   (a) Yes – Class *Aves* (birds)
   (b) No – go to 6
6. Skin has hairs, milk glands produce milk for young
   (a) Yes – Class *Mammalia* (mammals)
   (b) No (skin has scales) – Class *Reptilia* (reptiles)

Figure 1: A binary key for the classification of common vertebrate animals.

cation (Payne & Preece, 1980); keys can be used advantageously in a variety of fields: medical diagnosis, machine fault location, pattern recognition, and, as we show here, quantitative behaviour analysis. This use of keys is not unknown in education and behaviour research. Bekoff (1977) notes some examples in animal behaviour research, and Gay (1996, pg 19) gives a key for classifying types of research.

2 Construction of a binary key

While a great deal has been written on design of optimum keys, automated key construction programs developed, etc (Dallwitz, 1974; Osborne, 1963a, 1963b; Payne & Preece, 1980), most applications in behaviour research will deal with a relatively small number of categories. In cases like these, it is easiest to simply construct our key by hand, rather than using mathematical or computer tools.

The basic principles of constructing a successful key are few (Metcalf, 1954): the classification decisions must be based on observables, with the most prominent observables dealt with earlier in the key rather than later, the choices at each point must be mutually exclusive, the number of decisions required for any identification (especially common identifications) should be as small as possible, and the key should be written as simply an clearly as possible. Supporting illustrations or examples can be used to clarify specific questions asked in the key. Note that mutual exclusivity is readily obtained in a binary key if the questions with yes/no answers can be asked. The shortness of the key is important, since the likelihood of error in answering the questions increases with the number of questions. A short key results from each decision dividing the relevant group of categories into two subgroups of comparable size.

The process of constructing of a key from a list of definitions can be summarised as:

1. Determine the observable features for each category.
2. Choose one observable that will provide a suitable starting point for the key. Ideally, the first decision in the key should be the least error-prone, and should divide the categories into two roughly equal groups.
3. Continue repeating the previous step for the remaining observables within a group of behaviours produced by a previous decision, until only one behaviour is left.
4. Check the observable features of the remaining behaviour category. If it is possible to reach this end-point of the key without the remaining observables (if any) being present, then the original list of categories must be incomplete. Create a new category of necessary.
5. Repeat steps 3 and 4 until all categories have been separated.

If two categories cannot be distinguished from each other by observable features, they must be combined into a single category. If a particular category contains two distinct sets of observables, the category can be split in two separate categories. This can result in a more logical classification scheme. However, the combined category may be more useful for the intended

---

Non-play
Absence of play

Exploratory
Explores play environment (examines or searches for play objects or watches play partner's play)

Functional Play
Obvious play
Uses play object
Play object used in purely physical manner

Constructive Play
Obvious play
Uses play object
Play object used in purely physical manner
Spatial arrangement of play object(s)

Pretend Play
Obvious play
Symbolic or representational elements to play

Game with Rules
Obvious play
Play involves rules
Plays with partner

Figure 2: List of observable features of behaviours

analysis. For the purposes of constructing the key, it is best to temporarily split the category, since the two sub-categories might be reachable by very different paths in the final key.

We illustrate this method by constructing a key for the classification of a set of behaviours. The example comes from a study of the effect of training on the ability of peer play partners to play with children with autism. One component of this research is the analysis of the cognitive level of play shown by the children with autism. The categories into which the researcher intends to classify the behaviour are shown in figure 2.

Non-play
: The target child shows undirected, problem, and stereotype behaviours, or general social interactions without play activities.

Exploratory
: The target child searches around or examines play objects in an ambiguous manner or sensory-motor activities.

Functional Play
: The target child uses play objects in repetitive motor movements, physically appropriate manner as its function denoted without any creativity or flexibility, or relational using more than two play objects with no symbolic representation.

Constructive Play
: The target child organizes play objects in some type of spatial format to design an object or a model of a real object.

Pretend Play
: The target child lets a play object or person symbolize or represent a thing or person that it is not, in a make believe manner.

Game with Rules
: The target child and play partner play together with a set of their own rules including obligations or prohibitions.

Figure 3: List of category definitions

Next, we list the observable features of the behaviours in each category (listed in figure 3).

The next step is to choose the first decision to be made in the key. It should be the least ambiguous decision, and should divide the behaviours as equally as possible. From the list in figure 3, we choose to see whether the behaviour is obviously play. We can begin writing the key:

1. Is the child obviously playing?
   (a) Yes – Functional, Constructive, Pretend or Game with Rules
   (b) No – Non-play or Exploratory

Since we expect non-play and exploratory behaviour to be frequent, it is useful to separate these behaviours next:

1. Is the child obviously playing?
   (a) Yes – Functional, Constructive, Pretend or Game with Rules

   (b) No – go to 2
2. Is the child exploring or examining play objects or the play partner's play?
   (a) Yes – Exploratory
   (b) No – Non-play

The purely physical use of play objects characterises both functional and constructive play, and provides a suitable next decision in the key:

1. Is the child obviously playing?
   (a) Yes – go to 3
   (b) No – go to 2
2. Is the child exploring or examining play objects or the play partner's play?
   (a) Yes – Exploratory
   (b) No – Non-play
3. Does the play consist only of play objects being used in a physical manner?
   (a) Yes – Functional or Constructive
   (b) No – Pretend or Game with Rules

Now we can simply complete the key:

1. Is the child obviously playing?
   (a) Yes – go to 3
   (b) No – go to 2
2. Is the child exploring or examining play objects or the play partner's play?
   (a) Yes – Exploratory
   (b) No – Non-play
3. Does the play consist only of play objects being used in a physical manner?
   (a) Yes – go to 4
   (b) No – go to 5
4. Are the play objects being organised spatially?
   (a) Yes – Constructive Play
   (b) No – Functional Play
5. Are rules obviously being followed?
   (a) Yes – Game with Rules
   (b) No – Pretend Play

The categories into which the play is now to be classified are not the same as the original categories. We note that the original definition of the game with rules category specified that the child plays with the play partner. This is not a requirement of the new category. If we were to use the original set of category definitions, it would not be overly difficult to imagine plausible behaviours that would, for example, simultaneously fit the pretend play, game with rules, and constructive play categories. If the coder works from the definition list, they must judge which of the categories the behaviour should be classified into. The key removes this ambiguity – in this case, the behaviour would be classified in the game with rules category. To code any observed behaviour using this key requires only two or three decisions to be made, depending on the particular behaviour. To code observed behaviour from the definition list requires six decisions, each with multiple sub-decisions, to determine whether or not the behaviour fits each category, or more, if ambiguity forces a second round of choices to determine the most appropriate category.

The key can be given either in one of the traditional list formats (Metcalf (1954) discusses the advantages and disadvantages of a variety of list formats), as shown in figure 4, or as a graphical tree, as shown in figure 5. The graphical tree format is only practical for short keys. Since the key is intended for use by various observers, with minimal specific training, the key is best accompanied by some note to clarify what is meant by each decision in the key (see figure 6).

1. Is the child obviously playing?
   (a) Yes – go to 3
   (b) No – go to 2
2. Is the child exploring or examining play objects or the play partner's play?
   (a) Yes – Exploratory
   (b) No – Non-play
3. Does the play consist only of play objects being used in a physical manner?
   (a) Yes – go to 4
   (b) No – go to 5
4. Are the play objects being organised spatially?
   (a) Yes – Constructive Play
   (b) No – Functional Play
5. Are rules obviously being followed?
   (a) Yes – Game with Rules
   (b) No – Pretend Play

Figure 4: Key for classification of cognitive level of play – list format

3 Conclusion

The use of a binary key for the classification of behaviour can improve the consistency of classification. Since the decisions required of the classifier are prioritised and simplified, less knowledge of the classification scheme being used is required. Thus, improved consistency and inter-observer agreement can be obtained, without the need for extensive training of observers.

In addition, the act of constructing a key guarantees that the categories into which the observed behaviour is to be classified are mutually exclusive, complete (ie exhaustive), and can be distinguished from each other by observation. These properties are required for any set of categories to be considered correct. The key can also warn the researcher if any single category contains multiple distinguishable behaviours. This strongly suggests that the construction of a key is a valuable tool for the validation of a classification scheme, even if there is no intent to use the key for actual classification work.

Lastly, we warn against regarding the binary key as a panacaea for classification difficulties. Ambiguous behaviour is still ambiguous behaviour, and while the key prioritises and simplifies the decisions that need to be made, behaviour that is on the borderline between two categories will always vex the classifier.

References

Barlow, D. H. and Hersen, M. (1984). *Single case experimental designs: strategies for studying behavior change*, 2nd ed., Pergamon Press, New York.

Bekoff, M. (1977). Three areas of classical ethology. pp 1–46 in Hazlett, B. A. (ed), *Quantitative methods in the study of animal behaviour*, Academic Press, New York.

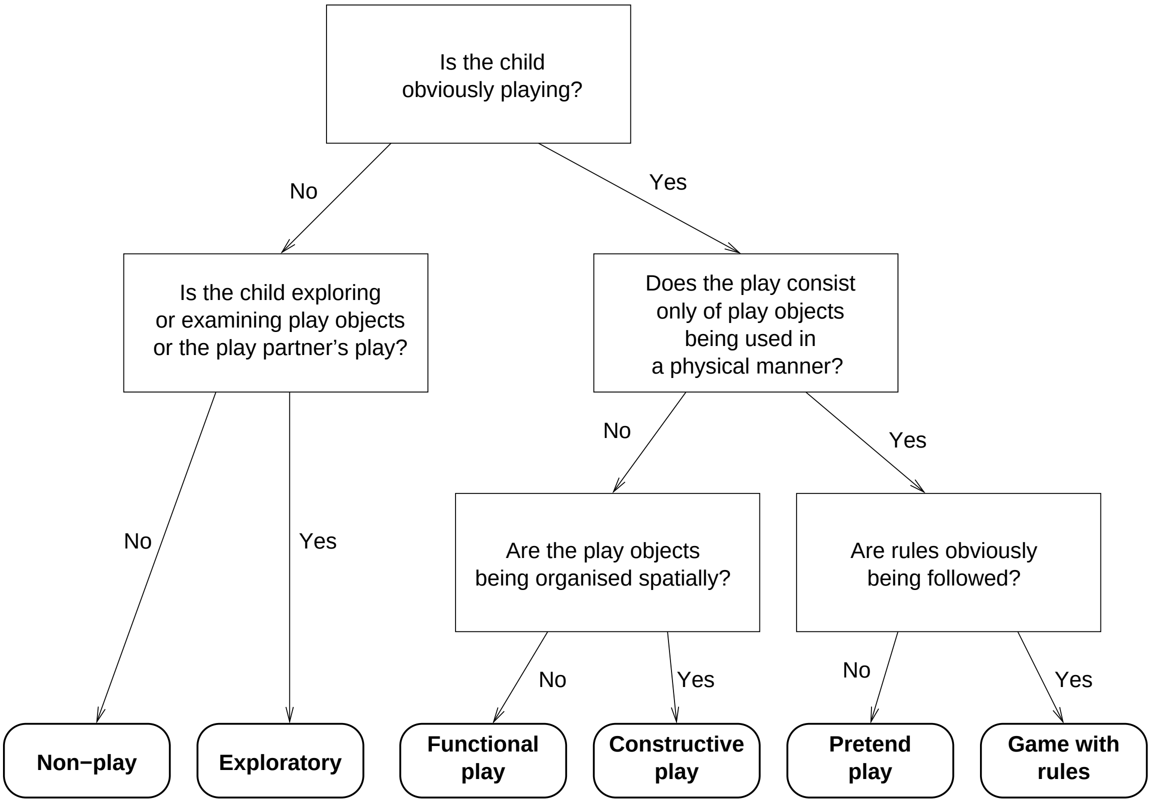

Figure 5: Key for classification of cognitive level of play – graphical tree format

Dallwitz, M. J. (1974). A flexible computer program for generating identification keys. *Syst. Zool.*, **23** 50–57.

Gay, L. R. (1996). *Educational research: competencies for analysis and application*, 5th ed., Prentice-Hall, Upper Saddle River, NJ.

Gittleman, J. L. and Decker, D. M. (1994). The phylogeny of behaviour. pp 80–105 in Slater, P. J. B. and Halliday, T. R. (eds), *Behaviour and Evolution*, Cambridge University Press, Cambridge, UK.

Herbert, J. and Attridge, C. (1975). A guide for developers and users of observation systems and manuals. *American Educational Research Journal*, **12** 1–20.

Meltzoff, J. (1998). *Critical thinking about research: psychology and related fields*, American Psychological Association, Washington, DC.

Metcalf, Z. P. (1954). The construction of keys. *Syst. Zool.*, **3** 38–45.

Mitchell, M. and Jolley, J. (2001). *Research design explained*, 4th ed., Harcourt, Fort Worth, TX.

Osborne, D. V. (1963a). A numerical representation for taxonomic keys. *New Phytologist*, **62** 35–43.

Osborne, D. V. (1963b). Some aspects of the theory of dichotomous keys. *New Phytologist*, **62** 144–160.

Is the child obviously playing?

If a positive decision cannot be made quickly, the child is not *obviously* playing. Therefore, choose "No." Problem behaviours such as self-directed or other violent behaviour are not play, even if play objects are used.

Is the child exploring or examining play objects or the play partner's play?

This includes watching, listening, touching and so on. If the child's attention is directed towards the play partner, the child must be watching the play partner's play, rather than the play partner.

Does the play consist only of play objects being used in a physical manner?

Play objects includes all things played with, such as the sand in a sandpit, as well as conventional toys. If any abstract, symbolic, or representational elements are present, decide "No." Examples include the use of wooden blocks to represent cars, play that assumes the presence of non-existent play objects, or non-existent properties of play objects. If no play objects are used, decide "No."

Are the play objects being organised spatially?

Examples include making piles of sand, stacking blocks and drawing a picture.

Are rules obviously being followed?

If a positive decision cannot be made quickly, rules are not *obviously* being followed. Therefore, choose "No." If either the child or play partner dictates the play according to whim, this doesn't count as rules. Decide "No."

Figure 6: Notes for the coder

Payne, R. W. and Preece, D. A. (1980). Identification keys and diagnostic tables: a review. *J. R. Statist. Soc. A*, **143** 253–292.

Slater, P. J. B. (1978). Data collection. pp 7–24 in Colgan, P. W. (ed), *Quantitative Ethology*, John Wiley, New York.

Whitley, B. E., Jr. (2002). *Principles of research in behavioral science*, 2nd ed., McGraw-Hill, Boston.