Tierärztliche Hochschule Hannover

Erstellung und Evaluierung eines Verhaltenstests zur Erkennung von Futteraggression gegenüber Menschen durch Hunde in einem Tierheim.

INAUGURAL - DISSERTATION

Zur Erlangung des Grades einer
Doktorin der Veterinärmedizin
- Doctor medicinae veterinariae -
(Dr. med. vet.)

vorgelegt von
Silvia Unglaube
Kapellendorf

Hannover 2015

Wissenschaftliche Betreuung: Prof. Dr. Hackbarth
Institut für Tierschutz und Verhalten
(Heim-, Labortiere und Pferde)

1. Gutachter: Prof. Dr. Hackbarth

2. Gutachter: PD Dr. rer. nat. Schmidt

Tag der mündlichen Prüfung: 12.11.2015

gefördert durch die Hans-Helga-Stiftung

FÜR MEINE FAMILIE

FÜR TINO UND ALICE

Inhaltsverzeichnis

1 Einleitung 1

2 Literatur 2

2.1 Kanine Aggression 2

2.1.1 Kanine Aggression aus Sicht der Gesellschaft 2

2.1.2 Agonistisches und aggressives Verhalten von Hunden 3

2.1.3 Klassifizierung von aggressivem Verhalten 7

2.1.4 Ressourcenbezogene Aggression 10

2.1.5 Futteraggression 11

2.2 Temperament- und Wesenstests 16

2.2.1 Grundidee von Verhaltenstests 16

2.2.2 Einsatz von Verhaltenstests in der Literatur 18

2.3 Evaluierung von Verhaltenstests 23

2.3.1 Evaluierung der Verhaltenstests in der Literatur 23

2.3.2 Standardisierung von Verhaltenstests 25

2.3.3 Reliabilität von Verhaltenstests 27

2.3.4 Validität von Verhaltenstests 31

2.3.5 Praktikabilität und Tierschutz 35

3 Material und Methoden 37

3.1 Überblick 37

3.2 Tiere 38

3.3 Testperson und Gutachter 39

3.4 Testgelände 41

3.5 Hilfsmittel 42

3.6 Testvorbereitung 44

3.7 Testablauf 44

3.7.1 Prätest 45

3.7.2 Test auf Futteraggression 48
3.7.3 Retest..... 53
3.7.4 Referenztest..... 54
3.8 Datenaufnahme und -auswertung 57
3.8.1 Auswertung des Prätests 57
3.8.2 Auswertung des Tests auf Futteraggression..... 57
3.8.3 Bestimmung der Reliabilität des Tests auf Futteraggression 60
3.8.4 Bestimmung der Kriteriumsvalidität des Tests auf Futteraggression..... 60
3.8.4.1 Auswertung des Datenerhebungsbogens……..61
3.8.4.2 Auswertung der Referenztests……....61
3.8.4.3 Auswertung der C-BARQ-Fragebögen……...…..62
3.8.5 Auswertung der Stresssignale…...……63
3.9. Statistische Methoden 63
4 Ergebnisse..... 66
4.1 Ergebnisse des Prätests..... 66
4.2. Ergebnisse des Tests auf Futteraggression 69
4.2.1 Verhalten in den Subtests 69
4.2.2 Inter-rater-Reliabilität des Tests auf Futteraggression 72
4.2.3 Intra-rater-Reliabilität des Tests auf Futteraggression 77
4.2.4 Test-Retest-Reliabilität des Tests auf Futteraggression..... 81
4.2.5 Kriteriumsvalidität des Tests auf Futteraggression..... 83
4.2.5.1 Datenerhebungsbogen..... 83
4.2.5.2 Referenztest..... 85
4.2.5.3 C-BARQ 87
4.3 Ergebnisse der Auswertung der Stresssignale 89
5 Diskussion 91
5.1 Der Prätest 91

5.2 Das Verhalten in den Subtests des Tests auf Futteraggression 92
5.3 Die Evaluierung des Tests auf Futteraggression 94
5.3.1 Zur Standardisierung und Objektivität des Tests 94
5.3.2 Die Reliabilität des Tests auf Futteraggression 97
5.3.2.1 Die Inter-rater-Reliabilität des Tests auf Futteraggression 97
5.3.2.2 Die Intra-rater-Reliabilität des Tests auf Futteraggression 102
5.3.2.3 Die Test-Retest-Reliabilität des Tests auf Futteraggression 104
5.3.3 Die Validität des Tests auf Futteraggression 105
5.3.3.1 Die Inhaltsvalidität des Tests auf Futteraggression 105
5.3.3.2 Die Kriteriumsvalidität des Tests auf Futteraggression 107
5.4 Die Stresssignale 111
5.5 Praktikabilität und Tierschutz 113
5.6 Schlussfolgerung 114
6 Zusammenfassung 118
7 Summary 122
8 Literaturverzeichnis i
9 Anhang xiv
9.1 Datenerhebungsbogen xiv
9.2 Ethogramm (Test auf Futteraggression) xviii
9.3 Auswertungstabellen (Test auf Futteraggression) xxi
9.3.1 Subtest 1 xxi
9.3.2 Subtest 2 xxii
9.3.3 Subtest 3 xxii
9.4 Ethogramm (Prätest) xxiii
9.5 Auswertungstabellen (Prätest) xxvi
9.5.1 Prätest 1 xxvi
9.5.2 Prätest 2 xxvii
Danksagung xxviii

Abbildungs- und Tabellenverzeichnis

Abbildung 1: Schematische Darstellung des Testgeländes…….....41

Abbildung 2: "Assess-A-Hand" nach STERNBERG (2002)…….....43

Abbildung 3: Handhabung der künstlichen Hand……..….43

Abbildung 4: Häufigkeiten der gezeigten Verhaltensweisen in Prätest 1 (insgesamt 270 Verhaltensweisen) und Prätest 2 (insgesamt 281 Verhaltensweisen). N=17 Hunde……..…67

Abbildung 5: Prozentualer Anteil der gezeigten Verhaltensweisen an der Gesamtheit der gezeigten Verhaltensweisen in Prätest 1 und 2. Verhaltensweisen aller Testsituationen zusammengefasst in Kategorien. Statistischer Vergleichs zwischen den einzelnen Kategorien von Prätest 1 und 2 als p-Werte des Wilcoxon-signed-rank-Tests. N=17 Hunde…….68

Abbildung 6: Verteilung von Häufigkeiten der Verhaltensweisen in Prätest 1 und 2, zusammengefasst in Kategorien. Boxplots mit Medianen, Interquartilsabständen sowie Minima und Maxima. N=17 Hunde.....68

Tabelle 1: Ausdrucksverhalten bei offensiv- und defensivbedingter Aggression (FEDDERSEN-PETERSEN, 2008, mod. nach ZIMEN, 1971)…….9

Tabelle 2: Ablauf Prätest 1……..46

Tabelle 3: Ablauf Prätest 2……..47

Tabelle 4: Ablauf Subtest 1…….49

Tabelle 5: Ablauf Subtest 2…….50

Tabelle 6: Ablauf Subtest 3…….51

Tabelle 7: Ablauf Referenztest Teil 1……..…55

Tabelle 8: Ablauf Referenztest Teil 2…….55

Tabelle 9: Ablauf Referenztest Teil 3…….56

Tabelle 10: Punktvergabe für Verhaltensweisen…….58

Tabelle 11: Kategorieneinteilung von Verhaltensweisen……59

Tabelle 12: Anzahl und prozentualer Anteil der von den drei Gutachtern erkannten Verhaltensweisen in Subtest 1. N=17 Hunde……69

Tabelle 13: Anzahl und prozentualer Anteil der von den drei Gutachtern erkannten Verhaltensweisen in Subtest 2. N=16 Hunde……70

Tabelle 14: Anzahl und prozentualer Anteil der von den drei Gutachtern erkannten Verhaltensweisen in Subtest 3. N=17 Hunde……...……70

Tabelle 15: Punkteergebnisse der drei Subtests sowie des Gesamtergebnisses der drei Gutachter für die einzelnen Hunde. N=17 Hunde…….71

Tabelle 16: Kategorienergebnisse der drei Subtests sowie die Gesamtkategorien der drei Gutachter für die einzelnen Hunde. N=17 Hunde…….72

Tabelle 17: p-Werte für die Übereinstimmung der Punktergebnisse zwischen den Gutachtern (Inter-rater) in Subtest 1. N=17 Hunde……..73

Tabelle 18: p-Werte für die Übereinstimmung der Punktergebnisse zwischen den Gutachtern (Inter-rater) in Subtest 2. N=16 Hunde……..73

Tabelle 19: p-Werte für die Übereinstimmung der Punktergebnisse zwischen den Gutachtern (Inter-rater) in Subtest 3. N=17 Hunde……..74

Tabelle 20: p-Werte für die Übereinstimmung der Punktergebnisse zwischen den Gutachtern (Inter-rater) im Gesamtergebnis. N=17 Hunde……74

Tabelle 21: p-Werte für die Übereinstimmung der Kategorien zwischen den Gutachtern (Inter-rater) in Subtest 1. N=17 Hunde……..75

Tabelle 22: κ-Werte für die Übereinstimmung der Kategorien zwischen den Gutachtern (Inter-rater) in Subtest 1. N=17 Hunde……..…75

Tabelle 23: p-Werte für die Übereinstimmung der Kategorien zwischen den Gutachtern (Inter-rater) in Subtest 2. N=16 Hunde……..75

Tabelle 24: κ-Werte für die Übereinstimmung der Kategorien zwischen den Gutachtern (Inter-rater) in Subtest 2. N=16 Hunde……..75

Tabelle 25: p-Werte für die Übereinstimmung der Kategorien zwischen den Gutachtern (Inter-rater) in Subtest 3. N=17 Hunde……..…76

Tabelle 26: κ-Werte für die Übereinstimmung der Kategorien zwischen den Gutachtern (Inter-rater) in Subtest 3. N=17 Hunde……..76

Tabelle 27: p-Werte für die Übereinstimmung der Kategorien zwischen den Gutachtern (Inter-rater) im Gesamtergebnis. N=17 Hunde……76

Tabelle 28: κ-Werte für die Übereinstimmung der Kategorienzwischen den Gutachtern (Inter-rater) im Gesamtergebnis. N=17 Hunde……77

Tabelle 29: Punkteergebnisse der drei Subtests sowie des Gesamtergebnisses der drei Gutachter für die einzelnen Hunde der zweiten Auswertung. N=17 Hunde…..…77

Tabelle 30: Kategorienergebnisse der drei Subtests sowie die Gesamtkategorien der drei Gutachter für die einzelnen Hunde der zweiten Auswertung. N=17 Hunde…..…78

Tabelle 31: p-Werte für die Übereinstimmung der Punkteergebnisse zwischen Auswertung 1 und 2 (Intra-rater) von Gutachter 1. N=17 Hunde……..79

Tabelle 32: p-Werte für die Übereinstimmung der Punkteergebnisse zwischen Auswertung 1 und 2 (Intra-rater) von Gutachter 2. N=17 Hunde……..……79

Tabelle 33: p-Werte für die Übereinstimmung der Punkteergebnisse zwischen Auswertung 1 und 2 (Intra-rater) von Gutachter 3. N=17 Hunde……..…..80

Tabelle 34: p- und κ-Werte für die Übereinstimmung der Kategorienergebnisse zwischen Auswertung 1 und 2 (Intra-rater) von Gutachter 1. N=17 Hunde……....80

Tabelle 35: p- und κ-Werte für die Übereinstimmung der Kategorienergebnisse zwischen Auswertung 1 und 2 (Intra-rater) von Gutachter 2. N=17 Hunde…….80

Tabelle 36: p- und κ-Werte für die Übereinstimmung der Kategorienergebnisse zwischen Auswertung 1 und 2 (Intra-rater) von Gutachter 3. N=17 Hunde…….81

Tabelle 37: Punkt- und Kategorienergebnisse des zweiten Testdurchlaufs (Re-Test), ausgewertet von Gutachter 1...……81

Tabelle 38: p- und κ-Werte für die Übereinstimmung der Punkte- bzw. Kategorienergebnisse des ersten und zweiten Testdurchlaufs (Test-Retest) in Subtest 1. N=17 Hunde……82

Tabelle 39: p- und κ-Werte für die Übereinstimmung der Punkte- bzw. Kategorienergebnisse des ersten und zweiten Testdurchlaufs (Test-Retest) in Subtest 2. N=16 Hunde……82

Tabelle 40: p- und κ-Werte für die Übereinstimmung der Punkte- bzw. Kategorienergebnisse des ersten und zweiten Testdurchlaufs (Test-Retest) in Subtest 3. N=17 Hunde……....83

Tabelle 41: p- und κ-Werte für die Übereinstimmung der Punkte- bzw. Kategorienergebnisse des ersten und zweiten Testdurchlaufs (Test-Retest) im Gesamtergebnis. N=17 Hunde……83

Tabelle 42: Kategorienergebnisse des Datenerhebungsbogens laut Angaben der Tierheimleitung. N=17 Hunde……..83

Tabelle 43: p- und κ-Werte für die Übereinstimmung der Gesamtkategorien mit dem Datenerhebungsbogen. N=17 Hunde.....85

Tabelle 44: Häufigkeiten der gezeigten Verhaltensweisen in den drei Teilen des Referenztests. N=16 Hunde.....85

Tabelle 45: Punkte- und Kategorienergebnisse des Referenztests. N=16 Hunde....86

Tabelle 46: p-Werte für die Übereinstimmung der Punktergebnisse mit dem Referenztest. N=16 Hunde.....87

Tabelle 47: p- und κ-Werte für die Übereinstimmung der Kategorien mit dem Referenztest. N=16 Hunde.....87

Tabelle 48: Ergebnisse des C-BARQs in vier ausgewählten Kategorien. N=17 Hunde.....88

Tabelle 49: Spearmans-Korrelationskoeffizient (r) und p-Werte zwischen den vier C-BARQ-Kategorien und den Gesamt-Punkteergebnissen und Gesamt-Kategorienergebnissen des Tests auf Futteraggression. N=17 Hunde.....89

Tabelle 50: Anzahl der von den drei Gutachtern beobachteten Stresssignale je Subtest. N=17 Hunde.....89

Tabelle 51: Spearmans-Korrelationskoeffizient (r) und p-Werte zwischen der Anzahl der Stresssignale und den Punkte- und Kategorienergebnissen des Tests auf Futteraggression. N=17 Hunde (Subtest 2 N=16 Hunde).....90

1 Einleitung

Dank vieler erschreckender Medienberichte über Beißvorfälle rücken aggressive Hunde und die von ihnen ausgehende Gefährlichkeit für die Bevölkerung immer wieder in den Fokus der Politik und der Wissenschaft (JONES, 2009). Im Zuge der Gefahrenvermeidung werden mehr und mehr Verhaltenstests für Hunde entwickelt, mit dem Ziel, aggressives Verhalten zuverlässig zu erkennen und vorherzusagen (VAN DER BORG, 1991; NETTO und PLANTA 1997; LEDGER und BAXTER, 1997; VAN DEN BERG et al., 2003; PLANTA und DEMEESTER, 2007; BOLLEN und HOROWITZ, 2008; VAN DER BORG et al., 2010; DOWLING-GEYER et al., 2011; VALSECCHI et al., 2011; BENNETT et al., 2012). Vor allem in Tierheimen kommen solche Tests häufig zum Einsatz, wobei sie einen großen Einfluss auf das Überleben und die Zukunft von tausenden Hunden jährlich haben. Viele dieser Verhaltenstests sind jedoch nicht ausreichend evaluiert, sodass ihre Aussagekraft und ihr Nutzen infrage gestellt werden muss (VAN DEN BERG et al., 2003; JONES und GOSLING, 2005; TAYLOR und MILLS, 2006; DIESEL et al., 2008; BENNETT et al., 2012; MOHAN-GIBBONS et al., 2012).

Eine große Rolle bei solchen Verhaltenstests nimmt die Erkennung von futterverteidigendem Verhalten ein, welches von Hunden gezeigt wird, die in Anwesenheit von Futter aggressives Verhalten gegenüber sich annähernden Menschen zeigen. Diese Form von Aggression ist weit verbreitet (BORCHELT; 1983; GUY et al., 2001) und stellt insbesondere für Kinder im selben Haushalt eine große Gefahr dar (VOITH und BORCHELT, 1982; BORCHELT, 1983; WRIGHT, 1996). Darum sollte im Rahmen dieser Studie ein standardisierter Verhaltenstest für Tierheimhunde entwickelt und evaluiert werden, der agonistisches Verhalten von Hunden während der Futteraufnahme gegenüber sich annähernden Menschen messen soll. So wurde ein Verhaltenstest entwickelt, welcher in drei Subtests agonistisches Verhalten sowie Stresssignale gegenüber einer sich annähernden Testperson mit einer künstlichen Hand messen soll. Dieser wurde anschließend in einem Tierheim an 17 Hunden getestet und mithilfe der Ergebnisse evaluiert, wobei sowohl die Inter-rater-Reliabilität, die Intra-rater-Reliabilität, die Test-Retest-Reliabilität, als auch die Inhalts- und Kriteriumsvalidität statistisch ermittelt wurden. Die Verwendung der künstlichen Hand wurde in einem Prätest ebenfalls evaluiert, um eine Beeinflussung der Ergebnisse auszuschließen.

2 Literatur

2.1 Kanine Aggression

2.1.1 Kanine Aggression aus Sicht der Gesellschaft

In unserer modernen Gesellschaft haben domestizierte Haustiere einen hohen Nutzwert für uns Menschen. Eine besondere Stellung darunter nehmen Hunde ein, welche uns in besonders vielen Bereichen unseres Lebens zur Seite stehen. Vor mindestens 14.000 Jahren domestiziert (SERPELL, 1995), entwickelte sich die Rolle des Hundes vom Fleisch- und Pelzlieferanten über den Gehilfen bei der Jagd, Beschützer des Viehs und Wächter von Haus und Hof bis hin zum Gefährten, Freund und Familienmitglied (SERPELL, 1995). Ob als Diensthund bei Polizei, Zoll, Rettungsdienst und Militär, als Begleithund für Blinde, Behinderte oder psychisch Kranke oder als Bestandteil der Familie und bester Freund des Menschen - der Hund ist heutzutage aus unserem Leben nicht mehr weg zu denken (BERGLER, 2000).
Laut dem Geschäftsbericht von 2012 des Verbandes für das Deutsche Hundewesen (VDH) e. V. leben rund 5,3 Millionen Hunde in Deutschland. Gerade in Großstädten leben die einst vom Wolf abstammenden Tiere eng mit uns Menschen zusammen. Umso problematischer ist es daher, wenn Hunde aggressives Verhalten gegenüber Menschen, anderen Hunden oder Tieren zeigen.
In den vergangenen Jahren hat aggressives Verhalten von Hunden gegenüber Menschen immer wieder Schlagzeilen gemacht. Solche Medienberichte über fatale Vorfälle, in denen Menschen und Kinder schwer verletzt oder getötet wurden, erweckten in der Öffentlichkeit den Eindruck, dass Hunde eine große Gefahr darstellen (JONES, 2009). Während Vorfälle mit als ungefährlich geltenden Rassen bisher kaum eine Erwähnung in den Medien fanden, waren einzelne tragische Ereignisse mit den sogenannten "Kampfhunderassen" besonders häufig in den Schlagzeilen (JONES, 2009).
Es ist jedoch äußerst schwierig, genaue Zahlen von Beißvorfällen und den darin verwickelten Hunderassen zu ermitteln, da in Deutschland bis heute keine einheitlichen und systematischen Datenerhebungen von Beißvorfällen durchgeführt werden. Stattdessen werden Hundebisse vorwiegend auf kommunaler Ebene zur

Anzeige gebracht und es obliegt den einzelnen Bundesländern, Vorfälle zu dokumentieren und zu veröffentlichen (MIKUS, 2006).

In vielen Ländern führte die Forderung der Bevölkerung nach mehr Sicherheit zu mehr oder weniger sinnvollen politischen Reglementierungen, deren Sinn und Nutzen von Fachleuten kritisch bewertet werden müssen (JONES, 2009). Diese Reglementierungen schließen Rasselisten, Verbote einzelner Rassen, Wesenstests sowie Sicherungsmaßnahmen und Tötungen von als gefährlich geltenden Tieren mit ein. Denn gleich welcher Art und welchen Ausmaßes, aggressive Hunde und gefährliche Hund-Mensch-Beziehungen werden in der Öffentlichkeit nicht toleriert (BOLLEN und HOROWITZ, 2008; FEDDERSEN-PETERSEN, 2008; JONES, 2009).

Da die Gesellschaft aggressives Verhalten von Hunden kaum duldet und sensibilisiert ist gegenüber Tieren, die dieses zeigen (BOLLEN und HOROWITZ, 2008; FEDDERSEN-PETERSEN, 2008; JONES, 2009), ist leicht nachvollziehbar, dass auch Hundehalter jegliche aggressive Kommunikation ihrer Hunde unterbinden möchten (JONES, 2009). Laut zweier Umfragen unter Tierärzten und Verhaltensmedizinern sind Aggressionen bei im Haushalt lebenden Hunden das am häufigsten beschriebene Problemverhalten ihrer Kunden (GUY et al., 2001, FATJO, 2006), was zeigt, dass auch die Besitzer aggressiver Tiere derlei Verhaltensweisen als störend und problematisch einstufen (TIEFENBACH, 2001).

Sowohl in der Gesellschaft als auch unter Verhaltensbiologen, Soziologen und Psychologen werden Aggressionen überwiegend negativ bewertet und gelten im Allgemeinen als "böse" (FEDDERSEN-PETERSEN, 2008). Doch aggressives Verhalten ist allgegenwertig und in allen sozialen Verbänden auch zwingend notwendig für die Etablierung von Beziehungen und Kommunikation, der Bildung einer Hierarchie und die Sicherung von Ordnung und Stabilität (FEDDERSEN-PETERSEN, 2004, 2008; JONES, 2009), was im nächsten Kapitel genauer erläutert wird.

2.1.2 Agonistisches und aggressives Verhalten von Hunden

Definition der Begriffe Agonistik und Aggression

Der Begriff "Aggression" wird häufig, vor allem aber in der anglo-amerikanischen Literatur, als Synonym für "Agonistik" (auch "Antagonistik" oder "agonales Verhalten") verwendet, obwohl beide Begriffe in der Literatur unterschiedlich definiert werden (IMMELMANN, 1982; FEDDERSEN-PETERSEN, 2004). So gehören zum

Aggressionsverhalten laut IMMELMANN (1982) alle Elemente des Angriffs-, Abwehr- und Drohverhaltens.
Einheitliche Definitionen haben sich jedoch für beide Begriffe bisher nicht durchgesetzt (FEDDERSEN-PETERSEN, 2008; JONES, 2009). Nach SCOTT und FREDERICSON (1951) stellt aggressives Verhalten nur einen Teilbereich der Agonistik dar, welche als Überbegriff für Aggression, Submission und Flucht Verwendung findet. Reizungsversuche an Katzenhirnen von HUNSBERGER (1956) sowie von VON HOLST und VON SAINT PAUL (1960) belegen ein zusammenhängendes, funktionelles System für Angriffsverhalten, Abwehrverhalten und Flucht im Mittelhirn und Hypothalamus, was einen Zusammenhang dieser Verhaltensweisen auf neuronaler Ebene beweist. Aus diesem Grund wird heute der Begriff der Agonistik von den meisten Autoren einheitlich verwendet als Sammelbezeichnung für alle Verhaltensweisen gegenüber Sozialpartnern, die das eigene Verhalten störend beeinflussen (GATTERMANN, 1993; FEDDERSEN-PETERSEN, 1993; BERNAUER-MÜNZ und QUANDT, 1995; FEDDERSEN-PETERSEN, 2008) So beinhaltet der Begriff Agonistik heute die beiden gegensätzlichen Anteile des Angriffsverhalten (offensives Verhalten) und des Abwehr- (defensives) bzw. Fluchtverhalten (IMMELMANN, 1982; GATTERMANN, 1993; FEDDERSEN-PETERSEN, 2004 und 2008). Es kommt häufig zu Überlagerungen beider Anteile und es ist selten, dass lediglich Verhaltensweisen eines Anteils gezeigt werden, da in Konfliktsituationen sowohl Angriffs- als auch Fluchtbereitschaft gleichzeitig aktiviert werden (IMMELMANN et al., 1996).

Die "4 F's"

Verhaltensweisen, die gezeigt werden, um den Abstand zu einer Bedrohung zu vergrößern, sind jene des Kampfes und der Flucht, sowie Erstarren und das Zeigen von Übersprungshandlungen (ARCHER, 1988; BERNAUER-MÜNZ und QUANDT, 1995, FEDDERSEN-PETERSEN, 2004, 2008; JONES, 2009; SCHÖNING und RÖHRS, 2013). Zum Kampf gehören der Offensivangriff und alle Elemente des Drohverhaltens, welche direkt distanzvergrößernd wirken, ebenso wie die Flucht, welche bei ausbleibenden Erfolg auch in einen Defensivangriff übergehen kann. Erstarren wird häufig als Initialreaktion auf eine Bedrohung gezeigt, wenn eine Auseinandersetzung vermieden werden soll in der Hoffnung, dass diese vorübergeht. Übersprungshandlungen sind dagegen Verhaltensweisen, die scheinbar nicht in den

gegenwärtigen Kontext passen und stammen meist aus anderen Funktionskreisen wie dem Spielverhalten. Sie sollen die Situation entspannen.
Diese vier Strategien werden oft auch als die "Vier F's" - Fight (Kampf), Flight (Flucht), Freeze (Erstarren) und Flirt (Übersprungshandlungen) - bezeichnet (FEDDERSEN-PETERSEN, 2008; JONES, 2009; SCHÖNING und RÖHRS, 2013). LINDSAY (2005) beschreibt jedoch noch eine fünfte Konfliktlösungsstrategie, das Dulden und Tolerieren einer Situation (Forbear). Welche der vier bzw. fünf Strategien in einer bedrohlichen Situation zum Einsatz kommt, hängt von unterschiedlichen Faktoren ab. So spielen die äußeren Umstände einer Situation, die Motivation des Tieres, seine angeborenen Eigenschaften, die bisherigen Erfahrungen und die Erfolgsaussichten der einzelnen Strategien eine große Rolle (LINDSAY, 2000; FEDDERSEN-PETERSEN, 2008; SCHÖNING und RÖHRS, 2013).

Biologischer Sinn von Agonistik und Aggression

Agonistisches Verhalten dient damit in erster Linie dazu, eine subjektiv empfundene Bedrohung zu eliminieren, eine Distanz zu ihr zu erhalten oder zu erschaffen (TEMBROCK, 1992; FEDDERSEN-PETERSEN und OHL, 1995; JONES, 2009), weshalb es von TEMBROCK (1987) auch als "sozio-negatives Verhalten" bezeichnet wird. Mithilfe von offensiven und defensiven Verhaltensweisen können Störungen beseitigt und raumzeitliche Distanzierungen aufrechterhalten werden. Damit dient agonistisches Verhalten auch zur Etablierung von Rangordnungen in sozialen Verbänden, was zur Stabilität und Ordnung in der Gruppe sowie zur Schaffung von Aktionsräumen für das Verhalten der Individuen beiträgt (TEMBROCK, 1992; GATTERMANN, 1993; FEDDERSEN-PETERSEN, 2004). Damit stellt agonistisches Verhalten einen lebensnotwendigen Bereich von sozialem Verhalten dar (FEDDERSEN-PETERSEN, 2008).
Als Teil der Agonistik ist Aggressionsverhalten laut FEDDERSEN-PETERSEN (2008) ein typisches multifaktorielles Verhalten und dient der Selbstverteidigung, dem Schutz der Nachkommen, Verwandten und Partner, dem Wettbewerb um Ressourcen wie Nahrung, Geschlechtspartner oder Territorium und dem individuellen Statusgewinn. Es reguliert soziale Beziehungen, etabliert Rangordnungen und ist damit zwingend notwendig zum Überleben. Auch nach JONES (2009) dient aggressives Verhalten vor allem dazu, eine Bedrohung auf Distanz zu halten und die eigenen Interessen bei einem Konflikt um lebenswichtige Ressourcen durchzusetzen. Es ist angeboren,

normal und artgemäß (IMMELMANN, 1996; FEDDERSEN-PETERSEN, 2004, 2008; JONES, 2009; SCHÖNING und RÖHRS, 2013).

Aggressive Kommunikation von Hunden

Um Verletzungen und unnötige Energieverschwendung zu vermeiden, haben sich in der Evolution Mechanismen herausgebildet, damit die Konflikte zwischen Individuen ohne Ernstkampf gelöst werden (JONES, 2009). Bei Wölfen und Haushunden haben sich zur Konfliktvermeidung Kommunikation und ritualisiertes Droh- und Aggressionsverhalten durchgesetzt (FEDDERSEN-PETERSEN, 2008, JONES, 2009). Dazu werden verschiedene optische und akustische Signale gezeigt, die Verteidigungsbereitschaft (defensiv) oder Angriffsbereitschaft (offensiv) signalisieren (FEDDERSEN-PETERSEN, 2008).

Zur Hemmung aggressiven Verhaltens eines Sozialpartners und zur Deeskalation von bedrohlichen Situationen dient das Demuts- oder Submissionsverhalten und das Beschwichtigungsverhalten (FEDDERSEN-PETERSEN, 2008; JONES, 2009). Gesten wie Blick- und Kopfabwenden, Zurücklegen der Ohren, Einknicken der Gliedmaßen, das tiefe Tragen der Rute und das Lecken der eigenen Schnauze sollen das Streben nach freundlicher und harmonischer sozialer Interaktion verdeutlichen (SCHENKEL, 1967; FEDDERSEN-PETERSEN, 2008; JONES, 2009). Führen diese Signale von Seiten des Hundes nicht zum Erfolg, kann der Hund schließlich aggressives Verhalten wie Erstarren, Drohfixieren, Knurren, Schnappen und Beißen als weitere Lösungsstrategie zeigen, wobei fließende Übergänge möglich sind (FEDDERSEN-PETERSEN, 2008).

Mithilfe dieser Eskalationskaskade von agonistischen Verhaltensweisen ist ein fein graduiertes und hochritualisiertes Drohverhalten möglich (FEDDERSEN-PETERSEN, 2004; SCHÖNING und RÖHRS, 2013). Dazu werden zum Teil subtile ritualisierte Intentionsbewegungen, Stimmungsbewegungen und Intentionsbewegungen zum Kämpfen gezeigt (FEDDERSEN-PETERSEN und OHL, 1995).

2.1.3 Klassifizierung von aggressivem Verhalten

Aggressives Verhalten ist vielursächlich. Art und Ausmaß hängen vom auslösenden Reiz und inneren Bedingungen wie Reaktionsbereitschaft oder Motivation ab (FEDDERSEN-PETERSEN, 2008), ebenso von der genetischen und physiologischen Ausstattung der Hunde, der Sozialisation, früheren Erfahrungen und vielen anderen Faktoren wie Haltung und Umweltbedingungen (WRIGHT, 1999; FEDDERSEN-PETERSEN, 2004). Vor allem Angst aufgrund von negativen oder mangelnden Erfahrungen kann aggressives Verhalten auslösen. Alternativ kann aggressives Verhalten ein Symptom für organische Ursachen wie Schmerz, verminderte Sinnesfunktionen, Stoffwechselerkrankungen, hormonelle Störungen, Infektionskrankheiten oder genetische Defekte sein.

In der Vergangenheit wurden daher viele Versuche unternommen, aggressives Verhalten zu klassifizieren. Zum einen kann das aggressive Verhalten eines Hundes als "normal" oder "unnormal" bezeichnet werden, wobei "normal" art- und erwartungsgemäß, sinnvoll und nachvollziehbar, "unnormal" aber pathologisch, übersteigert, deplatziert oder biologisch nicht nachvollziehbar bedeutet (ALOFF, 2002). Zum anderen werden Aggressionen nach dem Ziel eingeteilt, wobei nicht nur zwischen Menschen, anderen Hunden oder Tieren unterschieden wird, sondern auch, ob diese dem aggressiven Hund bekannt oder unbekannt sind (ALOFF, 2002; FEDDERSEN-PETERSEN, 2008; JONES, 2009).

Am häufigsten jedoch wurde versucht, die Auslöser bzw. Stimuli für aggressives Verhalten zu klassifizieren, was zu vielen verschiedenen "Aggressions-Modulen" führte (FEDDERSEN-PETERSEN, 2008). HOUPT (1979), HART (1980), BEAVER (1983) und BLACKSHAW (1985) beschrieben beispielsweise Konkurrenz-motivierte Aggression, sexuell-motivierte Aggression, territoriale Aggression, Beuteaggression[1], durch Schmerz ausgelöste Aggression, Angstaggression, maternale Aggression, Dominanzaggression und erlernte Aggression. 1983 definierte BORCHELT ähnliche acht Arten von Aggressionen: Aggression bezogen auf Furcht, Dominanz und Ressourcen, Territoriumsaggression, Beuteaggression, straf- und schmerzbedingte Aggression, sowie intraspezifische Aggression. HART und HART (1985) und

[1] Beutefangverhalten wird mittlerweile klar von aggressivem Verhalten getrennt, da hier eine völlig andere Motivation zugrunde liegt, andere neurophysiologische Prozesse ablaufen, die Handlungsketten verschieden sind und mit der Beute keinerlei agonistische Kommunikation stattfindet (FEDDERSEN-PETERSEN, 2008; JONES, 2009).

BLACKSHAW (1987) beschrieben außerdem eine idiopathische Aggression. Laut WRIGHT (1999) sind die Arten von offensiver Aggression Dominanzaggression, territoriale Aggression, strafbezogene Aggression, intraspezifische Aggression, Beuteaggression und ressourcenbezogene Aggression. Auch JONES (2009) teilt Aggressionen ein in aggressives Verhalten aus Angst, in Verbindung mit Ressourcen, aufgrund von Frustration oder wegen organischer Ursachen.

Heutzutage jedoch wird Aggression als komplexes Interaktionsgefüge zwischen äußeren und inneren Bedingungen sowie Lernerfahrungen betrachtet (WRIGHT, 1999; FEDDERSEN-PETERSEN, 2004, 2008; HASSENSTEIN, 2007). So können aggressive Verhaltensweisen gleich oder ähnlich aussehen und trotzdem auf unterschiedlichen inneren Bedingungen beruhen. Sie sind daher nicht statisch in einzelne Kategorien zu pressen, sondern jeweils situativ zu bewerten (FEDDERSEN-PETERSEN, 2008). Des Weiteren stellt aggressives Verhalten keinen eigenen Funktionskreis des Verhaltensrepertoires dar, sondern kann je nach Motivation und Ziel verschiedenen Funktionskreisen zugeordnet werden (TEMBROCK, 1983; FEDDERSEN-PETERSEN, 1993; GATTERMANN, 1993; SCHÖNING, 2013). Eine Auflistung von Aggressionsmodulen ist deshalb überholt und kaum hilfreich, da sie in den meisten Fällen subjektiv, reduktionistisch und für das situative Verständnis des Hundes ungeeignet sind (FEDDERSEN-PETERSEN, 2008). Außerdem kommt es oftmals vor, dass der Aggression von Hunden nicht nur eine, sondern mehrere Motivationen zugrunde liegen (WRIGHT, 1999). Solche sogenannten Misch-motivationen sind zumeist schwer den Aggressions-Modulen zuzuordnen und führen häufig zu Fehlinterpretationen von Seiten der Hundebesitzer (FEDDERSEN-PETERSEN, 2008).

Soll nur anhand der Körpersprache des Hundes eine Aussage über die Emotionen des Hundes getroffen werden, kann aggressives Verhalten jedoch generell in offensive und defensive Aggression eingeteilt werden (LINDSAY, 2001; FEDDERSEN-PETERSEN, 2008; JONES, 2009). Beide Strategien sind wertfrei zu sehen (FEDDERSEN-PETERSEN, 2008). Sie beinhalten jeweils freies und gehemmtes Aggressionsverhalten sowie Drohverhalten. Die Angriffsbereitschaft bei der offensiven Aggression ist hoch, was durch Anschleichen, Drohfixieren und Knurren bis hin zur Beißdrohstellung und einem freien Angriff mit Beißen und einem Ernstkampf gezeigt wird. Dabei kommen selbstsichere, aus dem Imponierverhalten entstehende Signale zum Ausdruck, wie etwa nach vorn aufgestellte Ohren, eine über der Rückenlinie

getragene Rute, durchgestreckte Gliedmaßen oder eine kurze und runde Maulspalte (FEDDERSEN-PETERSEN, 2004, 2008). Dagegen resultiert defensives Aggressionsverhalten aus Situationen von Angst, Furcht oder einer akuten Bedrohung und wird eher von unsicheren Hunden gezeigt (LINDSAY, 2000). Sie deutet soziale Unsicherheit und beginnenden Rückzug bei voller Verteidigungsbereitschaft an (FEDDERSEN-PETERSEN, 2004). Die aggressiven Signale wie Knurren, Schnappen und Beißen werden hierbei mit submissiven Gesten wie einer eingeklemmter Rute, angelegten Ohren und einer langer Maulspalte kombiniert (WRIGHT, 1999).
FEDDERSEN-PETERSEN (2008) beschreibt eine Übersicht über das Ausdrucksverhalten von unterschiedlich bedingtem Aggressionsverhalten (Tabelle 1, mod. nach ZIMEN, 1971):

Tabelle 1: Ausdrucksverhalten bei offensiv- und defensivbedingter Aggression (FEDDERSEN-PETERSEN, 2008, mod. nach ZIMEN, 1971)

Drohen offensiv	**Drohen defensiv**
Anschleichen	Gebissklappern
Blickkontakt	Wegsehen
Überfalldrohung	Abwehrschnappen
Haarsträuben	Haarsträuben
Knurren	Knurren
Vorn-Zähneblecken	Voll-Zähneblecken
Beißdrohstellung	Abwehrdrohen
Gehemmt offensiv	**Gehemmt defensiv**
Über die Schnauze beißen	Abwehr mit gekrümmtem Hals
Gegenstand abnehmen	Abwehrkreisel
Schieben, Anrempeln, Aufreiten	Abwehr auf dem Rücken
Runterdrücken, Umstellen, Überfall	
Vorderbeinstoßen, Anspringen	
Hochkampf, Rückenbiss, Verfolgen	
Frei offensiv	**Frei offensiv**
Angriff	Abwehrbeißen
Beißen	

Ernstkampf

Auch bei dieser Einteilung von aggressivem Verhalten sind jedoch Mischformen möglich, bei denen die Hunde offensive und defensive Signale in ambivalentem Verhalten kombiniert zeigen (LINDSAY, 2001; FEDDERSEN-PETERSEN, 2008).
Außerdem bilden die Sozialpartner während einer aggressiven Auseinandersetzung häufig komplexe Funktionseinheiten, bei denen die Rollen des Angreifers und des Verteidigers mitunter wechseln können (FEDDERSEN-PETERSEN, 2008).

2.1.4 Ressourcenbezogene Aggression

Bei ihren unterschiedlichen Einteilungen der Aggressionsauslöser beschreiben nahezu alle Autoren eine Aggression in Verbindung mit Ressourcen. Diese ressourcenbezogene Aggression kann auftreten, wenn ein Hund im Besitz einer für ihn wichtigen Ressource wie Futter oder Spielzeug ist und diese mit aggressivem Verhalten gegenüber einer subjektiven Bedrohung verteidigt (WRIGHT, 1999; LINDSAY, 2001; NEILSON, 2007; FEDDERSEN-PETERSEN, 2008; JONES, 2009). Ressourcen, die von Hunden verteidigt werden, können alle überlebenswichtigen Dinge sein, wie Futter, Wasser, Territorium oder Fortpflanzungspartner, aber auch Spielzeug, Schlafplätze, Durchgänge, der Besitzer oder andere in den Augen des Hundes wichtige oder interessante Objekte (ALOFF, 2002; LINDSAY, 2005; FEDDERSEN-PETERSEN, 2008; JONES, 2009). Dabei gilt, je seltener, und damit attraktiver und wichtiger, eine Ressource und für den Hund ist, desto eher wird ein hohes Maß an Aggressivität gezeigt (MOHAN-GIBBONS et al., 2012).
WRIGHT (1999), ALOFF (2002) und NEILSON (2007) unterteilen diese Art von Aggression ebenfalls nach offensiver oder defensiver Motivation und Körperhaltung. Am häufigsten aber ist ressourcenbezogenes Aggressionsverhalten von Angst und Furcht motiviert und nicht etwa von Dominanz, so wie BORCHELT (1983) behauptet (NEILSON, 2007, MARDER et al., 2013). So hat eine Untersuchung von MECH (1970) gezeigt, dass auch rangniedere Tiere in einem Wolfsrudel ihr Futter erfolgreich gegen die ranghohen Tiere verteidigen können. Dabei kommen häufig auch Mischmotivationen vor, bei denen ambivalente, wechselnde Signale beobachtet werden.
Ressourcenbezogene Aggression ist ein häufiges Verhaltensproblem von Hunden. In einer älteren Untersuchung von BORCHELT (1983) an 524 Hunden mit

Verhaltensproblemen kam ressourcenbezogene Aggression mit einer Prävalenz von 17% vor, wobei intakte Rüden mehr ressourcenbezogene Aggression zeigten als kastrierte Rüden oder Hündinnen. In einer aktuelleren Umfragestudie von GUY et al. (2001) zeigten 20% von 3226 Hunden ressourcenbezogene Aggression.
Opfer von ressourcenbezogener Aggression sind dabei meistens Menschen im selben Haushalt, darunter vor allem Kinder unter fünf Jahren, da diese häufig um die Ressourcen der Hunde konkurrieren, welche am Boden und in Reichweite der Kinder liegen. Dabei kommt es oft zu Verletzungen am Kopf, Hals oder Gesicht des Kindes (VOITH und BORCHELT, 1982; BORCHELT, 1983; WRIGHT, 1996).

2.1.5 Futteraggression

Biologischer Sinn und Risikofaktoren von Futteraggression

Eine häufig vorkommende Form von ressourcenbezogener Aggression ist die futterbezogene Aggression. Von dieser sogenannten Futteraggression wird gesprochen, wenn ein Hund, der im Besitz oder in der Nähe von Futter ist, Drohsignale oder beschädigendes Verhalten gegen ein von ihm als Bedrohung empfundenes Individuum richtet, mit der Absicht, dieses vom Futter fernzuhalten (DONALDSON, 2002; NEILSON 2007).
Es ist normal für einen Hund, Futter gegen eine als solche empfundene Bedrohung zu verteidigen, denn Futter ist überlebenswichtig und die aggressive Kommunikation ist ein geeignetes Mittel, eine Bedrohung von dieser wichtigen Ressource fernzuhalten (DONALDSON, 2002; NEILSON, 2007). Die Intensität von Futter-aggression kann dabei von dezenten Drohsignalen wie beschleunigtem Verzehr, Erstarren, Drohfixieren und dem Abschirmen des Futters mit dem Körper über stärker ausgeprägtes Drohverhalten wie Knurren und Zähneblecken bis hin zu starker Aggressivität wie Schnappen oder Beißen reichen, wenn entweder der fressende Hund oder das Futtermittel von Personen berührt werden oder der Versuch gemacht wird, diese zu berühren (DONALDSON, 2002; NEILSON, 2007; MOHAN-GIBBONS et al., 2012; MARDER et al., 2013).

Viele Besitzer erhöhen auf verschiedenen Wegen unbewusst das Risiko von Futteraggression ihres Hundes: So kann die Gabe von hochwertigem, seltenem oder attraktivem Futter bei Hunden eher das Verlangen auslösen, dieses zu verteidigen, als

Futter, welches weniger attraktiv für den Hund ist (LINDSAY, 2005; NEILSON, 2007). Daneben kann das Füttern an ungünstigen Stellen mit viel genutzten Laufwegen für den Hund sehr bedrohlich wirken (NEILSON, 2007). Außerdem sind manche Präventivmaßnahmen, die häufig bei Welpen angewendet werden, kontraproduktiv: So kann das häufige Wegnehmen von Futter während des Fressens Futteraggression auslösen oder verstärken, selbst wenn das Futter wieder hingestellt wird. Denn hier lernt der Welpe nur, dass die Annäherung eines Menschen beim Fressen etwas Unangenehmes zur Folge hat, nämlich den Verlust des Futters, was aus lerntheoretischer Sicht einer negativen Strafe[2] gleichkommt (LINDSAY, 2005; NEILSON, 2007). Auch das Hineinhalten der Hand in die Futterschüssel während des Fressens, um einen Gewöhnungseffekt zu erreichen, kann das Gegenteil bewirken und den Hund eher sensibilisieren, also reaktiver machen (NEILSON, 2007). Kommt es schließlich zum futterverteidigenden Verhalten und ist dieses erfolgreich, indem sich die Bedrohung entfernt, wird es dadurch schnell negativ verstärkt[3], also belohnt. Das Verhalten wird dann zukünftig in solchen Situationen öfter gezeigt (LANDSBERG, 2003; NEILSON, 2007).

In einer Untersuchung von MCGREEVY und MASTERS (2008), bei denen sie in einer Umfrage über das Verhalten von 690 australischen Hunden die Risikofaktoren für Trennungsangst und futterbezogener Aggression ermittelten, stellten sie fest, dass Mischlinge mit höherer Wahrscheinlichkeit stärker ausgeprägte futterbezogene Aggression zeigen, als reinrassige Hunde. Auch ein höheres Alter bei Erwerb der Hunde lässt demnach das Risiko für das Auftreten von stärkerer futterbezogener Aggression steigen. Bei den halterbedingten Einflussfaktoren stellten MCGREEVY und MASTERS (2008) außerdem fest, dass tägliche Spaziergänge mit dem Hund das Risiko für futterbezogene Aggression senken, wogegen Hunde, die gleich nach Ankunft ihrer Besitzer ausgeführt werden, ein höheres Risiko haben, futterbezogene Aggression zu zeigen. Liebkosen Besitzer ihre Hunde innerhalb von 30 min nach ihrer Ankunft zuhause, entwickeln die Hunde laut ihren Untersuchungen mit höherer Wahrscheinlichkeit futterbezogene Aggression, ebenso, wenn die Hunde Futterbelohnungen während des Essens der Besitzer erhielten. Dagegen zeigten sie auf, dass Hunde, die während des Trainings mit Futter belohnt wurden, mit geringerer Wahrscheinlichkeit futterbezogene Aggression entwickelten. Dieses Ergebnis ließ

[2] negative Strafe = etwas Angenehmes wird entfernt
[3] negative Belohnung = etwas Unangenehmes wird entfernt

MCGREEVY und MASTERS (2008) vermuten, dass Hunde, die viel Zeit mit ihren Besitzern verbringen und auch mit Futterbelohnungen trainiert werden, die Priorität von Ressourcen und damit den Wert von Futter anders einschätzen als Hunde, die nicht regelmäßig ausgeführt und trainiert werden. Tägliches Training mit Futterbelohnungen stärke demnach außerdem die Impulskontrolle des Hundes und übe den Umgang mit Menschen in der Gegenwart von Futter. Als umweltbedingten Einflussfaktor auf die Entwicklung von futterbezogener Aggression fanden MCGREEYVY und MASTERS (2008) die Anzahl der Hunde im Haushalt, wobei die Wahrscheinlichkeit für futterbezogene Aggression eines Hundes mit der Anzahl der im Haus lebenden Hunde stieg. Dies erklärten sie damit, dass mit der Anzahl der Hunde im Haushalt auch die Konkurrenz um Ressourcen steigt und somit die Notwendigkeit, Futter als Ressource zu verteidigen. Ebenso stieg die Wahrscheinlichkeit von intensiver ausgeprägter Futteraggression mit der Anzahl der im Haushalt lebenden weiblichen Personen, was sie mit dem generellen Einfluss des Geschlechts von Personen auf das Verhalten der Hunde erklärten. MCGREEVY und MASTERS (2008) stellten außerdem einen signifikanten Zusammenhang von futterbezogener Aggression mit dem Auftreten einer Trennungsangst fest. Dieses Ergebnis erklärten sie mit einem engen Zusammenhang der Motivationen von Hunden, in Besitz und Kontrolle von Futter sowie dem Besitzer als Ressource zu sein.

Bedeutung von Futteraggression im Tierheim

Da Futteraggression beim Hund mit einem hohen Verletzungsrisiko für alle Menschen in dessen Umgebung einhergehen kann, tun sich viele Tierheime schwer, solche Hunde zu vermitteln. Futteraggression ist stattdessen der häufigste Grund, selbst junge und gesunde Tiere als nicht vermittelbar einzustufen und sofort zu euthanasieren, was rund die Hälfte (51%) der Tierheime in den USA praktizieren (MOHAN-GIBBONS et al., 2012). Nur 34% der Tierheime versuchen eine Verhaltensmodifikation und die Vermittlung solcher Hunde, zumeist unter strengen Auflagen und Sicherheitsvorkehrungen (MOHAN-GIBBONS et al., 2012; MARDER et al., 2013). Zum Beispiel dürfen oftmals keine Kinder im Haushalt sein, der neue Besitzer muss ein gewisses Maß an Erfahrung mitbringen oder er muss Erziehungskurse absolvieren. Diese Bedingungen erschweren die Vermittlung erheblich oder machen sie ganz unmöglich (MOHAN-GIBBONS et al., 2012; MARDER et al., 2013).

Diese Vorgehensweise der Tierheime ist besorgniserregend, da das gezeigte futteraggressive Verhalten während eines provokativen Verhaltenstests in einem Tierheim, mit all seinen begleitenden stressreichen Umweltbedingungen, nicht zwingend in einem stabilen häuslichen Umfeld mit familiären Bindungen bestehen bleiben muss. MARDER et al. (2013) gelangten in einer aktuellen Studie über futterverteidigendes Verhalten zu dem Schluss, dass viele im Tierheim als futteraggressiv bewertete Hunde im neuen Zuhause kein futteraggressives Verhalten mehr zeigten. So waren nur wenig mehr als die Hälfte der Hunde (55%), die im Tierheim als futteraggressiv getestet wurden, auch im neuen Zuhause futteraggressiv. Dagegen zeigten 22% der Hunde, die im Tierheim als nicht futteraggressiv getestet wurden, Futteraggression im neuen Zuhause. 78% der im Tierheim als nicht futteraggressiv getesteten Hunde zeigten auch im neuen Zuhause keine Anzeichen für Futteraggression (MARDER et al., 2013). Das Nichtauftreten von Futteraggression im Verhaltenstest im Tierheim ist damit von größerem informativen Wert als das Auftreten von Futteraggression (MARDER et al., 2013).

In einem Tierheim kann das Maß an Futteraggressivität von den Tierheimmitarbeitern während des täglichen Umgangs mit den Hunden oder aber in einem Verhaltenstest festgestellt werden, welcher dieses Verhalten gezielt provoziert (MARDER et al., 2013). So konnte in einer Studie von MOHAN-GIBBONS et al. (2012), in der 77 Tierheime zum Umgang mit futteraggressiven Hunden befragt wurden, eine durchschnittliche Prävalenz von 14% (7-30%) ermittelt werden.
In einer Studie zur Vorhersagekraft solcher Verhaltenstests in den Niederlanden stellten VAN DER BORG et al. (1991) jedoch fest, dass futterbezogene Aggression das am schlechtesten vorhersagbare Verhalten war. Dagegen zeigte sich in einer Untersuchung von MARDER et al. (2013) ein signifikanter Zusammenhang zwischen dem Auftreten von Futteraggression in einem standardisierten Verhaltenstest und dem Auftreten von Futteraggression im neuen Zuhause.

Bedeutung von Futteraggression im häuslichen Umfeld

Häufig zeigen Hunde im häuslichen Umfeld nur geringgradiges futteraggressives Verhalten und nur wenige Hunde zeigen schwerwiegende Aggression (GUY et al., 2001). Zum selben Ergebnis gelangten auch MARDER et al. (2013), welche feststellten, dass die meisten futteraggressiven Hunde (82%) nur moderates

Aggressionsverhalten (Knurren oder Zähneblecken) zeigten und nur 18% auch starkes Aggressionsverhalten (Schnappen oder Beißen). Außerdem stellte sich in dieser Studie heraus, dass die Mehrzahl der Hunde (93%), die nach der Adoption in ein neues Zuhause noch Futteraggression zeigten, dies nur selten taten und nur 7% der Hunde regelmäßig oder immer futteraggressiv waren.

Außerdem muss berücksichtigt werden, ob die neuen Besitzer Futteraggression als händelbar oder gefährlich betrachten. Ein Viertel der Besitzer (25%) gab in Untersuchungen an, nie mit ihrem fressenden Hund zu interagieren und futteraggressives Verhalten, selbst wenn ihr Hund dieses zeigte, als wenig problematisch zu betrachteten (VAN DER BORG et al., 1991).

MOHAN-GIBBONS et al. (2012) und MARDER et al. (2013) ermittelten ebenfalls einen großen Unterschied zwischen den Ergebnissen des Verhaltenstests auf Futteraggression im Tierheim und der Einschätzung des Problems durch die neuen Besitzer. So bewerteten in der Studie von MARDER et al. (2013) 28% der Hundehalter Futteraggression als gar nicht problematisch und nur 5% der Halter schätzten Futteraggression als großes Problem ein. Fast alle Hundehalter (96%) gaben in einer Befragung nach der Adoption an, ihre Hunde seien nicht futteraggressiv, darunter selbst jene Besitzer, die aggressives Verhalten bei ihren Hunden gesehen hatten. Auch 48% der Hundehalter, die Futteraggression bei ihren Hunden bereits beobachtet hatten, schätzten dieses Verhalten als wenig oder gar nicht problematisch ein, während nur 25% der Hundehalter, deren Hunde nicht futteraggressiv waren, dieser Meinung waren (MARDER et al., 2013).

Wenn Hunde futteraggressiv sind, ist eine Therapie im Rahmen einer Verhaltensmodifikation möglich. In einer Studie von MOHAN-GIBBONS et al. (2012) wurde der Einfluss einer Verhaltensmodifikation auf erfolgreich vermittelte, futteraggressive Hunde untersucht. Hierbei wurde festgestellt, dass nach einem Verhaltensmodifikationsprogramm im Tierheim und im neuen Zuhause nach drei Wochen nur noch 10% der Hunde futteraggressiv waren und dass nach drei Monaten laut Angaben der Besitzer kein Hund mehr futteraggressives Verhalten zeigte.

So wird Futteraggression aus Sicht der Hundehalter weniger problematisch eingeschätzt als aus Sicht der Tierheime und ist noch dazu gut zu therapieren. Die gängige Praxis der Vermittlung unter strengsten Auflagen aber vor allem der sofortigen Euthanasie von futteraggressiven Tieren in Tierheimen in den USA ist damit als

unbegründet und tierschutzrelevant zu betrachten (MOHAN-GIBBONS et al., 2012; MARDER et al., 2013).

2.2 Temperament- und Wesenstests

2.2.1 Grundidee von Verhaltenstests

Beständigkeit der Reaktionen des Hundes auf einen Stimulus charakterisiert den Begriff "Temperament" oder "Wesen" (TAYLOR und MILLS, 2006). Viele Autoren benutzen die Wörter Wesen, Temperament oder auch Charakter und Persönlichkeit als ein und denselben Begriff, wogegen andere versucht haben, die einzelnen Begriffe voneinander abzugrenzen und zu definieren (DIEDERICH und GIFFROY, 2006; BENNETT, 2010). In der Literatur finden sich daher viele verschiedene Begriffserklärungen, von denen sich bisher keine gegenüber den anderen durchgesetzt hat. FEDDERSEN-PETERSEN (2004) versteht den Begriff des Wesens als das Verhalten und die Bindung eines Hundes seinem Halter gegenüber sowie das Verhalten gegenüber fremden Menschen und Artgenossen. Außerdem rechnet sie auch die Reaktionsbereitschaft und -geschwindigkeit zum Wesen, die Ausdauer und schließlich auch das Verhaltensinventar des Hundes, welches mehr oder weniger vollständig sein oder im Bereich bestimmter Funktionskreise übersteigerte oder reduzierte Verhaltensweisen aufweisen kann. DIEDERICH und GIFFROY (2006) definieren den Begriff Temperament als Verhaltensunterschiede von Individuen, welche unter ähnlichen Bedingungen relativ konstant ab einem frühen Alter, über längere Zeit hinweg und in verschiedenen Situationen gezeigt werden. Einflussfaktoren auf das Verhalten und auf das zugrunde liegende Temperament können dabei genetischer, pränataler und umweltbedingter Natur sein, sowie von frühen Erfahrungen und Lernvorgängen geprägt werden. Betrachtet man diese verschiedenen Definitionen der Begriffe Temperament und Wesen, fällt auf, dass zumeist das Verhalten des Hundes als Ganzes gemeint ist, welches von einzelnen Verhaltensweisen gebildet wird (DIEDERICH und GIFFROY, 2006).

Um Eigenschaften, Fähigkeiten und Verhaltensbesonderheiten von Hunden zu untersuchen, wurden daher Wesens- oder Verhaltenstests entwickelt, in denen das Verhalten von Hunden gegenüber ihrer Umwelt ermittelt werden soll. Hierzu werden Hunde mit einer Vielzahl von verschiedenen optischen, akustischen und olfaktorischen Stimuli konfrontiert, die bekanntermaßen aggressives Verhalten auslösen können

(FEDDERSEN-PETERSEN, 2004). Damit bieten sie die Möglichkeit, Verhalten von Hunden zu messen.

In seiner einfachsten Form ermittelt ein Verhaltenstest das Verhalten eines individuellen Tieres auf einen bestimmten Reiz, zu einer bestimmten Zeit und an einem bestimmten Ort (BENNETT, 2010), wodurch ein solcher Test immer nur als eine Momentaufnahme des Hundeverhaltens in genau dieser Situation und mit genau diesen äußeren und inneren Bedingungen zu betrachten ist (BENNETT, 2010; FEDDERSEN-PETERSEN, 2004). Ändert sich eine dieser Gegebenheiten, kann die Reaktion des Hundes eine andere sein (FEDDERSEN-PETERSEN, 2004). So kann ein Hund, der im Test nicht aggressiv reagiert, in anderen Situationen durchaus aggressives Verhalten zeigen (PAROZ et al., 2008), da Verhalten jederzeit Veränderungen unterworfen ist durch unterschiedliche Umweltreize, Lernprozesse sowie dem Verhalten des Hundehalters (FEDDERSEN-PETERSEN, 2004; TAYLOR und MILLS, 2006). Es ist daher nicht immer konstant und kann je nach Situation variieren.

Da es aber nicht möglich ist, alle Umweltsituationen oder komplexe Situationskonstellationen in einem Verhaltenstest zu simulieren, sind diese wenig geeignet, mit absoluter Sicherheit das gesamte zukünftige Hundeverhalten vorherzusagen (LEDGER und BAXTER, 1997; KAHN et al., 2003; FEDDERSEN-PETERSEN, 2004; BARNARD et al., 2012).

Der Sinn eines Verhaltenstests liegt stattdessen darin, das Wissen über das Verhalten eines Hundes in bestimmten Situationen zu verbessern und damit sein zukünftiges Verhalten über den Zufall hinaus vorhersagen zu können (TAYLOR und MILLS, 2006). Deshalb wird versucht, aus den spezifischen Reaktionen der Hunde in einem Verhaltenstest einen Rückschluss auf zukünftiges Verhalten in ähnlichen Situationen zu ziehen und darüber hinaus auf das Wesen oder Temperament eines Tieres zu schließen (BENNETT, 2010).

Der Nutzen solcher Tests wird jedoch kontrovers diskutiert: Einerseits wird angenommen, dass Verhaltenstests geeignete und weithin akzeptierte Mittel sind, um ein relativ gutes Bild des Verhaltens, der kompetitiven Möglichkeiten und der aggressiven Kommunikation einzelner Hunde sowie die Beziehung zu deren Haltern zu ermöglichen (FEDDERSEN-PETERSEN, 2004; JONES und GOSLING, 2005; BARNARD et al., 2012). Andererseits gelten sie als nicht aussagekräftig, solange sie

nicht ausreichend standardisiert und evaluiert wurden, was viele dieser Verhaltenstests, insbesondere für Tierheimhunde, nicht sind (VAN DEN BERG et al., 2003; JONES und GOSLING, 2005; TAYLOR und MILLS, 2006; DIESEL et al., 2008; BENNETT et al., 2012).

2.2.2 Einsatz von Verhaltenstests in der Literatur

Mit ihrer zunehmenden Anzahl werden Vergleiche zwischen den Verhaltenstests immer schwieriger (DIEDERICH und GIFFROY, 2006), vor allem, wenn diese mit unterschiedlichen Zielsetzungen entwickelt wurden. So gibt es in der Literatur viele verschiedene Tests für Zuchthunde sowie für Gebrauchshunde wie Polizei-, Jagd-, Blinden-, Behindertenbegleit- und Therapiehunde (WILSSON und SUNDGREN, 1997a, 1997b; NETTO und PLANTA, 1997; WEISS und GREENBERG, 1997; MURPHY, 1998; SLABBERT und ODENDAAL, 1999; SERPELL und HSU, 2001; JONES und GOSLING, 2005). In den vergangenen 15 Jahren wuchs aber auch das Interesse an Verhaltenstests, um die Eignung von Hunden als Haustiere zu bestimmen. Die meisten dieser Tests beschäftigen sich vor allem mit der Feststellung von Problemverhalten wie Aggressionen, was auch politisch im Zuge der Gefahrenvermeidung durch gefährliche Hunde von Bedeutung ist (FEDDERSEN-PETERSEN, 2004; TAYLOR und MILLS, 2006). NETTO und PLANTA (1997), VAN DEN BERG et al. (2003) und KROLL et al. (2004) entwickelten beispielsweise Verhaltenstests, um spezielle Wesensmerkmale wie Umgänglichkeit und Aggressivität zu messen. VAN DER BORG (1991) und LEDGER und BAXTER (1997) haben dagegen Tests entwickelt, welche allgemeine Informationen über das Wesen des Hundes ermitteln sollten, um optimale Besitzer zu finden.

Einsatz und Nachteile von Testbatterien

Am häufigsten kommen dabei Test-Batterien zum Einsatz, bei denen die Reaktionen der Hunde in einer Reihe von standardisierten Testsituationen oder Subtests ermittelt werden, wodurch bestimmte Verhaltensweisen erkannt werden können (JONES und GOSLING, 2005; DOWLING-GUYER et al., 2011). Diese sind als relativ objektiv zu betrachten (JONES und GOSLING, 2005). Viele dieser Test-Batterien sind bemerkenswert ähnlich aufgebaut, was darauf zurückzuführen ist, dass oftmals Elemente oder sogar ganze Tests von früheren Arbeiten übernommen werden. So

haben zum Beispiel NETTO und PLANTA (1997) und VAN DEN BERG et al. (2003) einige Elemente von dem 1991 von VAN DER BORG entwickelten Test übernommen (TAYLOR und MILLS, 2006). Auch der 2003 von VAN DEN BERG entwickelte Verhaltenstest zur Korrelation von Genetik und Aggressivität bei Golden Retrievern sowie der 2007 von DEMEESTER und PLANTA entwickelte, sogenannte SAB-Test ("sociable acceptable behavior"-Test) für aggressives Verhalten bei Haushunden bauen auf dem Test von NETTO und PLANTA (1997) auf. Dieser hatte sich in der Pilotstudie als zu lang und stressreich erwiesen, was zu vermehrten Aggressionsreaktionen geführt hatte. So zeigten 97% der Hunde im Test aggressives Verhalten, obwohl nicht alle eine aggressive Vorgeschichte laut Besitzerangabe hatten. Außerdem bissen 67% der Hunde in mindestens einer der Testsituationen, wobei das aggressive Verhalten im Laufe des Tests zunahm (NETTO und PLANTA, 1997). Im Vergleich löste der relativ lange Verhaltenstest von NETTO und PLANTA (1997) mit seinen zahlreichen Subtests signifikant häufiger aggressives Verhalten aus und hatte eine höhere falsch-positive Ergebnisrate als die gekürzte Variante dieses Tests von VAN DEN BERG et al. (2003). Dieses Ergebnis lässt darauf schließen, dass mit steigender Anzahl der Subtests auch der Stresslevel der Hunde ansteigt, was zu einer erhöhten Reaktionsbereitschaft und zu aggressivem Verhalten führt (VAN DER BORG et al., 2010). Aber auch der kürzere SAB-Test hat sich nur als suboptimal herausgestellt (VAN DER BORG et al., 2010), da er eine geringe Sensitivität[4] und Spezifität[5] im Vergleich mit einem Besitzerfragebogen aufweist, um Aggressionsverhalten vorherzusagen. Andere Verhaltensweisen, vor allem Angstreaktionen, werden dagegen recht gut erkannt, wodurch der SAB-Test durchaus eine grobe Einschätzung von zukünftigem Problemverhalten geben kann (PAROZ et al., 2008).

Aber trotz der hohen Anzahl solcher Verhaltenstests in der Literatur haben sich bisher keine Standard-Stimuli durchgesetzt, um ein bestimmtes Verhalten zu provozieren. Die Auswahl der Stimuli liegt stattdessen in den Händen des Testentwicklers (DIEDERICH und GIFFROY, 2006).

---

[4] = der Anteil der korrekt als vorhanden erkannte Verhaltensweisen an der Gesamtheit der tatsächlich vorhandenen Verhaltensweisen

[5] = der Anteil an korrekt als nicht vorhandenen erkannten Verhaltensweisen an der Gesamtheit der tatsächlich nicht vorhandenen Verhaltensweisen

Ein großer Nachteil von solchen Test-Batterien ist jedoch, dass häufig keine Aussagen über bestimmte Verhaltensmerkmale wie intraspezifische Aggressionen, Jagdverhalten oder territoriale Aggressionen getroffen werden können, da ihr Testaufbau solche Situationen nicht möglich machen (BENNETT, 2010; VAN DER BORG et al., 2010). Auch müssen bei anderen Testsituationen, insbesondere wenn Angst oder aggressives Verhalten provoziert werden soll, aufgrund von Sicherheitsmaßnahmen Modelle wie eine künstliche Hand oder eine Puppe benutzt werden (KROLL et al., 2004; DIEDERICH und GIFFROY, 2006). Diese Modelle sollen Kinder oder Hände simulieren, können aber nie genaue Duplikationen sein, da sie nicht exakt gleich aussehen, riechen oder sich bewegen (KROLL et al., 2004; BARNARD et al., 2012). Daher kommt es häufig zu falsch-positiven oder falsch-negativen Ergebnissen, wie KROLL et al. (2004) in einer Studie festgestellt haben. Laut dieser Studie wirken die Modelle mitunter nicht echt genug auf die Hunde, um ähnliche Reaktionen in vergleichbaren realen Situationen hervorzurufen. Zudem kann das Modell von den Hunden auch als Spielzeug betrachtet werden und nicht als Bedrohung, die es in bestimmten Testsituationen darstellen soll. Die Benutzung eines Verhaltenstests mit Modellen kann daher nur einen Trend im Aggressionsverhalten eines Hundes wiedergeben, aber nie absolut sicher alle Hunde ermitteln, die aggressiv reagieren. Hunde, die bei den Modellen aggressiv reagieren, müssen nicht zwangsläufig auch in der Realität aggressiv sein (KROLL et al., 2004).

Futteraggressionstests für Tierheimhunde in der Literatur

Neben den oben erwähnten Tests wurden viele weitere Verhaltenstests entwickelt, die vor allem zukünftiges Problemverhalten bei Tierheimhunden erfassen sollen. Vor allem Tierheime verlassen sich mehr und mehr auf die Ergebnisse solcher Tests, um ihre Hunde bestmöglich zu vermitteln (JONES und GOSLING, 2005). Ein Subtest, bei dem Futteraggression ermittelt werden soll, gehört dabei zu den 20 am häufigsten eingesetzten Subtests (TAYLOR und MILLS, 2006). So ist ein solcher Subtest zur Ermittlung von Futteraggression zum Beispiel Teil des SAFER Aggression Assessment von WEISS (2007), des von STERNBERG (2002) entwickelten "Assess-A-Pet", welcher erst 2008 von BOLLEN und HOROWITZ evaluiert wurde oder des Match-Up II Shelter Dog Rehoming Programm des Center for Shelter Dogs von DOWLING-GUYER et al. (2011). Diese Subtests haben gemeinsam, dass ressourcenbezogenes Aggressionsverhalten bei den Hunden provoziert wird, indem

Futter oder ein Kauartikel gereicht und anschließend mithilfe einer Gummihand an einem Stock mit dem Hund und dem Futter interagiert wird, während der Hund frisst (MARDER et al., 2013). Leider hat sich in einer weiterführenden Studie gezeigt, dass der SAFER und der Assess-a-Pet, die beide in den USA am häufigsten zur Beurteilung von Tierheimhunden eingesetzt werden (JONES und GOSLING, 2005), nur eine geringfügige Sensitivität und Spezifität aufweisen, wobei der Assess-a-Pet ein wenig besser aggressive Hunde erkennen kann als der SAFER (BENETT et al., 2012). Die Ergebnisse des SAFER korrelierten bei der Evaluierung durch BENETT et al. (2012) nicht signifikant mit den Beiß-Historien der Hunde, während der Assess-a-Pet zumindest statistisch mit den Historien korrelierte. Der SAFER erkannte außerdem nicht zuverlässig Hunde mit moderater Aggression, die mithilfe einer Verhaltenstherapie durchaus adoptionsfähig gemacht werden könnten. So stellte sich der Assess-a-Pet von STERNBERG (2002) bei der vergleichenden Untersuchung von BENNETT et al. (2012) in allen Punkten der Evaluierung als besser heraus als der SAFER von WEISS (2007). Da jedoch in beiden Tests als aggressiv eingestufte Tiere oftmals nicht vermittelt und stattdessen sofort euthanasiert werden, kann keine Aussage über ihre falsch-positiven Ergebnisse getroffen werden (BOLLEN und HOROWITZ, 2008).

Notwendigkeit der Ergänzung und Evaluierung von Verhaltenstests

Die meisten Autoren jüngerer Studien sind sich darum mittlerweile einig, dass die Ergebnisse eines Verhaltenstests nur als zusätzliches Hilfsmittel zur Beurteilung von Hundeverhalten zu betrachten und diese durch zusätzliche Informationen zu ergänzen sind (KROLL et al., 2004; VAN DER BORG et al., 2010; VALSECCHI et al., 2011; BENNETT et al., 2012; MARDER et al., 2013). Diese können Besitzerfragebögen oder die Erfahrungen des Tierheimpersonals mit dem betreffenden Hund einschließen (BENNETT, 2010).

So wird es mittlerweile auch mit dem "Niedersächsischen Wesenstest" gehandhabt, bei dem die Ergebnisse der Testsituationen mit einem ergänzenden, umfangreichen Besitzerfragebogen in Zusammenhang gebracht werden. Der "Niedersächsische Wesenstest", welcher ursprünglich als Zuchttauglichkeitstest für Hunde bestimmter Rassezugehörigkeit entwickelt worden war, wird in leicht veränderter Form bis heute als Test auf gestörte aggressive Kommunikation eingesetzt. Ziel dieses Verhaltenstests ist es, Hunde mit gestörtem Sozialverhalten, insbesondere einem

unakzeptablem und situativ unangemessenem Aggressionsverhalten, zu erkennen (NIEDERSÄCHSICHES MINISTERIUM FÜR ERNÄHRUNG, LANDWIRTSCHAFT UND FORSTEN, 2000). Unangemessenes Aggressionsverhalten wird hierbei definiert als das Fehlen von bestimmten Eskalationsstufen, wonach Aggressionsverhalten nicht mehr als Anpassungsmechanismus und soziales Regulativ dient, sondern stattdessen biologisch und in der Genese nicht mehr nachvollziehbar scheint und unvermittelt, plötzlich und in extremer Ausprägung auftritt (FEDDERSEN-PETERSEN, 2004). Laut FEDDERSEN-PETERSEN (2004) haben diese Hunde, die ohne vorheriges Drohverhalten angreifen, ein gestörtes Sozialverhalten und stellen für ihre Umwelt ein erhöhtes Gefährdungspotenzial dar. Auch die Einwirkungsmöglichkeit und -fähigkeit des Halters auf den Hund ist wichtig bei der Beurteilung des Niedersächsischen Wesenstests. Deswegen wird bei diesem auch besonderes Augenmerk auf das Hund-Halter-Gespann und ihre Beziehung gelegt, um ein mögliches Gefahrenpotential durch die Hund-Halter-Kombination zu erfassen (FEDDERSEN-PETERSEN, 2004).

Da jedoch alle Verhaltenstests mehr oder weniger lang und stressreich für die Tiere sind, sind falsch-positive Ergebnisse immer möglich. Einige Autoren empfehlen daher mittlerweile, keine voreiligen Schlüsse aus den Ergebnissen eines Verhaltenstests zu ziehen oder deren Ergebnisse über zu bewerten und nur aufgrund eines einzelnen Tests über die Zukunft oder das Leben des Hundes zu schließen (KROLL et al., 2004; VAN DER BORG et al., 2010; VALSECCHI et al., 2011; BENNETT et al., 2012; MARDER et al., 2013). Stattdessen sollten die Verhaltenstests auf ihre Aussagekraft hin geprüft und evaluiert werden (TAYLOR und MILLS, 2006; DIEDERICH und GIFFROY, 2006). Außerdem könnten einzelne Tests wiederholt werden, um die Beständigkeit von Verhaltensweisen zu überprüfen (TAYLOR und MILLS, 2006; FEDDERSEN-PETERSEN, 2008) oder es könnten die Ergebnisse verschiedener Tests miteinander kombiniert werden, um ein möglichst breites Bild vom Verhaltensrepertoire der Hunde zu erhalten (FEDDERSEN-PETERSEN, 2004; BOLLEN und HOROWITZ, 2008; BENNETT et al., 2012).

2.3 Evaluierung von Verhaltenstests

2.3.1 Evaluierung der Verhaltenstests in der Literatur

Der Wert eines Verhaltenstests und die Interpretation seiner Ergebnisse hängen von seiner Standardisierung, seiner Reliabilität und seiner Validität ab. Bei einem Verhaltenstest sollten die Durchführung des Tests, die Aufnahme des Verhaltens und die Auswertung der Daten so standardisiert sein, sodass das Verhalten des Hundes zur einzigen Variablen wird. Zudem muss ein Verhaltenstest reliabel sein, wodurch seine Ergebnisse bei wiederholter Durchführung reproduzierbar werden. Außerdem sollte ein Verhaltenstest valide sein, also wirklich nur jenes Verhalten messen, welches er messen soll (JONES und GOSLING, 2005; DIEDERICH und GIFFROY, 2006; BENNETT, 2010). Diese drei Qualitätsmerkmale stehen in logischem Zusammenhang, wonach ein Test erst valide sein kann, wenn er reliabel ist, was wiederum erst erreicht werden kann, wenn er standardisiert ist (TAYLOR und MILLS, 2006).

Die Überprüfung dieser Qualitätsmerkmale wurde jedoch bei den bestehenden Tests in der Literatur teilweise vernachlässigt. So gibt es wenige, bei denen Standardisierung, Reliabilität, Validität und Praktikabilität ausreichend geprüft wurde (VAN DEN BERG et al., 2003; JONES und GOSLING, 2005; TAYLOR und MILLS, 2006; DIESEL et al., 2008; BENNETT et al., 2012). Meistens wurde nur ein Aspekt dieser Qualitätsmerkmale für Verhaltenstests näher untersucht, vor allem die Kriteriumsvalidität im Vergleich mit Besitzerfragebögen oder die Vorhersagekraft mithilfe von Sensitivität und Spezifität (PLANTA und DEMEESTER, 2007; BOLLEN und HOROWITZ, 2008; VAN DER BORG et al., 2010; BENNETT et al., 2012). Die Qualitätspunkte der Reliabilität wurden dabei aber häufig außer Acht gelassen, wodurch nicht klar ist, inwieweit die Ergebnisse von Unbeständigkeiten im Hinblick auf den Tester, den Hund, den Gutachter oder das Tierheim beeinflusst wurden (TAYLOR und MILLS, 2006). So findet sich in der Literatur kein Test, bei dem alle Qualitätsmerkmale der Validität und Reliabilität untersucht worden sind (VAN DEN BERG et al., 2003; JONES und GOSLING, 2005; TAYLOR und MILLS, 2006; DIESEL et al., 2008; BENNETT et al., 2012).
Viele Tierheime und Tierfangauffangstationen in den USA benutzen zudem ihre eigens entwickelten Tests, wovon viele ohne Hilfe von Verhaltensspezialisten entwickelt und

nicht publiziert wurden. Diese gängige Praxis in den Tierheimen der USA ist zu bedenken, da sie den Ergebnissen ihrer nicht oder nur unzureichend evaluierten und auf ihre Aussagekraft untersuchten Tests große Bedeutung zusprechen und von diesen die Zukunft und das Leben vieler tausend Hunde pro Jahr abhängt (JONES und GOSLING, 2005; TAYLOR und MILLS, 2006; MOHAN-GIBBONS et al., 2012). Wichtig wäre außerdem, dass die Verhaltenstests nicht nur in den einzelnen Tierheimen evaluiert werden, in denen sie Anwendung finden, um Unterschiede bei der Durchführung zwischen den Tierheimen zu reduzieren. Es müssen auch der Gesamterfolg und die eventuell vorgenommenen Veränderungen des Tests evaluiert werden. Nur so ist eine Optimierung des Verhältnisses zwischen Aufwand und Nutzen möglich (TAYLOR und MILLS, 2006).

Von den wenigen evaluierten Tests in der Literatur kann ein hoher Anteil dennoch eine moderate bis gute Aussagekraft aufweisen, vor allem jene, die spezielle Verhaltensprobleme wie Furcht oder Aggression messen sollten (TAYLOR und MILLS, 2006). Dennoch variiert die Methode der Evaluierung stark und nicht alle Autoren bestimmen die gleichen, eindeutigen Qualitätsmerkmale, was einen Vergleich zwischen den verschiedenen Tests erschwert (TAYLOR und MILLS, 2006; DIEDERICH und GIFFROY, 2006). Denn obwohl es zahlreiche Richtlinien und Regeln für die Durchführungen von Tests in anderen Wissenschaften gibt, waren solche für Verhaltenstests bei Hunden bis ins Jahr 2006 nicht beschrieben (DIEDERICH und GIFFROY, 2006). TAYLOR und MILLS (2006) veröffentlichten jedoch in diesem Jahr eine Arbeit, in denen sie allgemeine, wissenschaftlich korrekte Herangehensweisen an die Entwicklung von Verhaltenstests und deren Evaluierung an Beispielen von verschiedenen in der Literatur zu findenden Tests erörterten. Dazu beschrieben sie den Sinn und die Vorgehensweise zur Standardisierung von Tests und zur Ermittlung der Reliabilität, welche in Test-Retest-Reliabilität, Inter-rater-Reliabilität und Intra-rater-Reliabilität unterteilt werden kann. Außerdem wird darin die Bedeutung und die Herangehensweise zur Bestimmung der Validität beschrieben, welche wiederum in Inhaltsvalidität, Konstruktvalidität und Kriteriumsvalidität eingeteilt werden kann, sowie die Notwendigkeit der Beachtung der Praktikabilität und des Tierwohls bei der Entwicklung von Verhaltenstests diskutiert. Auf diese Qualitätsmerkmale wird darum auch im Folgenden näher eingegangen.

2.3.2 Standardisierung von Verhaltenstests

Die Standardisierung ist eine wichtige Voraussetzung für die Reliabilität, die Reproduzierbarkeit, eines Verhaltenstests. Sie wird mit einem Durchführungs-protokoll erreicht, welches alle erdenklichen Varianzen zwischen den Tests ausschließt, sodass das Verhalten des Hundes die einzige Variable wird (DIEDERICH und GIFFROY, 2006). Die Varianzen schließen dabei den Ort, die Zeit und alle äußeren Umweltreize mit ein. Wichtig ist, dass diese Bedingungen bei allen Hunden gleich sind, weswegen sie im Voraus festgelegt werden müssen. So kann es je nach Fragestellung und Testziel in Bezug auf die Standardisierung sinnvoll sein, sich zunächst auf eine einzige Rasse zu beschränken, um Variationen im Verhalten aufgrund unterschiedlicher Rassedispositionen und Genetik zu reduzieren. Um aber eine Aussage über Hundeverhalten im Allgemeinen zu treffen, ist es notwendig, verschiedene Rassen zu testen, um Rassedispositionen zu übergehen (DIEDERICH und GIFFROY, 2006). Um eine gute Standardisierung zu erreichen, müssen außerdem die Datenaufnahme und die Auswertung der gezeigten Verhaltensweisen vereinheitlicht werden, egal ob jedes Verhalten deskriptiv aufgenommen oder mittels Skalen, Punkten oder Kategorien bewertet wird. Dabei muss darauf geachtet werden, dass zum Teil komplexes Verhalten nicht zu sehr reduziert wird, was zu einem schlechteren Vorhersagewert führen würde (TAYLOR und MILLS, 2006).

Das gezeigte Verhalten kann dabei auf unterschiedliche Weise aufgezeichnet und bewertet werden: objektives Bewerten von Frequenz und Dauer des Verhaltens, Einteilung mit abgestuften Punkten nach Intensität des Verhaltens, oder subjektiv durch direkte Beobachtung, welche abhängig ist vom Übungsgrad und Wissensstand des Beobachters. Dabei ist die subjektive Datenaufnahme meist schwer zu vereinheitlichen, wenn andere Personen auswerten als der Entwickler des Tests (DIEDERICH und GIFFROY, 2006). Es können starke Varianzen in der Bewertung auftreten, da die obligatorische objektive und vorurteilsfreie Beschreibung des Verhaltens häufig fehlt und das beobachtete Verhalten stattdessen sogleich gewertet und interpretiert wird (FEDDERSEN-PETERSEN, 2004). Das Problem der subjektiven Bewertung von Verhalten kann aber umgangen werden, indem die einzelnen Verhaltensweisen in einem Ethogramm genau beschrieben, nach Intensität eingeteilt und diese wiederum genau definiert werden. Doch auch bei der Definition und Einteilung von einzelnen Verhaltensweisen existieren bisher keine Vorgaben oder

Richtlinien, weshalb sich diese auch noch vielfach bei den Bewertern unterscheiden (FEDDERSEN-PETERSEN, 2004; JONES und GOSLING, 2005; DIEDERICH und GIFFROY, 2006). Daher ist eine Weiterbildung für das Tierheimpersonal in der Ethologie von Haushunden empfehlenswert, um ein biologisch korrektes und objektives Zuordnen von Ausdrucksverhalten in bestimmte situative Kontexte zu ermöglichen (FEDDERSEN-PETERSEN, 2004).

Die Standardisierung kann als Minimum an Qualität für einen Test angesehen werden, welche von den meisten in der Literatur zu findenden Tests weitestgehend erreicht wird (TAYLOR und MILLS, 2006; DIEDERICH und GIFFROY, 2006). Es kann sich jedoch als schwierig herausstellen, einen Test komplett zu standardisieren, da das Verhalten des Hundes auch stark von der Interaktion mit der Testperson abhängt. So können auch kleine Unterschiede im Verhalten der Testperson das Ergebnis beeinflussen. Ein Hund ist daher besser zu beurteilen, wenn die Testperson ein stabiles und einheitliches Verhalten während des Tests an den Tag legt (TAYLOR und MILLS, 2006). Besonders schwer ist es deshalb, einen Verhaltenstest im häuslichen Umfeld und mit dem Besitzer zu standardisieren (SVARTBER et al., 2005), da die Testbedingungen und das Verhalten des Besitzers stark variieren können.

Die Tests in der Literatur erfüllen in der Regel den Anspruch auf Standardisierung, indem die jeweilige Durchführung protokolliert und festgelegt ist. Messfehler sind dennoch immer möglich. So können Unterschiede in Anzahl und Art von beobachteten Verhalten an der Durchführung des Tests an sich oder der Ermittlung der Daten liegen, hierbei insbesondere an der Anzahl der Bewerter, einer unterschiedlichen Beobachtung (direkt vs. Video) oder an der Länge und Frequenz von Testeinheiten (DIEDERICH und GIFFROY, 2006).
Ein weiterer Kritikpunkt ist die schnelle Abfolge der zum Teil recht stressreichen Testsituationen, die kaum Zeit für Entspannungsmomente lässt (FEDDERSEN-PETERSEN, 2004). Die meisten Hunde befinden sich in einem Tierheim ohnehin wegen der Haltungsbedingungen, der Hektik, dem hohen Maß an Unbekannten und dem Mangel an Fluchtmöglichkeit in einer höheren Bereitschaft für defensiv aggressives Verhalten (TAYLOR und MILLS, 2006). In den Verhaltenstests werden dann zum Teil sehr gefährliche Momente erzeugt, in denen sich das Verhalten der

Hunde durch Erlernen von Stressbewältigungsstrategien im schlimmsten Fall auch negativ verändern kann (FEDDERSEN-PETERSEN, 2004).

2.3.3 Reliabilität von Verhaltenstests

Bei der Überprüfung der Reliabilität wird die Frage untersucht, ob die Messungen reproduzierbare, also wiederholbare Ergebnisse erzielen. Hierbei werden drei Arten von Reliabilität untersucht: die Test-Retest-Reliabilität, die Inter-rater-Reliabilität und die Intra-rater-Reliabilität (BENNETT, 2010). Diese untersuchen die Beständigkeit der Ergebnisse innerhalb der einzelnen Testobjekte bei wiederholter Durchführung (Test-Retest-Reliabilität), zwischen unterschiedlichen Bewertern (Inter-rater-Reliabilität) und innerhalb der Bewerter bei wiederholter Auswertung (Intra-rater-Reliabilität) (TAYLOR und MILLS, 2006). Damit ist die Reliabilität ein Maß für die Wiederholbarkeit und Beständigkeit einer Messung. Mit ihrer Bestimmung wird außerdem untersucht, ob die Ergebnisse frei von Messfehlern und Verzerrungen sind (MARTIN und BATESON, 2007). Die Bestimmung der Reliabilität eines Verhaltenstests ist sehr wichtig, da sie eine Voraussetzung für die Validität eines Tests ist (DIEDERICH und GIFFROY, 2006). Sie wird mithilfe von Korrelations-Koeffizienten angegeben, wobei diese sowohl für jede einzelne Verhaltensweise als auch für Kategorien von Verhaltensweisen berechnet werden können (MARTIN und BATESON, 2007).

Test-Retest-Reliabilität

Die Test-Retest-Reliabilität ist - im Falle eines Verhaltenstests für Hunde - ein Maß für die Wahrscheinlichkeit, dass sich ein Hund bei einer Wiederholung desselben Tests zu einem anderen Zeitpunkt gleich verhält (TAYLOR und MILLS, 2006). Bei der Ermittlung der Test-Retest-Reliabilität wird also derselbe Test mit derselben Testpopulation wiederholt und die Ergebnisse beider Testdurchläufe werden auf Übereinstimmungen oder Abweichungen bezüglich des gezeigten Verhaltens, deren Frequenz, Dauer oder Intensität hin untersucht. So wird überprüft, ob die Testergebnisse reproduzierbar sind oder ob das im Test gezeigte Verhalten nur zufällig gezeigt wurde (DIEDERICH und GIFFROY, 2006).

Über eine Zeit hinweg gleichbleibendes Verhalten ist ein Schlüsselpunkt für das Konzept von Verhaltensmerkmalen und damit auch von Verhaltenstests. Eine hohe Test-Retest-Reliabilität belegt, dass ein Test resistent ist gegen Abweichungen im

Hinblick auf die verstrichene Zeit, weshalb sie einen hohen Vorhersagewert hat. (ROST, 2004; SVARTBERG et al., 2005; TAYLOR und MILLS, 2006). Denn erst wenn ein Hund in derselben Situation auch dasselbe Verhalten zeigt, kann mit einer bestimmten Wahrscheinlichkeit vorhergesagt werden, dass er auch in ähnlichen Situationen genau dieses Verhalten zeigt. Trotzdem sind viele Einflussfaktoren auf das Hundeverhalten zu berücksichtigen, z.B. Stress, Hunger, Krankheit oder andere äußere Einflussfaktoren. Deswegen ist es wichtig, den Test so detailliert wie möglich zu beschreiben und zu standardisieren, damit alle Varianzen bei der Durchführung des Tests ausgeschlossen werden können, was die Test-Retest-Reliabilität erhöht.

Der Nachteil an einem Retest ist der Lerneffekt, dem alle Probanden unterliegen und der zu zum Teil erheblichen Abweichungen in den Ergebnissen führen kann (DIEDERICH und GIFFROY, 2006). So kann zum einen eine Gewöhnung an neue Reize, z.B. an unbekannte Objekte, die Testperson oder die Testbedingungen im ersten Test, das Verhalten im zweiten Test beeinflussen und zu einer schlechteren Übereinstimmung führen (SVARTBERG, 2005, TAYLOR und MILLS, 2006). Zum anderen kann erfolgreiches Verhalten im ersten Test zu einem veränderten erlernten Verhalten im zweiten Test führen (TAYLOR und MILLS, 2006). Zudem kommt, dass Versuchsergebnisse schon allein aufgrund der natürlichen Variabilität nicht genau reproduzierbar sind. So erzielen Verhaltenstests in der Wiederholung nie dasselbe Ergebnis. Der Grund dafür liegt in den Rahmenbedingungen der Tests, die nie gänzlich standardisierbar und mehr oder weniger mit Zufallsvariablen belastet sind. Ein standardisierter Test unter Laborbedingungen wäre jedoch ebenfalls unzweckmäßig und nicht verwertbar, da ein Verhaltenstest die Reaktion eines Hundes in Alltagssituationen erfassen soll (FEDDERSEN-PETERSEN, 2004).

In der Literatur werden Retests nicht systematisch durchgeführt. So untersuchten lediglich NETTO und PLANTA (1997), SVARTBERG (2005), VAN DER BERG et al. (2010), OTT (2010) und VALCSECCHI et al. (2011) die Test-Retest-Reliabilität, wobei die Zeitabstände zwischen einer Stunde und einem halben Jahr variierten. Aus diesem Grund ist bisher nicht festgelegt, in welchem Zeitrahmen ein Retest durchgeführt werden sollte, ob nach Stunden, Tagen oder Wochen. Aus lerntheoretischer Sicht sollte ein Retest nicht unmittelbar nach dem ersten Test durchgeführt werden, da die Wahrscheinlichkeit von Gewöhnungs- und Lerneffekten hoch wäre (TAYLOR und

MILLS, 2006). Ist die Zeit zwischen den Tests jedoch zu lang, könnten Unterschiede im Verhalten auf anderweitige Lernerfahrungen, innere Umständen wie Krankheit oder Alter oder veränderte Haltungsbedingungen wie etwa nach einem Besitzerwechsel zurückzuführen sein.

Inter-rater-Reliabilität

Bei der Inter-rater-Reliabilität wird überprüft, ob mehrere Bewerter bei der Beurteilung desselben Tests auch zu denselben Ergebnissen kommen (TAYLOR und MILLS, 2006; DIEDERICH und GIFFROY, 2006; MARTIN und BATESON, 2007; BENNETT, 2010). Dazu bewerten verschiedene Personen denselben Test unabhängig voneinander, woraufhin die Ergebnisse auf Übereinstimmung geprüft werden (DIEDERICH und GIFFROY, 2006; MARTIN und BATESON, 2007).

Vor allem in Tierheimen und Auffangstationen werden Verhaltenstests von verschiedenen Personen durchgeführt und bewertet. Darum ist es besonders wichtig, die Übereinstimmung der Beurteilungen zwischen verschiedenen Bewertern für diese Verhaltenstests zu überprüfen, da große Varianzen zwischen den Bewertern auftreten können (MURPHY, 1998; TAYLOR und MILLS, 2006; DIEDERICH und GIFFROY, 2006; MARTIN und BATESON, 2007). Auch hier sind Videoaufnahmen hilfreich, wenn nicht alle Bewerter zur selben Zeit denselben Hund bewerten können (TAYLOR und MILLS, 2006; MARTIN und BATESON, 2007). Vergleiche zwischen Echtzeit-Beobachtungen und Videoaufnahmen sollten vermieden werden, da beide Methoden unterschiedliche Arten von Schwachstellen aufweisen, wodurch die Übereinstimmung in der Regel eher schlecht ist (TAYLOR und MILLS, 2006). Um die Inter-rater-Reliabilität zu verbessern, kann es hilfreich sein, wenn die Bewerter vor der eigentlichen Beurteilung unterwiesen werden und eventuell Übungsvideos auswerten (TAYLOR und MILLS, 2006; DIESEL et al., 2008; VALSECCHI et al., 2011). Außerdem ist es sinnvoll, Bewerter mit unterschiedlichem Maß von Vorerfahrung in der Auswertung von Verhaltenstests zu wählen, um eine Verzerrung (Bias) der Ergebnisse aufgrund von Erfahrung und Übung in der Auswertung zu reduzieren (KROLL et al., 2004).

Eine weitere, häufig vorkommende Quelle für eine mangelhafte Inter-rater-Reliabilität liegt in der Definition und Beschreibung von Verhaltensweisen. Darum ist es

besonders wichtig, im Vorfeld die zu messenden Verhaltensweisen festzulegen und genauestens zu definieren (MARTIN und BATESON, 2007).

Die Überprüfung der Inter-rater-Reliabilität wurde in vielen der in der Literatur zu findenden Tests nicht durchgeführt. Lediglich VAN DER BORG et al. (2010) überprüfte die Inter-rater-Reliabilität für den 2007 von PLANTA und DEMEESTER entwickelten SAB-Test und VALSECCHI et al. (2011) für einen Test für Tierheimhunde.

Intra-rater-Reliabilität

Die Intra-rater-Reliabilität beschreibt das Maß an Einheitlichkeit der Ergebnisse eines einzelnen Bewerters (MARTIN und BATESON, 2007). Es wird also überprüft, ob ein Bewerter bei wiederholter Auswertung desselben Tests erneut zu denselben Ergebnissen kommt (TAYLOR und MILLS, 2006; DIEDERICH und GIFFROY, 2006). Dazu wertet ein Bewerter einen Test ein zweites Mal oder mehrere Male aus, woraufhin die Ergebnisse anschließend auf Übereinstimmung geprüft werden (MARTIN und BATESON, 2007). Um sicherzustellen, dass mögliche Abweichungen der Beurteilungen nicht auf Unterschiede des Hundeverhaltens sondern auf den Bewerter zurückzuführen sind, ist es auch hier sinnvoll, Videoaufnahmen der Tests zu verwenden. So kann der Bewerter ein und denselben Test mehrfach zu unterschiedlichen Zeiten auswerten (TAYLOR und MILLS, 2006; MARTIN und BATESON, 2007).

Die Bestimmung der Intra-rater-Reliabilität ist eine grundlegende Messung der Reliabilität und ein wichtiges Qualitätsmerkmal, da sie Voraussetzung für eine hohe Test-Retest-Reliabilität ist. Sie ist vor allem von Bedeutung, wenn subjektive Meinungen und Interpretationen Teil des Bewertungssystems des Tests sind (TAYLOR und MILLS, 2006). Nur wenige in der Literatur zu findende Tests wurden explizit auf ihre Intra-rater-Reliabilität hin überprüft. So wurde sie für den 2007 von PLANTA und DEMEESTER entwickelten SAB-Test 2010 von VAN DER BORG et al. ermittelt, sowie von einem Test für Tierheimhunde von VALSECCHI et al. 2011.

2.3.4 Validität von Verhaltenstests

Die Validität behandelt die Frage, in welchem Maß ein Test auch das misst, was er messen soll (JONES und GOSLING, 2005; TAYLOR und MILLS, 2006; MARTIN und BATESON, 2007; BENNETT, 2010). Um als valide zu gelten, muss ein Test das messen, wozu er zu messen entwickelt wurde. Wenn jedoch nicht genau bekannt oder definiert ist, was er messen soll, kann der Test im Umkehrschluss unmöglich valide sein (TAYLOR und MILLS, 2006). Außerdem muss ein Test zunächst reliabel sein, um als valide gelten zu können. Damit ist die Reliabilität eine notwendige Voraussetzung für die Validität eines Tests (JONES und GOSLING, 2005).
Die Validitätsmessung für Verhaltenstests ist schwierig, denn es gilt als unwahrscheinlich, dass ein Test hundertprozentig das Verhalten eines Hundes in allen erdenklichen Situationen und Umweltkonstellationen vorhersagen kann (LEDGER und BAXTER, 1997; KAHN et al., 2003; TAYLOR und MILLS, 2006). Um die Validität eines Verhaltenstests zu bestimmen, wird dessen Inhalts-, Kriteriums- und Konstruktvalidität bestimmt (DIEDERICH und GIFFROY, 2006; TAYLOR und MILLS, 2006; BENNETT, 2010).

Inhaltsvalidität

Die Inhaltsvalidität bewertet die Frage, ob der Test jenes Zielverhalten misst, welches er messen soll und ob die einzelnen Messkomponenten das gesamte Zielverhalten erfassen. Die Beurteilung der Inhaltsvalidität ist besonders wichtig für Verhaltenstests, da diese eine genaue Aussage treffen sollen über Verhaltensmuster des Hundes mithilfe weniger Teststimuli und nur über einen kurzen Zeitraum hinweg (TAYLOR und MILLS, 2006). Sie wird von Verhaltensspezialisten und Fachleuten bewertet, welche prüfen, ob ein logischer Zusammenhang zwischen Testverfahren und Zielstellung gegeben ist (DIEDERICH und GIFFROY, 2006; TAYLOR und MILLS, 2006).
Leider wurden bisher keine der in der Literatur zu findenden Tests auf ihre Inhaltsvalidität hin überprüft, was jedoch schon am Anfang der Entwicklung eines Verhaltenstests berücksichtigt werden sollte (TAYLOR und MILLS, 2006).

Ziel eines Verhaltenstests ist es immer, möglichst in hohem Maße reliabel und valide zu sein, also möglichst genau und präzise ein bestimmtes Verhalten vorherzusagen (JONES und GOSLING, 2005). Es wird jedoch teilweise angenommen, dass die

Validität eines Tests sogar sinkt, wenn Maßnahmen zur Erhöhung der Reliabilität vorgenommen werden (ROST, 2004). Diese Inkompatibilität von gleichzeitiger Erhöhung der Reliabilität und Validität wird von ROST (2004) als Reliabilitäts-Validitäts-Dilemma bezeichnet. Dieses beschreibt, dass eine zu starke Steigerung der Reliabilität durch Anpassung der Testdurchführung und Ausschließen aller möglicher Störfaktoren und Messfehlern dazu führt, dass der Test nicht mehr das misst, was er ursprünglich messen soll (ROST, 2004). Übertragen in die Praxis von Verhaltenstests für Hunde würde dies bedeuten, dass zu sehr theoretisierte und standardisierte Testbedingungen zu weit entfernt sind von realitätsnahen und komplexen Umwelteinflüssen, die ursprünglich im Verhaltenstest simuliert werden sollten. Diese Gratwanderung zwischen der Optimierung der Messgenauigkeit und der inhaltlichen Logik des Tests ist eine große Herausforderung bei der Entwicklung neuer Verhaltenstests.

Kriteriumsvalidität

Bei der Bestimmung der Kriteriumsvalidität (auch externe Validität, Vorhersagevalidität) wird überprüft, inwiefern die Ergebnisse eines Verhaltenstests mit denen einer externen Messung für dasselbe Zielverhalten übereinstimmen (JONES und GOSLING, 2005; TAYLOR und MILLS, 2006; DIEDERICH und GIFFROY, 2006; BENETT, 2010). In der Regel wird dazu der aktuelle Goldstandard zum Vergleich herangezogen, der selbst reliabel und valide sein sollte. Im Bereich der Verhaltenstests für Hunde existiert jedoch kein solcher Test als Goldstandard, weshalb auf andere externe Messungen als Referenz zurückgegriffen werden muss, deren Reliabilität und Validität unter Umständen nicht überprüft wurden (JONES und GOSLING, 2005; TAYLOR und MILLS, 2006). In diesem Fall wird auch von Konkurrenzvalidität gesprochen (HSU und SERPELL, 2003; SVARTBERG, 2005).

Als Referenztests werden in der Literatur häufig Besitzerfragebögen eingesetzt, die das Verhalten der Hunde erfragen sollen (TAYLOR und MILLS, 2006), so zum Beispiel bei den Arbeiten von VAN DER BORG et al. (1991), NETTO und PLANTA (1997), LEDGER und BAXTER (1997), VAN DEN BERG et al. (2003), PLANTA und DEMEESTER (2007), BOLLEN und HOROWITZ (2008), VAN DER BORG et al. (2010), VALSECCHI et al. (2011) und BENNETT et al. (2012). Hierbei wird meistens nach aggressivem Verhalten der Hunde in der Vergangenheit gefragt oder aber nach

Problemverhalten, sobald der Hund aus dem Tierheim von einem neuen Besitzer adoptiert wurde (TAYLOR und MILLS, 2006). LEDGER und BAXTER fanden 1997 Übereinstimmungen zwischen den Angaben der Besitzer und ihrem Verhaltenstest, ebenso SVARTBERG (2005). Jedoch ist die Validität von Besitzerfragebögen fraglich, da auch diese eigentlich evaluiert werden müssten (VAN DER BORG, 1991; TAYLOR und MILLS, 2006). VAN DER BORG (1991) merkte außerdem an, dass Besitzer gerade bei der Frage nach Problemverhalten wenig aussagekräftige Angaben machen, da sie das Problemverhalten häufig entweder nicht als solches erkennen oder es selbst nicht als problematisch einschätzen, weshalb sie dieses dann nicht im Fragebogen angeben. So hängt die Übereinstimmung von Verhaltenstest und Besitzerfragebogen erheblich vom Besitzer selbst ab, von seinem Wissenstand und seinen Erfahrungen über und mit dem Hund sowie von seinem eigenen Charakter.

Wird ein Testergebnis mit einer externen Messung oder einer Referenz verglichen, die den wahren Zustand des Hundes wiedergibt, ist eine Aussage über die Vorhersagekraft, welche als Teil der Kriteriumsvalidität anzusehen ist, möglich. Diese beschreibt das Maß, mit welcher ein Verhaltenstest ein in der Zukunft liegendes Testergebnis korrekt vorhersagen kann (TAYLOR und MILLS, 2006). Sie wird häufig mit den Begriffen Sensitivität und Spezifität beschrieben, so zum Beispiel bei VAN DER BORG et al. (1991), NETTO und PLANTA (1997), PLANTA und DEMEESTER (2007), BOLLEN und HOROWITZ (2008), VAN DER BORG et al. (2010) und BENNET et al. (2012). Diese beiden Punkte werden oft getrennt von der Validität ermittelt, können aber wichtige Gütekriterien für einen Test darstellen.
Manchmal werden parallel zur Sensitivität und Spezifität oder stattdessen auch Aussagen zu den falsch positiven[6] und falsch negativen[7] Ergebnissen gemacht, welche die Anteile der nicht korrekt ermittelten Ergebnisse im Vergleich zu einem Goldstandard oder einer Referenz wiedergeben (TAYLOR und MILLS, 2006).

In der Literatur wurde die Kriteriumsvalidität von VAN DER BORG et al. (1991), NETTO und PLANTA (1997), LEDGER und BAXTER (1997), VAN DEN BERG et al. (2003),

---

[6] =Anteil der falsch als vorhanden erkannte Verhaltensweise an der Gesamtheit der tatsächlich vorhandenen Verhaltensweisen

[7] =Anteil der falsch als nicht vorhanden erkannte Verhaltensweise an der Gesamtheit der tatsächlich nicht vorhandenen Verhaltensweisen

PLANTA und DEMEESTER (2007), BOLLEN und HOROWITZ (2008), VAN DER BORG et al. (2010), VALSECCHI et al. (2011) und BENNETT et al. (2012) ermittelt. Durch diese Vielzahl an unterschiedlichen Herangehensweisen, Messungen oder statistischen Berechnungen ist ein Vergleich der Ergebnisse zwischen den Studien jedoch nahezu unmöglich (JONES und GOSLING, 2005).

Konstruktvalidität

Neben der externen Validität in Form der Kriteriumsvalidität kann aber auch die interne Validität oder auch Konstruktvalidität ermittelt werden, bei der geprüft wird, wie gut einzelne Bestandteile des Tests ein allgemeines Konstrukt oder Merkmal messen (BENNETT, 2010). Im Falle von Verhaltenstests für Hunde ist ein allgemeines Konstrukt oder Merkmal eine bestimmte Verhaltensweise oder ein Funktionskreis von Verhaltensweisen, wie etwa Aggressivität oder Furcht. Dabei wird geprüft, ob die Ergebnisse einzelner Testsituationen sinngemäß in Zusammenhang stehen. So wird geprüft, ob Testsituationen, die dasselbe Merkmal messen sollen, auch zum selben Ergebnis kommen, was Konvergenzvalidität genannt wird, bzw. ob Testsituationen, die gegenteilige Merkmale messen sollen, auch zum gegenteiligen Ergebnis kommen, was als Diskriminanzvalidität bezeichnet wird (TAYLOR und MILLS, 2006; DIEDERICH und GIFFROY, 2006; BENNETT, 2010). Demnach ist die Bestimmung der Konstruktvalidität nur sinnvoll und notwendig, wenn eine Testbatterie mehrere Verhaltensweisen aus verschiedenen Funktionskreisen messen will. Bei Testbatterien, die lediglich eine Verhaltensweise oder einen Funktionskreis messen wollen, zum Beispiel in denen ausschließlich aggressives Verhalten provoziert werden soll, wird die Konstruktvalidität nicht bestimmt.

Obwohl die meisten Tests in der Literatur mehrere Verhaltensmerkmale oder Funktionskreise messen, wurde die Konstruktvalidität lediglich für die Tests von VAN DER BORG et al. (2010), DOWLING-GUYER et al. (2011) und VALSCCHI et al. (2011) überprüft.

2.3.5 Praktikabilität und Tierschutz

Zusätzlich zu den beschriebenen Qualitätsmerkmalen der Standardisierung, Reliabilität und Validität muss ein Verhaltenstest zusätzlich auch praktikabel, einfach und wirtschaftlich in seiner Anwendung sein, egal ob er als Test für Tierheimhunde, zur Zuchtauswahl oder zur Selektion von Diensthunden verwendet wird. Nicht praktikable Tests, die zu lang oder zu kompliziert sind, werden vermutlich nicht oder nicht richtig ausgeführt, wodurch die Reliabilität und der Vorhersagewert sinken. Viele wissenschaftlich entwickelte Tests bedürfen daher einer Überprüfung der Praktikabilität. Um als praktikabel zu gelten, muss ein Verhaltenstest standardisiert, kurz und leicht durchzuführen sein und eine einfache Aufnahme des Hundeverhaltens möglich machen (TAYLOR und MILLS, 2006).

Viele der in der Literatur zu findenden Verhaltenstests sind zwar ursprünglich für Tierheimhunde entwickelt worden, so zum Beispiel die Tests von VAN DER BORG et al. (1991), LEDGER und BAXTER (1997), BOLLEN und HOROWITZ (2008), DOWLING-GUYER et al. (2011), VALSECCHI et al. (2011) und BENNETT et al. (2012), sind aber zu einem großen Teil zu lang, um im Tierheimalltag Einsatz finden zu können. So nehmen einige der Tests bis zu eine Stunde je Hund in Anspruch, etwa jene von VAN DER BORG et al. (1991) und LEDGER und BAXTER (1997). Um wirklich durchführbar zu sein, benötigen die meisten entwickelten Tests daher eine Vereinfachung, bei der die Anzahl an Testsituationen oder Subtests reduziert, die Methode der Aufnahme von Verhalten vereinfacht und die Interpretation der Ergebnisse für individuelle Hunde ermöglicht werden sollten. Gerade wenn der Test eine schlechte Reliabilität aufweist, sollte er gekürzt und vereinfacht werden. Danach muss erneut eine Validitäts- und Reliabilitätsbestimmung vorgenommen werden, da diese für den veränderten Test nicht mehr gelten.

Bei der Bearbeitung eines Verhaltenstests darf auch der Aspekt des Tierwohls nicht außer Acht gelassen werden, vor allem, wenn darin Angst oder Aggression gemessen werden soll (TAYLOR und MILLS, 2006).
Da die Umweltbedingungen in Tierheimen für die Hunde sehr belastend und stressreich sein können, kann das wahre Verhalten unter Umständen verschleiert werden (WEISS und GREENBERG, 1997). Denn Stress hat auf das gezeigte Verhalten der Hunde einen starken Einfluss (BENNETT et al., 2012), sodass das

gezeigte Verhalten in einem Verhaltenstest nicht dem Verhalten entspricht, welches in einer normalen häuslichen Situation gezeigt werden würde. Tatsächlich können Tests, die in Tierheimen durchgeführt werden, aufgrund der dortigen Haltungsbedingungen keine häuslichen Situationen simulieren, was sich negativ auf die Inhaltsvalidität auswirkt. So kann die Hektik, das hohe Maß an Unbekannten und der Mangel an Fluchtmöglichkeiten die Bereitschaft für defensiv aggressives Verhalten von Hunden erhöhen (TAYLOR und MILLS, 2006).

Laut VAN DER BORG et al., (2010), TAYLOR und MILLS (2006) und MARDER et al. (2013) kann es zu vermehrt ängstlichen oder angstaggressiven Reaktionen führen. Demnach kann angstaggressives Verhalten, zum Beispiel gegenüber fremden Personen oder der verwendeten künstlichen Hand in einem Verhaltenstest gezeigt werden, welches dann als anders motivierte (offensive) Aggression fehlinterpretiert wird und zu falsch-positiven Ergebnissen führt (KROLL et al., 2004; VAN DER BORG et al., 2010). So waren in einer Studie von DEMEESTER (2008) 31% der beißenden Hunde von Angst motiviert, was sich durch eine niedrige Schwanzposition auszeichnete. Auch die Studie von MARDER et al. (2013) zeigte, dass das Verhalten von Hunden in einem Tierheim sehr verschieden sein kann von dem Verhalten in einem sicheren und vorhersagbaren Umfeld eines Zuhauses. Dieses Ergebnis korreliert mit den Ergebnissen von VALSECCHI et al. (2011), wo aggressives Verhalten von Hunden gegenüber unbekannten Personen in einem Tierheim mit aggressivem Verhalten gegenüber vertrauten Personen in einem Zuhause verglichen wurde. Diese Ergebnisse haben Auswirkungen auf die Interpretation von allen provokativen Verhaltenstests, vor allem aber in Tierheimen, in denen Hunde gestresst sind und keine enge Bindung an Betreuungspersonen haben.

In diesen Fällen wäre die Validität des Verhaltenstests anzuzweifeln (WEISS und GREENBERG, 1997).

Darum ist es besonders wichtig, dass Verhaltenstests in Tierheimen so kurz und stressarm wie möglich gehalten werden, um eine Beeinträchtigung der Testergebnisse zu verhindern und um eine gute Inhaltsvalidität zu gewährleisten (BENNETT et al., 2012).

3 Material und Methoden

3.1 Überblick

Ziel dieser Arbeit war es, einen Test zur Erkennung von futterverteidigendem Verhalten von Hunden gegenüber sich annähernden Menschen zu entwickeln und zu validieren. Dieser Test auf Futteraggression wurde in Anlehnung an STERNBERGs (2002) Assess-A-Pet (BOLLEN und HOROWITZ, 2008) und dem MATCH-UP vom Center for Shelter Dogs (DOWLING-GUYER et al., 2011) entwickelt. Der Test auf Futteraggression sollte die Reaktionen von Hunden eines Tierheims in Anwesenheit von Futter auf eine sich annähernde Testperson prüfen und bewerten. Dabei wurden drei Testsituationen abgefragt, im Folgenden als "Subtests" bezeichnet (Subtest 1 – leere Futterschüssel, Subtest 2 – normales Futter, Subtest 3 – hochwertiger Kauartikel). In 13 (Subtest 1 und 2) bzw. zwölf (Subtest 3) festgelegten Testsituationen näherte sich eine Testperson auf einem eingezäunten Testgelände schrittweise den angebundenen Hunden an und interagierte mit ihnen und der Futterquelle.

Die Reaktionen der Hunde wurden mit einer Videokamera erfasst und anschließend ausgewertet. Mit den gewonnenen Daten war daraufhin eine Überprüfung der Tests im Hinblick auf ihre Reliabilität und Validität möglich.

So konnte mithilfe der Videoaufnahmen der Tests die Reproduzierbarkeit der Beurteilung von einer auswertenden Person (Intra-rater-Reliabilität), bzw. drei auswertenden Personen (Inter-rater-Reliabilität) untersucht werden. Außerdem wurde die Reproduzierbarkeit der Ergebnisse (Test-Retest-Reliabilität) überprüft, indem der Test nach acht Wochen mit denselben Hunden wiederholt wurde und die Ergebnisse mit denen des ersten Tests verglichen wurden.

Um sicherzustellen, dass der Test lediglich Futteraggression misst (Inhaltsvalidität), beinhaltete er eine kurze Allgemeinuntersuchung, um medizinische Ursachen für Aggressivität auszuschließen sowie einen vorangehenden Prätest. Dieser Prätest wurde vor dem Test auf Futteraggression durchgeführt und sollte jene Hunde vom Test ausschließen, die schon ohne Anwesenheit von Futter agonistisches Verhalten gegenüber der Testperson, der verwendeten künstlichen Hand oder aufgrund der Testbedingungen zeigten. Durch einen direkten Vergleich des Verhaltens der Hunde in diesem Prätest gegenüber der Testperson mit und ohne künstlicher Hand konnte

außerdem auf die Unterscheidungsfähigkeit der Hunde von realer und künstlicher Hand geschlossen werden.
Als letztes wurde die Kriteriumsvalidität untersucht. Dazu wurden die Ergebnisse des Tests mit den Angaben der Tierheimmitarbeiter in einem Datenerhebungsbogen zum Verhalten der einzelnen Hunde verglichen. Außerdem wurde nach weiteren sieben Wochen ein weniger standardisierter Referenztest von der Tierheimleiterin durchgeführt, welcher als Goldstandard den wahren Status der Hunde in Bezug auf Futteraggression ermitteln sollte. Dieser wurde so konstruiert, dass er vergleichbar mit dem Test auf Futteraggression und gleichzeitig alltäglicher und weniger standardisiert war. Zuletzt wurden die Ergebnisse des Tests mit denen des von HSU und SERPELL (2003) entwickelten und evaluierten C-BARQ-Fragebogen auf Korrelationen überprüft.

3.2 Tiere

Mit dem in dieser Studie entwickelten Test auf Futteraggression wurden 17 Hunde der Rasse Bullterrier aus dem Tierheim "Bullterrier in Not" in Herford getestet. Dieses Tierheim betreut ausschließlich Hunde der Rassen Bullterrier, American Staffordshire Terrier, Staffordshire Bullterrier und Hunde vom Pit-Bull-Typ sowie deren Mischlinge. Die Beschränkung auf Hunde einer einzigen Rasse sollte Varianzen im Verhalten aufgrund unterschiedlicher Rassedisposition minimieren. Mit der Beschränkung auf nur ein Tierheim wurden außerdem Verhaltensabweichungen aufgrund unterschiedlicher Haltungsbedingungen ausgeschlossen. So konnte sichergestellt werden, dass die getestete Hundepopulation möglichst einheitlich war.
Für den Test auf Futteraggression wurden nur adulte, gesunde und an das Tierheim gewöhnte Hunde ausgewählt, die keine Verhaltensauffälligkeiten im Hinblick auf gesteigerte Aggressivität oder Furcht gegenüber Menschen oder andere konfliktbezogene Verhaltensstörungen aufwiesen.
So wurden nur Hunde getestet, die seit mindestens zwei Tagen aus der Quarantäne- bzw. Krankenstation entlassen und die mindestens ein Jahr oder älter waren. Von allen infrage kommenden Hunden wurden Daten bezüglich ihrer Demographik, Haltungsbedingungen, Gesundheitsstatus und auffälligen Verhaltensweisen beim Vorbesitzer und im Tierheim mithilfe eines Datenerhebungsbogens (Anhang 9.1) ermittelt, welcher von den Tierheimmitarbeitern ausgefüllt wurde. Aufgrund dieser Angaben wurden jene Tiere ermittelt und ausgeschlossen, welche in der

Vergangenheit aggressives Verhalten im Sinne von Erstarren, Drohfixieren, Knurren, Zähneblecken, Schnappen oder Beißen nach den Definitionen des Ethogramms für den Test auf Futteraggression (Anhang 9.2) gegenüber fremden oder bekannten Personen in einem anderen Kontext als Futterverteidigung gezeigt hatten. Ebenso wurden Hunde ausgeschlossen, die vorberichtlich mit Furcht im Sinne von geduckter Körperhaltung oder eingeklemmter Rute gegenüber sich freundlich annähernden Personen reagiert hatten. So sollte sichergestellt werden, dass keine anderen Motivationen außer die Verteidigung von Futter für agonistisches Verhalten der Hunde im Test verantwortlich waren. Bei positiver Beantwortung der Fragen nach anderen Aggressionen außer Futteraggression und Furcht vor Personen im Datenerhebungsbogen wurden die entsprechenden Situationen näher hinterfragt, um die genaueren Umstände und Auslöser von aggressivem oder ängstlichem Verhalten der Hunde zu klären. Konnte die Testperson hierbei eindeutig als Auslöser für Aggression oder Furcht ausgeschlossen werden, weil die Hunde beispielsweise ausschließlich aggressiv oder ängstlich auf Männer oder Kinder reagierten, wurden diese Hunde trotzdem zum Test zugelassen, damit Tiere nicht unnötig aus dem Test ausschieden. Des Weiteren wurden nur Hunde zum Test zugelassen, die weder unter bekannten schmerzhaften oder das Verhalten beeinflussenden Erkrankungen, noch unter konfliktbezogenen Verhaltensstörungen wie Stereotypien litten, was ebenfalls im Datenerhebungsbogen (Anhang 9.1) erfragt wurde. In diesen Fällen wurde individuell entschieden, ob eine Teilnahme am Test zumutbar oder das Risiko von weiteren physischen und psychischen Leiden oder Schäden für die Hunde zu groß war.
Nach Auswertung des Datenerhebungsbogens wurden schließlich 17 von insgesamt 34 im Tierheim befindlichen Hunde (50 %) zum Test zugelassen. Es wurden 13 weibliche und vier männliche Hunde getestet, von denen zwei weibliche Hunde und ein männlicher Hund nicht kastriert waren. Im Durchschnitt hatten sie 1,94 Vorbesitzer, waren 8,24 Jahre alt und seit 5,18 Jahren in diesem Tierheim.

3.3 Testperson und Gutachter

Um eine gute Standardisierung zu erreichen, wurden die Tests bei allen Hunden von derselben Testperson durchgeführt. Diese war in der Zeit der Studie approbierte Tierärztin, Doktorandin und Mitarbeiterin des Instituts für Tierschutz und Verhalten an

der Stiftung Tierärztliche Hochschule Hannover. Sie war allen Hunden unbekannt und trug an allen Testtagen dieselbe Kleidung.

Als Gutachter wurden in dieser Studie drei Personen eingesetzt, welche die Videoaufnahmen der einzelnen Tests auswerteten. Diese hatten ein unterschiedliches Maß an Vorerfahrung im Bereich Hundeverhalten und sollten unterschiedlich ausgebildete Personen simulieren, welche möglicherweise später den Test auf Futteraggression bei Hunden anwenden und auswerten könnten. Diese werteten die Videoaufnahmen der Tests in zufälliger Reihenfolge und unabhängig voneinander aus.

Die erste Gutachterin (1) war die Testperson und Doktorandin, welche die Tests durchführte. Um Voreingenommenheit soweit wie möglich zu vermeiden und um sicher zu stellen, dass sie sich möglichst nicht an das Verhalten der einzelnen Hunde im Test erinnerte, wertete diese die Videoaufnahmen frühestens zwei Wochen nach Durchführung des Tests aus.

Der zweite Gutachter (2) war ein Hundehalter ohne weitere Ausbildung im Bereich Hundeverhalten und war nicht mit den Hunden vertraut.

Die dritte Gutachterin (3) war Doktorin der Veterinärmedizin mit der Zusatzbezeichnung Verhaltenstherapie, Mitarbeiterin des Instituts für Tierschutz und Verhalten der Stiftung Tierärztlichen Hochschule Hannover und Inhaberin einer Verhaltenspraxis und Hundeschule in Hannover, welche ebenfalls nicht vertraut mit den Hunden des Tierheims war.

Alle drei Gutachter verwendeten für die Ermittlung der Inter- und Intra-rater-Reliabilität bei der Auswertung dieselben Videos und Auswertungstabellen (Anhang 9.3), dieselbe Bewertungssystematik (Kapitel 3.8.1) sowie dieselben Ethogramme für die Subtests (Anhang 9.2).

3.4 Testgelände

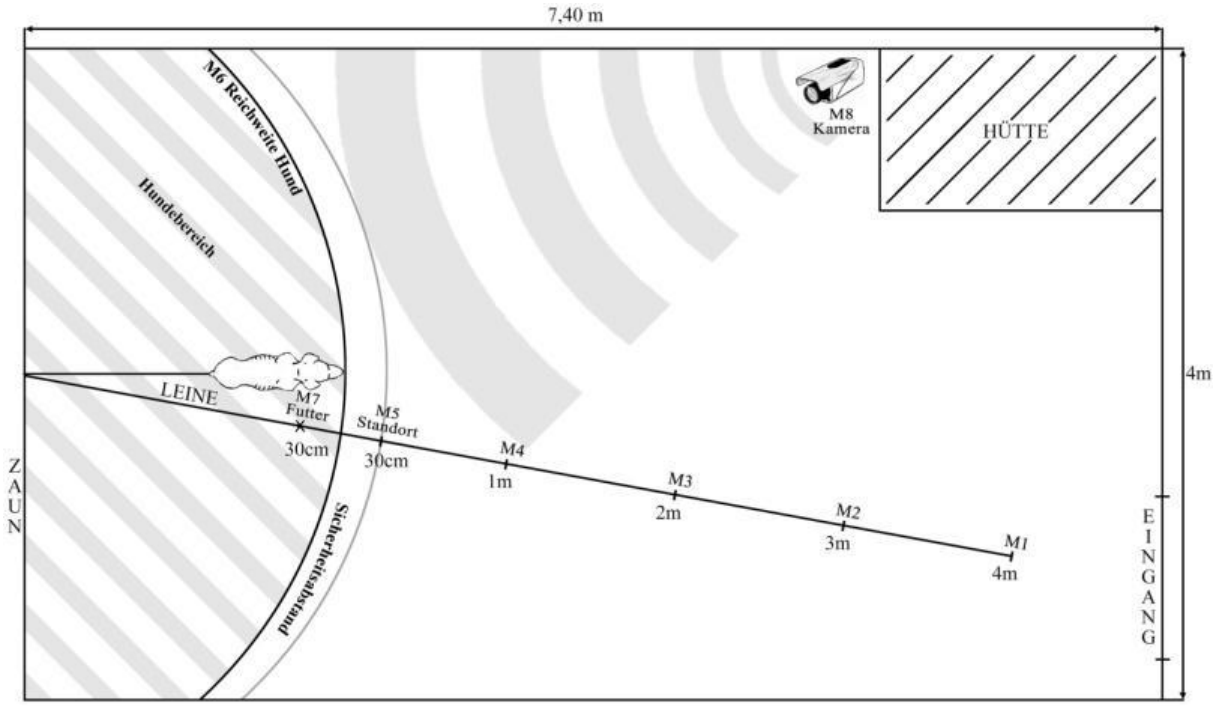

Abbildung 1: schematische Darstellung des Testgeländes

Alle Tests wurden aus Gründen der Standardisierung auf dem in Abb. 1 schematisch dargestellten Testgelände durchgeführt. Dieses befand sich auf dem Grundstück des Tierheims und war ein 7,4 x 4 Meter großes Auslaufgelände. Die Anbindevorrichtung bestand aus einer stabilen Eisenkette, die in Kopfhöhe der Hunde um einen fest im Boden verankerten Holzbalken in der Mitte eines vier Meter breiten Lamellenzauns gelegt war. Dort konnte eine zwei Meter lange Nylonleine befestigt werden, an der die Hunde während des Tests angeleint wurden.

Für den Testablauf wurden acht Markierungen mit schwarzem Farbspray auf dem Boden des Testgeländes angebracht. Markierung 6 stellte die absolute Reichweite der Hunde an der gespannten 2 Meter langen Leine mit zusätzlichen 30 cm für den Kopfbereich der Hunde dar und war in Form eines Halbkreises (r = 2,30 m) um die Anbindevorrichtung auf dem Boden aufgebracht. In diesem Bereich konnten sich die Hunde frei bewegen, weshalb er als "Hundebereich" bezeichnet wurde. Die Markierungen 1 bis 4 beschrieben die Entfernung in gerader Linie zum Halbkreis des Hundebereichs (siehe Abb. 1: Markierung 1: 4 Meter, Markierung 2: 3 Meter, Markierung 3: 2 Meter und Markierung 4: 1 Meter zur maximalen Reichweite der Hunde an der Leine). Zwischen Markierung 4 und der Halbkreislinie des

Hundebereichs (Markierung 6) war Markierung 5 angebracht, die den Standort der Testperson für die zweite Hälfte des Tests beschrieb, in der mit den Hunden interagiert werden sollte. Der Standort lag 30 cm von der maximalen Reichweite der Hunde an der Leine entfernt und war aus Sicherheitsgründen notwendig, um während des Tests versehentliches Hineinbeugen der Testperson in den Bereich der Hunde zu verhindern und um sicherzustellen, dass die Testperson während der einzelnen Tests zu jeder Zeit außerhalb des Hundebereichs blieb und ausschließlich mit der künstlichen Hand die Halbkreislinie von Markierung 6 überquerte. Markierung 7 beschrieb die Position der Futterschüssel und des abgelegten Rinderpansenstreifens und befand sich ebenfalls 30 cm von der maximalen Reichweite der Hunde an der Leine entfernt, jedoch innerhalb des Hundebereichs. Damit war sie vom Standort der Testperson (Markierung 5) genau 60 cm entfernt, sodass das Futter dort gut mit der 70 cm langen künstlichen Hand erreichbar war.
Markierung 8 bezeichnete den Standort der Kamera, welcher an der rechten Zaunseite in einem Abstand von ca. 3 Metern vom Hundebereich (Markierung 6) lag. Die Kamera zeichnete den kompletten Hundebereich sowie den hundenahen Bereich der Testperson ab Markierung 4 von schräg vorne und aus einer Höhe von ca. 1 Meter auf.

3.5 Hilfsmittel

Um Verletzungen der Testperson durch die Hunde zu vermeiden, wurde für den Test eine 70 cm lange künstliche Hand verwendet. Diese bestand aus einer lebensecht aussehenden Gummihand, die an einem Holzstiel befestigt war, wobei ein mit Watte ausgestopfter Hemdärmel den Unterarm simulierte. Es wurde die in den USA gebräuchliche "Assess-A-Hand" von STERNBERG (2002) verwendet, da sie ein sehr realitätsnahes Aussehen hat (Abb. 2). Die künstliche Hand wurde, wie in Abb. 3 dargestellt, mit dem Holzstiel in den Ärmel der Testperson geschoben und das Holzende im Schulterbereich festgehalten, sodass die künstliche Hand frei beweglich und der Arm der Testperson unter der Jacke angewinkelt war. Durch diese Handhabung wurde gewährleistet, dass die Hunde nicht unnötig durch einen stockähnlichen Gegenstand als verlängerten Arm verunsichert wurden, die künstliche Hand möglichst lebensecht verwendet werden konnte und der Arm der Testperson optimal vor Verletzungen geschützt wurde.

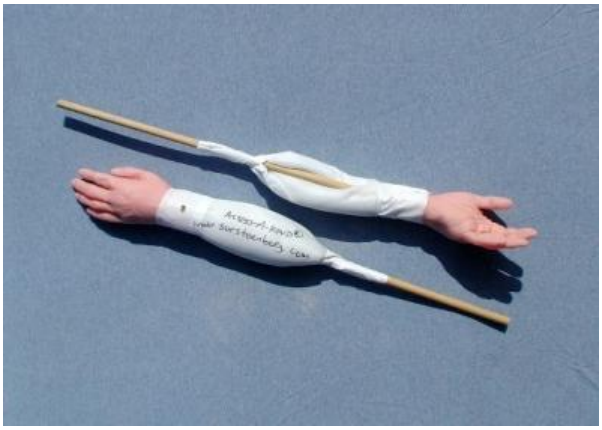

Abbildung 2: "Assess-A-Hand" nach STERNBERG (2002)

Abbildung 3: Handhabung der künstlichen Hand

Für das Futter wurden die gewohnten Futterschüsseln der einzelnen Hunde verwendet, um Verhaltensabweichungen aufgrund von fremdem Aussehen, Material und Geruch zu vermeiden. Aus demselben Grund und um eine höhere Wertigkeit des Futters für die Hunde zu vermeiden, wurde das gewohnte Futter der Hunde angeboten. Es wurden Trockenfutter und Nassfutter sowie ggf. spezielle Allergiker- oder Diätfuttermittel verschiedener Marken und Sorten verwendet. Für jeden Hund wurde eine Tagesportion des Futters angereicht, sodass der Hund während der gesamten Dauer des jeweiligen Tests fressen konnte. Außerdem wurden handelsübliche Streifen von getrocknetem Rinderpansen verwendet, welche für die getesteten Hunde eine hohe Wertigkeit aufgrund besserer Schmackhaftigkeit und seltener Gabe hatten. Die Streifen waren mit mehr als 10 cm lang genug, um die Gefahr des Verschluckens auszuschließen und die Hunde während der gesamten Dauer des jeweiligen Tests mit Kauen zu beschäftigen. War der Rinderpansen für Hunde nicht attraktiv genug oder durften sie diesen aufgrund von Allergien oder Unverträglichkeiten nicht fressen, wurden ersatzweise getrocknete Entenfleischstreifen bzw. Straußenfleischstücke verwendet.

Zur Videoaufzeichnung der Tests wurde ein Medion X47032 Life Camcorder als Kamera verwendet, welcher die Videos auf einer SDHC-Card im MOD-Format speicherte. Um die Kamera sicher aufzustellen, wurde ein Stativ verwendet, welches

auf eine optimale Höhe von ca. einem Meter eingestellt und sicher am Metallzaun befestigt, in einer Entfernung von ca. drei Metern vom Hundebereich aufgestellt wurde.

3.6 Testvorbereitung

Während des Tests musste sichergestellt werden, dass es zu keinen Ablenkungen der Hunde durch andere Personen, Hunde oder störende Geräusche kommen konnte. Dazu wurden Hunde aus benachbarten Zwingern umgesetzt und die Tierheimmitarbeiter um Rücksichtnahme während der Tests gebeten.
Die zu testenden Hunde wurden kurz vor ihrer gewohnten Fütterungszeit von der Testperson aus ihren Zwingern geholt und für fünf bis zehn Minuten auf einer dem Tierheim angrenzenden Wiese zum Lösen und Entspannen ausgeführt. Durch diese Vorgehensweise sollte sichergestellt werden, dass die Hunde zum Test hungrig, gelöst und entspannt waren. Anschließend wurden sie auf das Testgelände gebracht, welches sicher verschlossen wurde. Hier bekamen die Hunde drei Minuten Zeit, das Testgelände frei zu erkunden. Nach dieser Erkundungszeit wurden die Hunde von der Testperson an der zwei Meter langen Nylonleine der Anbindevorrichtung festgebunden und die Videoaufnahme gestartet.

3.7 Testablauf

Während der einzelnen Tests wurden die Hunde direkt angesehen, ohne sie bedrohlich anzustarren. Außerdem durfte die Testperson die Hunde während der Tests nicht ansprechen. Aufgrund der Kameraposition und für die Auswertung wurden aber sowohl die Namen der Hunde, die Prätest- und Subtestnummern, als auch die Nummern der jeweiligen Testsituationen laut angesagt.
Kam es vor, dass Hunde sehr unruhig waren oder sich im angebundenen Zustand in der Leine verwickelten, was ein erhöhtes Verletzungsrisiko für diese Tiere darstellte, wurde die Tierheimleiterin als den Hunden vertraute Person gebeten, sich in den Hundebereich zu stellen und hinter den Hunden andeutungsweise die Leine in die Hand zu nehmen. So blieben die Hunde zwar an der Anbindevorrichtung festgebunden, doch die Tierheimleitung konnte nötigenfalls Verletzungen der Hunde durch allzu starken Zug an der Leine, Herumspringen oder Umwickeln von Gliedmaßen verhindern. Ansonsten verhielt sich die Tierheimleitung während des

Tests den Hunden gegenüber passiv, ignorierte jedes Verhalten der Hunde ihr gegenüber und sprach die Hunde nicht an. Von den getesteten 17 Hunden wurden vier Hunde auf diese Weise getestet. Die Tierheimleiterin verblieb dabei sowohl für die Prätests, als auch für alle Subtests im Hundebereich.

3.7.1 Prätest

Vor dem Test auf Futteraggression wurden zwei Prätests durchgeführt, um die Unterscheidungsfähigkeit der Hunde von echter und künstlicher Hand zu untersuchen und um andere als durch Verteidigung von Futter motivierte Aggressionen als Ursache für agonistisches Verhalten im Test auszuschließen. Dazu wurden die Reaktionen der Hunde auf die Annäherung der Testperson und die Berührungen ihrer echten (Prätest 1) bzw. der künstlichen Hand (Prätest 2) erfasst und anschließend auf signifikante Unterschiede untersucht.

Im ersten Prätest näherte sich die Testperson ohne die künstliche Hand den Hunden schrittweise aus vier Metern Entfernung an und streichelte sie wie unten beschrieben an verschiedenen Körperstellen, wenn dies gefahrlos möglich war. Im zweiten Prätest wurde dieses Vorgehen genau wiederholt, mit dem einzigen Unterschied, dass die Testperson nun die künstliche Hand verwendete.

Das Verhalten der Hunde in beiden Prätests wurde dann in einer separaten Auswertungstabelle für den Prätest (Angang 9.5) und mithilfe eines eigenen Ethogramms (Anhang 9.4) erfasst und anschließend die Häufigkeiten der einzelnen Verhaltensweisen miteinander verglichen. Traten signifikante Unterschiede im Verhalten gegenüber der Testperson zwischen den Prätests eines Hundes auf, so konnte davon ausgegangen werden, dass diese Hunde in der Lage waren, die künstliche Hand von der echten Hand der Testperson zu unterscheiden.

Im Folgenden sind die Testsituationen der beiden Prätests beschrieben:

Tabelle 2: Ablauf Prätest 1

Prätest 1 Annäherung und Berührung der Hunde durch die Testperson mit der echten Hand

Testsituation	Beschreibung
1	Die Testperson steht 3 sec lang in einem Abstand von 4 m (Markierung 1) vom Hund entfernt, die echte Hand wird dabei tief an der Seite gehalten.
2	Die Testperson steht 3 sec lang in einem Abstand von 3 m (Markierung 2) vom Hund entfernt, die echte Hand wird dabei tief an der Seite gehalten.
3	Die Testperson steht 3 sec lang in einem Abstand von 2 m (Markierung 3) vom Hund entfernt, die echte Hand wird dabei tief an der Seite gehalten.
4	Die Testperson steht 3 sec lang in einem Abstand von 1 m (Markierung 4) vom Hund entfernt, die echte Hand wird dabei tief an der Seite gehalten.
5	Die Testperson steht 3 sec lang auf Markierung 5, die echte Hand wird dabei tief an der Seite gehalten.
6	Die Testperson hält die echte Hand 3 sec lang dem Hund zum Erkunden entgegen.
7	Die Testperson berührt mit der echten Hand 3 sec lang die Vorderbrust des Hundes, soweit diese erreichbar und es gefahrlos[8] möglich ist.
8	Die Testperson streicht mit der echten Hand 3 sec lang über die Seite des Hundes, soweit diese erreichbar und es gefahrlos möglich ist.
9	Die Testperson streicht mit der echten Hand 3 sec lang über den Rücken des Hundes, soweit dieser erreichbar und es gefahrlos möglich ist.

[8] = der Hund zeigt keine Drohsignale wie Erstarren, Drohfixieren, Knurren, Zähneblecken, Schnappen oder Beißen nach Definition des Ethogramms für den Prätest

10	Die Testperson berührt mit der echten Hand 3 sec lang den Kopf des Hundes zwischen den Ohren, soweit dieser erreichbar und es gefahrlos möglich ist.

Tabelle 3: Ablauf Prätest 2

Prätest 2 Annäherung und Berührung der Hunde durch die Testperson mit der künstlichen Hand

Testsituation	Beschreibung
1	Die Testperson steht 3 sec lang in einem Abstand von 4 m (Markierung 1) vom Hund entfernt, die künstliche Hand wird dabei tief an der Seite gehalten.
2	Die Testperson steht 3 sec lang in einem Abstand von 3 m (Markierung 2) vom Hund entfernt, die künstliche Hand wird dabei tief an der Seite gehalten.
3	Die Testperson steht 3 sec lang in einem Abstand von 2 m (Markierung 3) vom Hund entfernt, die künstliche Hand wird dabei tief an der Seite gehalten.
4	Die Testperson steht 3 sec lang in einem Abstand von 1 m (Markierung 4) vom Hund entfernt, die künstliche Hand wird dabei tief an der Seite gehalten.
5	Die Testperson steht 3 sec lang auf Markierung 5, die künstliche Hand wird dabei tief an der Seite gehalten.
6	Die Testperson hält die künstliche Hand 3 sec lang dem Hund zum Erkunden entgegen.
7	Die Testperson berührt mit der künstlichen Hand 3 sec lang die Vorderbrust des Hundes, soweit diese erreichbar ist.
8	Die Testperson streicht mit der künstlichen Hand 3 sec lang über die Seite des Hundes, soweit diese erreichbar ist.
9	Die Testperson streicht mit der künstlichen Hand 3 sec lang über den Rücken des Hundes, soweit dieser erreichbar ist.

10	Die Testperson berührt mit der künstlichen Hand 3 sec lang den Kopf des Hundes zwischen den Ohren, soweit dieser erreichbar ist.

Zeigten Hunde während einem dieser Prätests gegenüber der Testperson oder der künstlichen Hand aggressives Verhalten im Sinne von Erstarren, Drohfixieren, Knurren, Zähneblecken, Schnappen oder Beißen nach den Definitionen des Ethogramms, wurden sie vom weiteren Verlauf des Tests auf Futteraggression ausgeschlossen. Ein Ausschluss erfolgte ebenfalls, wenn Hunde sich mit eingeklemmter Rute oder geduckter Körperhaltung von der Testperson oder der künstlichen Hand weg bewegten oder sich mit einem dieser Signale außerhalb ihrer Reichweite aufhielten. Diese Verhaltensweisen deuteten auf Aggressivität oder Furcht gegenüber der Testperson oder der künstlichen Hand hin. Durch den Ausschluss dieser Hunde vom weiteren Testverlauf konnte gewährleitet werden, dass agonistisches Verhalten gegenüber der Testperson und der künstlichen Hand im darauf folgenden Test auf Futteraggression lediglich aufgrund der Motivation von Futterverteidigung, nicht aber aufgrund von Furcht oder anderen Arten von Aggressionen gezeigt wurde.
Um aggressive Verhaltensweisen oder Furcht während der Prätests festzustellen, wurden deren Videos gleich nach ihrer Durchführung von der Testperson auf der Kamera angesehen. Konnte sie dabei keine Anzeichen dieser Verhaltensweisen bei den Hunden feststellen, wurden anschließend die unten beschriebenen Subtests des Tests auf Futteraggression unmittelbar nacheinander folgend und in angegebener Reihenfolge durchgeführt.

3.7.2 Test auf Futteraggression

Vor den jeweiligen Subtests des Tests auf Futteraggression wurde die Videoaufnahme der Kamera gestartet und anschließend die Futterschüssel bzw. der Rinderpansenstreifen mithilfe der künstlichen Hand auf Markierung 7 geschoben. Während die Testperson zu Markierung 1 zurückging, hatten die Hunde drei bis vier Sekunden Zeit, die Schüssel zu begutachten oder mit Fressen bzw. Kauen anzufangen.

Hatten Hunde ihre Portionen bereits aufgefressen, bevor das Ende des jeweiligen Subtests erreicht war, wurde ihnen eine neue Portion gereicht und an selber Stelle des Tests fortgefahren.

Kam es weiterhin vor, dass Hunde kein Interesse am Futter oder den Rinderpansenstreifen zeigten, wurden sie bis zu dreimal zum Fressen animiert, indem das Futter erneut vor die Hunde geschoben wurde. War dies nicht erfolgreich, wurde attraktiveres Futter, wie zum Beispiel Dosenfutter in Subtest 2 oder getrocknetes Entenfleisch in Subtest 3, verwendet. Fraßen sie auch nach dem dritten Versuch und mit attraktiverem Futter nicht, wurde der jeweilige Subtest abgebrochen und mit dem nächsten fortgefahren. Aus diesem Grund abgebrochene Subtests wurden einmalig und komplett am nächsten Tag und unter Verwendung von attraktiverem Futter wiederholt.

Im Folgenden sind die Subtests mit ihren einzelnen Testsituationen beschrieben:

Tabelle 4: Ablauf Subtest 1

Subtest 1 Test auf Futteraggression mit leerer Futterschüssel

Testsitaution	Beschreibung
1	Die Testperson steht 3 sec lang in einem Abstand von 4 m (Markierung 1) vom Hund entfernt, die künstliche Hand wird dabei tief an der Seite gehalten.
2	Die Testperson steht 3 sec lang in einem Abstand von 3 m (Markierung 2) vom Hund entfernt, die künstliche Hand wird dabei tief an der Seite gehalten.
3	Die Testperson steht 3 sec lang in einem Abstand von 2 m (Markierung 3) vom Hund entfernt, die künstliche Hand wird dabei tief an der Seite gehalten.
4	Die Testperson steht 3 sec lang in einem Abstand von 1 m (Markierung 4) vom Hund entfernt, die künstliche Hand wird dabei tief an der Seite gehalten.
5	Die Testperson steht 3 sec lang auf Markierung 5, die künstliche Hand wird dabei tief an der Seite gehalten.
6	Die Testperson berührt mit der künstlichen Hand 3 sec lang die Vorderbrust des Hundes, soweit diese erreichbar ist.

7	Die Testperson streicht mit der künstlichen Hand 3 sec lang über die Seite des Hundes, soweit diese erreichbar ist.
8	Die Testperson streicht mit der künstlichen Hand 3 sec lang über den Rücken des Hundes, soweit dieser erreichbar ist.
9	Die Testperson berührt mit der künstlichen Hand 3 sec lang den Kopf des Hundes zwischen den Ohren, soweit dieser erreichbar ist.
10	Die Testperson beschreibt mit der künstlichen Hand einen Halbkreis um die leere Futterschüssel im Abstand von ca. 15 cm.
11	Die Testperson fährt mit der künstlichen Hand den Rand der leeren Futterschüssel ab, wenn möglich, ohne dabei den Hund zu berühren.
12	Die Testperson hält die künstliche Hand für 3 sec in die Futterschüssel, wenn möglich, ohne dabei den Hund zu berühren.
13	Die Testperson zieht mit der künstlichen Hand die Futterschüssel langsam aus der Reichweite des Hundes und nimmt sie in sicherem Abstand auf.

Tabelle 5: Ablauf Subtest 2

Subtest 2 Test auf Futteraggression mit gewohntem Futter

Testsituation	Beschreibung
1	Die Testperson steht 3 sec lang in einem Abstand von 4 m (Markierung 1) vom Hund entfernt, die künstliche Hand wird dabei tief an der Seite gehalten.
2	Die Testperson steht 3 sec lang in einem Abstand von 3 m (Markierung 2) vom Hund entfernt, die künstliche Hand wird dabei tief an der Seite gehalten.
3	Die Testperson steht 3 sec lang in einem Abstand von 2 m (Markierung 3) vom Hund entfernt, die künstliche Hand wird dabei tief an der Seite gehalten.
4	Die Testperson steht 3 sec lang in einem Abstand von 1 m (Markierung 4) vom Hund entfernt, die künstliche Hand wird dabei tief an der Seite gehalten.

5	Die Testperson steht 3 sec lang auf Markierung 5, die künstliche Hand wird dabei tief an der Seite gehalten
6	Die Testperson berührt mit der künstlichen Hand 3 sec lang die Vorderbrust des fressenden Hundes, soweit diese erreichbar ist.
7	Die Testperson streicht mit der künstlichen Hand 3 sec lang über die Seite des fressenden Hundes, soweit diese erreichbar ist.
8	Die Testperson streicht mit der künstlichen Hand 3 sec lang über den Rücken des fressenden Hundes, soweit dieser erreichbar ist.
9	Die Testperson berührt mit der künstlichen Hand 3 sec lang den Kopf zwischen den Ohren des fressenden Hundes, soweit dieser erreichbar ist.
10	Die Testperson beschreibt mit der künstlichen Hand einen Halbkreis um die Futterschüssel im Abstand von ca. 15 cm.
11	Die Testperson fährt mit der künstlichen Hand den erreichbaren Rand der Futterschüssel ab, wenn möglich, ohne dabei den Hund zu berühren.
12	Die Testperson hält die künstliche Hand für 3 sec in die Futterschüssel, wenn möglich, ohne dabei den Hund zu berühren.
13	Die Testperson zieht mit der künstlichen Hand die Futterschüssel langsam aus der Reichweite des Hundes und nimmt sie in sicherem Abstand auf.

Tabelle 6: Ablauf Subtest 3

Subtest 3 Test auf Futteraggression mit einem Streifen getrocknetem Rinder-pansen

Testsituation	Beschreibung
1	Die Testperson steht 3 sec lang in einem Abstand von 4 m (Markierung 1) vom Hund entfernt, die künstliche Hand wird dabei tief an der Seite gehalten.
2	Die Testperson steht 3 sec lang in einem Abstand von 3 m (Markierung 2) vom Hund entfernt, die künstliche Hand wird dabei tief an der Seite gehalten.

3	Die Testperson steht 3 sec lang in einem Abstand von 2 m (Markierung 3) vom Hund entfernt, die künstliche Hand wird dabei tief an der Seite gehalten.
4	Die Testperson steht 3 sec lang in einem Abstand von 1 m (Markierung 4) vom Hund entfernt, die künstliche Hand wird dabei tief an der Seite gehalten.
5	Die Testperson steht 3 sec lang auf Markierung 5, die künstliche Hand wird dabei tief an der Seite gehalten.
6	Die Testperson berührt mit der künstlichen Hand 3 sec lang die Vorderbrust des kauenden Hundes, soweit diese mit der künstlichen Hand erreichbar ist.
7	Die Testperson streicht mit der künstlichen Hand 3 sec lang über die Seite des kauenden Hundes, soweit diese mit der künstlichen Hand erreichbar ist.
8	Die Testperson streicht mit der künstlichen Hand 3 sec lang über den Rücken des kauenden Hundes, soweit dieser mit der künstlichen Hand erreichbar ist.
9	Die Testperson berührt mit der künstlichen Hand 3 sec lang den Kopf zwischen den Ohren des kauenden Hundes, soweit dieser mit der künstlichen Hand erreichbar ist.
10	Die Testperson beschreibt mit der künstlichen Hand einen Halbkreis um den Pansenstreifen im Abstand von ca. 15 cm.
11	Die Testperson berührt mir der künstlichen Hand 3 sec lang den Pansenstreifen, wenn möglich, ohne dabei den Hund zu berühren.
12	Wenn möglich, zieht die Testperson mit der künstlichen Hand den Pansenstreifen langsam aus der Reichweite des Hundes und nimmt ihn in sicherem Abstand auf.

Nach Ende von Subtest 3 wurde die Videoaufnahme gestoppt und die Hunde von der Testperson kurz allgemeinuntersucht. Die Untersuchung aller getesteten Hunde wurde erst nach dem Test von der Doktorandin durchgeführt, um bei den Hunden keine Furcht vor der Testperson zu provozieren und sie nicht vor Testbeginn in Stress zu versetzen. Wurden bei Hunden akute Schmerzzustände, Entzündungen oder Unwohlsein festgestellt, wurde der Test des jeweiligen Hundes nicht gewertet, da auch

diese Umstände zu einer gesteigerten Aggressivität und Reaktionsbereitschaft führen könnten. In diesem Fall wurden die zuständigen Tierpfleger über den Gesundheitsstatus der Hunde informiert. Nach der Untersuchung wurden die Hunde von der Anbindevorrichtung losgemacht und erneut ca. fünf Minuten zur Reduzierung von Stress auf dem angrenzenden Gelände ausgeführt. Danach wurden sie in ihre Zwinger zurück gebracht, wo sie das verbliebene Futter ungestört verzehren konnten. Sowohl die beiden Prätests als auch jeder der drei Subtests dauerte ca. 1 bis 1,5 Minuten, wodurch insgesamt eine Testdauer von ca. 5 bis 8 Minuten pro Hund erreicht wurde. Zwischen den einzelnen Prä- und Subtests wurde die Videoaufzeichnung angehalten und der nächste Prä- oder Subtest vorbereitet, was ca. 2 Minuten in Anspruch nahm. So erreichte der Test inklusive Prätests und ohne Ausführ- und Erkundungszeit eine Gesamtdauer von ca. 7 bis 10 Minuten. Inklusive vorherigem Ausführen und Erkunden sowie der nachstehenden Allgemeinuntersuchung und dem erneutem Ausführen war die Testperson ca. 25 bis 30 Minuten je Hund beschäftigt. Um alle 17 Hunde zu testen, wurde ein Zeitraum von drei Tagen benötigt, wobei am ersten Tag acht, am zweiten Tag vier und am dritten Tag fünf Hunde getestet werden konnten.

3.7.3 Retest

Acht Wochen nach der ersten Durchführung wurde der Test auf Futteraggression von derselben Testperson und mit denselben Hunden wie oben beschrieben wiederholt, um die Test-Retest-Reliabilität des Tests zu bestimmen.

Es konnten alle 17 Hunde nachgetestet werden. Der Ablauf des Re-Tests entsprach dem des ersten Testdurchgangs. Besonderheiten in der Durchführung, wie das Anbieten von attraktiverem Futter oder die Anwesenheit der Tierheimleiterin aufgrund der Verletzungsgefahr bei vier einzelnen Tieren, wurde ebenfalls auf gleiche Weise wiederholt, um Abweichungen des Verhaltens der Hunde aufgrund anderer Testbedingungen auszuschließen.

Durch den Vergleich des Verhaltens der Hunde im ersten und zweiten Testdurchlauf konnte die Test-Retest-Reliabiltät des Tests statistisch ermittelt werden.

3.7.4 Referenztest

Nach sieben weiteren Wochen wurde mit 16 der 17 Hunde ein Referenztest zur Bestimmung der Kriteriumsvalidität durchgeführt. Ein Hund musste in diesen sieben Wochen aufgrund von schweren Erkrankungen eingeschläfert werden, sodass dieser den Referenztest nicht absolvieren konnte und seine Daten für den Vergleich mit den Ergebnissen des Tests auf Futteraggression fehlten.

Der Referenztest sollte als Goldstandard den wahren Status der Hunde in Bezug auf Futteraggression ermitteln und wurde so konstruiert, dass er sowohl vergleichbar mit den Subtests des Tests auf Futteraggression, als auch mit einer alltäglichen und weniger standardisierten Situation war. Er wurde von der Tierheimleiterin als den Hunden vertraute Person durchgeführt und von der Doktorandin auf Video aufgezeichnet. Auf ein Testgelände wurde verzichtet und die Hunde stattdessen in ihren jeweiligen Zwingern an gewohnter Stelle und zu gewohnter Zeit gefüttert, um der alltäglichen Fütterungssituation so nahe wie möglich zu kommen.

Im Referenztest wurden dem Test auf Futteraggression entsprechende Teile simuliert, welche den Subtests des Tests auf Futteraggression entsprachen. Innerhalb dieser Teile wurden ebenso die Testsituationen der einzelnen Subtests simuliert, welche das Annähern und Streicheln der Hunde sowie die Interaktionen mit der Futterschüssel bzw. dem Futter beinhalteten (Tabellen 7 bis 9).

Da dieser Referenztest in den jeweiligen Zwingern der Hunde durchgeführt wurde und diese nicht alle groß genug für eine Annäherung aus vier Metern Entfernung waren, musste die Annäherungsphase, welche den Testsituationen 1 bis 5 im Test auf Futteraggression entsprach, gekürzt werden. Die Annäherung wurde deswegen auf nur eine Testsituation des Referenztests beschränkt, und zwar auf den Weg von der Zwingertür bis zum Fütterungsort.

Da keine künstliche Hand benutzt wurde, war die Tierheimleitung dazu angehalten, stets vorsichtig zu sein und ggf. den Test abzubrechen.

Gefilmt wurde der Referenztest von der Doktorandin, welche darauf achtete, dass die Beleuchtung der Zwinger ausreichend und die Reaktionen der Hunde möglichst gut im Bild der Kamera zu sehen waren.

Das Futter wurde, wie im Tierheim üblich, zur gewohnten Fütterungszeit der Hunde außerhalb der Zwinger zubereitet. Die Doktorandin betrat währenddessen die Zwinger, damit sich die Hunde an ihre Anwesenheit gewöhnten. Sie stellte sich so auf, dass eine Videoaufnahme gut möglich war und die Hunde möglichst wenig abgelenkt

wurden. Anschließend betrat die Tierheimleitung die Zwinger, schloss deren Türen und stellte die Futterschüsseln bzw. das entsprechende Futter bereit. Während die Hunde zu fressen begannen, ging die Tierheimleitung bis zur Zwingertür zurück und die Doktorandin startete die Videoaufnahme. Anschließend wurden die jeweiligen Tests in unten angegebener Reihenfolge durchgeführt.

Während der Tests verhielt sich die Doktorandin den Hunden gegenüber passiv, während die Tierheimleiterin den Hunden möglichst gewohntes Verhalten entgegen brachte. Nach jedem Teil des Referenztests verließ sie den Zwinger und bereitete den nächsten Teil vor, um im Anschluss damit fortzufahren.

Im Folgenden sind die Situationen des Referenztests beschrieben:

Tabelle 7: Ablauf Referenztest Teil 1

Teil 1 leerer Futterschüssel

Testsituation	Beschreibung
1	Die Tierheimleitung nähert sich dem Hund.
2	Die Tierheimleitung streichelt den Hund kurz an der Vorderbrust.
3	Die Tierheimleitung streichelt den Hund kurz an einer Seite.
4	Die Tierheimleitung streichelt dem Hund kurz über den Rücken.
5	Die Tierheimleitung streichelt dem Hund kurz über den Kopf.
6	Die Tierheimleitung umschreibt mit der Hand im Abstand von 15 cm einen Halbkreis um die Futterschüssel.
7	Die Tierheimleitung fährt mit der Hand den Rand der Futterschüssel ab.
8	Die Tierheimleitung hält die Hand in die Futterschüssel.
9	Die Tierheimleitung zieht die Futterschüssel langsam weg und nimmt sie auf.

Tabelle 8: Ablauf Referenztest Teil 2

Teil 2 normales Futter in der Schüssel

Testsituation	Beschreibung
1	Die Tierheimleitung nähert sich dem Hund.

2	Die Tierheimleitung streichelt den Hund kurz an der Vorderbrust.
3	Die Tierheimleitung streichelt den Hund kurz an einer Seite.
4	Die Tierheimleitung streichelt dem Hund kurz über den Rücken.
5	Die Tierheimleitung streichelt dem Hund kurz über den Kopf.
6	Die Tierheimleitung umschreibt mit der Hand im Abstand von 15 cm einen Halbkreis um die Futterschüssel.
7	Die Tierheimleitung fährt mit der Hand den Rand der Futterschüssel ab.
8	Die Tierheimleitung hält die Hand in die Futterschüssel.
9	Die Tierheimleitung zieht die Futterschüssel langsam weg und nimmt sie auf.

Tabelle 9: Ablauf Referenztest Teil 3

Teil 3 getrockneter Rinderpansenstreifen

Testsituation	Beschreibung
1	Die Tierheimleitung nähert sich dem Hund.
2	Die Tierheimleitung streichelt den Hund kurz an der Vorderbrust.
3	Die Tierheimleitung streichelt den Hund kurz an einer Seite.
4	Die Tierheimleitung streichelt dem Hund kurz über den Rücken.
5	Die Tierheimleitung streichelt dem Hund kurz über den Kopf.
6	Die Tierheimleitung umschreibt mit der Hand im Abstand von 15 cm einen Halbkreis um den Pansenstreifen.
7	Die Tierheimleitung berührt mit der Hand den Rand des Pansenstreifens.
8	Die Tierheimleitung versucht, den Pansenstreifen wegzunehmen.

Die Referenztests nahmen pro Hund ca. 10 bis 15 Minuten in Anspruch und wurde an zwei Tagen durchgeführt, wobei am ersten Tag zwölf und am zweiten Tag vier Hunde getestet wurden.

3.8 Datenaufnahme und -auswertung

Alle Tests, sowohl die beiden Prätests und die drei Subtests, als auch die Re- und die Referenztests von allen Hunden wurden mit Hilfe einer Digitalkamera (Medion X47032 Life Camcorder) auf Video aufgezeichnet. Die Studiennummern der Hunde von 1 bis 17 wurden auf den Datenerhebungsbögen zufällig vergeben.

3.8.1 Auswertung des Prätests

Die Prätests wurden von der Doktorandin zwei Wochen nach der Aufnahme ausgewertet. Hierbei waren Wiederholungen, Standbilder und Veränderungen der Abspielgeschwindigkeit erlaubt, um möglichst alle Signale der Hunde zu erfassen.
Die beobachteten Signale der Hunde wurden gemäß eines eigenen Ethogramms für die Prätests (Anhang 9.4) in eigens dafür vorgesehenen Auswertungstabellen für die Prätests (Anhang 9.5) eingetragen. Für jede Testsituation wurden alle beobachteten soziopositiven und agonistischen Verhaltensweisen sowie die Stresssignale als beobachtet bzw. nicht beobachtet notiert.
Um das Verhalten der Hunde in beiden Prätests zu vergleichen, wurden die Häufigkeiten der einzelnen Verhaltensweisen in beiden Tests ausgezählt und in die Kategorien "soziopositive Verhaltensweisen", "agonistische Vehaltensweisen" und "Stresssignale" eingeteilt. Die Häufigkeiten der Verhaltensweisen dieser drei Kategorien wurden anschließend zwischen beiden Prätests auf signifikante Unterschiede überprüft. Mit diesem Vergleich konnte eine Aussage darüber gemacht werden, ob sich Hunde im zweiten Prätest mit der künstlichen Hand signifikant anders verhielten als in Prätest 1 ohne künstliche Hand und ob Hunde ein unterschiedliches Maß an Stress in beiden Tests empfanden.

3.8.2 Auswertung des Tests auf Futteraggression

Die Auswertung des Tests auf Futteraggression wurde von drei Gutachtern durchgeführt. Die Gutachter werteten die Videos unabhängig voneinander aus, wobei die Doktorandin erneut eine Frist von mindestens zwei Wochen nach Videoaufnahme verstreichen lassen musste.

Bei der Auswertung der Videos waren auch hier Wiederholungen, Standbilder und Veränderungen der Abspielgeschwindigkeit erlaubt, um möglichst alle Signale der Hunde zu erfassen.
Dokumentiert wurden die beobachteten Signale der Hunde gemäß dem Ethogramm für den Test auf Futteraggression (Anhang 9.2) in dafür vorgesehenen Auswertungstabellen für die Subtests (Anhang 9.3).
Auch hier wurden für jede Testsituation alle beobachteten agonistischen Verhaltensweisen als beobachtet bzw. nicht beobachtet notiert und anschließend ihre Häufigkeiten ausgezählt, während die Stresssignale unabhängig vom Auftreten agonistischen Verhaltens erfasst und ihre Häufigkeit später gesondert ausgewertet wurde. Die soziopositiven Verhaltensweisen wurden für den Test auf Futteraggression nicht erfasst.
Den einzelnen im Ethogramm (Anhang 9.2) beschriebenen agonistischen Verhaltensweisen wurden im weiteren Verlauf in aufsteigender Reihenfolge Punkte ähnlich einem Schulnotensystem zugeordnet, wobei die Punktzahl mit steigender Eskalation des aggressiven Verhaltens zunahm. Verhaltensweisen mit ähnlichem Eskalationsniveau (z.B. knurren und zähneblecken) bekamen dabei dieselben Punktzahlen zugeordnet. Verhaltensweisen, die innerhalb der Agonistik dem Meideverhalten zuzuordnen waren sowie nicht agonistisches Verhalten wurden mit null Punkten bewertet. Auf diese Weise wurde das von den Hunden gezeigte agonistische Verhalten in einen ordinalen Datentyp umgewandelt.
So ergab sich folgende Punktevergabe für den Test auf Futteraggression:

Tabelle 10: Punktvergabe für Verhaltensweisen

Verhaltensweise	**Punktzahl**
Beschleunigter Verzehr	1 Punkt
Festhalten	2 Punkte
Abschirmen	2 Punkte
Wegtragen	2 Punkte
Erstarren	3 Punkte
Drohfixieren	4 Punkte
Knurren	5 Punkte
Zähneblecken	5 Punkte
Schnappen	6 Punkte

Beißen	7 Punkte
Abwenden	0 Punkte
Außerhalb der Reichweite	0 Punkte
Kein agonistisches Verhalten	0 Punkte

Bei der Bewertung der einzelnen Testsituationen mithilfe dieser Punktevergabe wurde lediglich das gezeigte Verhalten bewertet, welches innerhalb der Testsituationen die höchste Punktzahl erhielt. So wurden, wenn in einer Testsituation mehrere Verhaltensweisen gezeigt wurden, deren entsprechenden Punkte nicht summiert, sondern nur die Punktzahl der gezeigten Verhaltensweise in die Zeile "Punkte" für die jeweilige Testsituation übernommen, welche den höchsten Zahlenwert erhielt. Für jeden Subtest wurden dann die Punkte der einzelnen Testsituationen miteinander addiert. Anschließend wurden die Gesamtpunkte der einzelnen Subtests für jeden Hund miteinander addiert, was ein gesamtes Testergebnis für jeden einzelnen Hund für den Test auf Futteraggression ergab.
Zusätzlich wurde eine weitere Einteilung des gezeigten Verhaltens in Kategorien entwickelt. Diese sollte die Auswertung des Tests vereinfachen und die Praktikabilität im Feld erhöhen. Dazu wurden die oben angegebenen und in den Auswertungstabellen erfassten Verhaltensweisen in gröber abgestufte Kategorien als im Ethogramm eingeteilt (Tabelle 11).

Tabelle 11: Kategorieneinteilung von Verhaltensweisen

Kategorie	Interpretation	Verhaltensweisen
Kategorie 0	kein agonistisches Verhalten	kein agonistiches Verhalten
Kategorie 1	Meideverhalten	abwenden, außerhalb Reichweite
Kategorie 2	leichte Futteraggression	beschleunigter Verzehr, festhalten, abschirmen, wegtragen
Kategorie 3	moderate Futteraggression	erstarren, drohfixieren, knurren, zähneblecken
Kategorie 4	starke Futteraggression	schnappen, beißen

Die Kategorien wurden für jeden einzelnen Subtest bestimmt, wobei immer jene Verhaltensweise die Kategorie bestimmte, welche in diesem Subtest vom höchsten Eskalationsniveau war. Zeigte ein Hund während eines Subtests also die

Verhaltensweisen "abwenden", "beschleunigter Verzehr" und "erstarren", so wurde unabhängig ihrer Häufigkeit und ihres Auftretens in einzelnen Testsituationen die Kategorie der höchsten Eskalationsstufe, in diesem Fall "erstarren", mit Kategorie 3 gewertet. So wurde für alle drei Subtests vorgegangen. Für die Gesamtkategorie der einzelnen Hunde wurde die höchste Kategorie innerhalb der drei Subtests gewertet. Hatte ein Hund also im ersten Subtest Kategorie 0, im zweiten Subtest Kategorie 2 und im dritten Subtest Kategorie 3, so wurde als Gesamtkategorie Kategorie 3 gewertet.

3.8.3 Bestimmung der Reliabilität des Tests auf Futteraggression

Sowohl die Punktzahlen und Kategorien der einzelnen Subtests der Hunde, als auch die Gesamtpunktergebnisse und Gesamtkategorien des Tests wurden später zu den statistischen Vergleichen zur Bestimmung der Reliabilität herangezogen. So konnten die Punkt- und Kategorienergebnisse der einzelnen Subtests sowie der Gesamtergebnisse der einzelnen Gutachter jeweils miteinander verglichen werden, um die Inter-rater-Reliabilität zu ermitteln. Ebenso wurden diese Ergebnisse herangezogen, um die Gutachter nach einer erneuten Auswertung derselben Videos mit sich selbst zu vergleichen und so die Intra-rater-Reliabilität der einzelnen Gutachter zu bestimmen. Die Punkte- und Kategorienergebnisse der Subtests sowie die Gesamtergebnisse wurden außerdem mit denen des zweiten Testdurchlaufs mit derselben Hundepopulation (Retest) auf signifikante Unterschiede geprüft, um die Test-Retest-Reliabilität des Tests auf Futteraggression zu ermitteln.

3.8.4 Bestimmung der Kriteriumsvalidität des Tests auf Futteraggression

Die Punkte- und Kategorienergebnisse der einzelnen Subtests sowie die Gesamtergebnisse der ersten Auswertung der Doktorandin (Gutachter 1) wurden ebenfalls herangezogen, um die Validität des Tests zu ermitteln. Hierzu wurden die Kategorienergebnisse der ersten Auswertung der Doktorandin mit den Ergebnissen des Datenerhebungsbogens auf Übereinstimmung geprüft. Außerdem wurden die Punkte- und Kategorienergebnisse der ersten Auswertung der Doktorandin mit denen des Referenztests verglichen, welcher in Ermangelung eines Goldstandards dessen Rolle einnahm. Als dritte Referenz diente der von HSU und SERPELL (2003)

entwickelten C-BARQ-Fragebogen, dessen Ergebnisse für die am Test teilgenommenen Hunde mit den Gesamtpunkte- und Kategorienergebnissen des Tests auf Futteraggression in Korrelation gesetzt wurden.

3.8.4.1 Auswertung des Datenerhebungsbogens

Im Datenerhebungsbogen (Anhang 9.1) wurde die Einschätzung der Tierheimmitarbeiter über das Maß an Futteraggression der einzelnen Hunde erfragt. Hierzu wurde eine Tabelle ausgefüllt, welche das Verhalten der Hunde in den Situationen "leere Futterschüssel", "normales Futter" und "besonderes Futter" ermitteln sollte. Diese wurden dann den einzelnen Subtests des Tests auf Futteraggression gegenübergestellt. Als Auswahlmöglichkeiten konnten die Mitarbeiter innerhalb dieser drei Situationen die einzelnen im Ethogramm für den Test auf Futteraggression (Anhang 9.2) beschriebenen Verhaltensweisen wählen, welche auch von den Gutachtern zur Auswertung des Tests verwendet wurden. So wurde der statistische Vergleich von Datenerhebungsbogen und Testauswertung möglich. Zusätzlich konnten Angaben über die Personen (bekannte Person, fremde Person, Alter und Geschlecht) gemacht werden, gegenüber denen das angegebene Verhalten gezeigt wurde.
Die erfragten Verhaltensweisen des Datenerhebungsbogens wurden anschließend in die im Kapitel 3.8.2 beschriebenen Kategorien eingeteilt und diese statistisch mit den Kategorienergebnissen der ersten Auswertung der Doktorandin vergleichen. Eine Gegenüberstellung mithilfe der Punktevergabe war hierbei nicht möglich, da die Punkteergebnisse durch die Häufigkeit der auftretenden Verhaltensweisen beeinflusst wurden, welche mit dem Datenerhebungsbogen nicht ermittelt werden konnten. Verglichen wurden die Kategorien für jeden Subtest sowie die Gesamtkategorie der einzelnen Hunde. Dabei wurde Subtest 1 der ersten Frage ("leere Futterschüssel), Subtest 2 der zweiten Frage ("normales Futter") und Subtest 3 der dritten Frage ("besonderes Futter") gegenüber gestellt.

3.8.4.2 Auswertung der Referenztests

Die Videoauswertung der Referenztests wurde eine Woche nach Aufzeichnung der Videos und in zufälliger Reihenfolge von der Doktorandin durchgeführt. Es wurde das Ethogramm für die Auswertung der Tests auf Futteraggression (Anhang 9.2)

verwendet, da die Ergebnisse dieser beiden Tests anschließend miteinander verglichen wurden. Aus demselben Grund folgte die Videoauswertung des Referenztests auch denselben Regeln für die Auswertung der Tests auf Futteraggression.

Ebenso wurde dieselbe oben angegebene Bewertungssystematik nach Punkten sowie die oben angegebene Kategorieneinteilung der Verhaltensweisen verwendet wie im Test auf Futteraggression (Kapitel 3.8.2). Diese Punkt- und Kategorienergebnisse des Referenztests wurden anschließend mit den Punkten und Kategorien der einzelnen Subtestergebnisse der Tests auf Futteraggression auf statistisch signifikante Unterschiede geprüft.

Da der Referenztest vier Testsituationen kürzer war als der Test auf Futteraggression, weil die Annäherungsphase (Testsituation 1-5 im Test auf Futteraggression) zu nur einer Testsituation reduziert wurde, wurde für den Vergleich nur jenes Verhalten innerhalb der ersten fünf Testsituationen des Tests auf Futteraggression gewertet, welches die höchste Punktzahl nach der oben angegebenen Bewertungssystematik erhielt. Ebenso wurde bei Auftreten von Stresssignalen innerhalb dieser ersten fünf Testsituation nur ein Stresssignal jeder Art gewertet. Dieses Vorgehen wurde gewählt, um die ersten fünf Testsituationen des Tests auf Futteraggression mit einer Testsituation im Referenztest vergleichbar zu machen, in dem verhindert wird, dass sich Verhaltensweisen aus fünf Testsituationen summierten, was zu ungleich höheren Ergebnissen als in einer Testsituation im Referenztest geführt hätte.

Dieses Vorgehen war bei der Auswertung mithilfe der Kategorieneinteilung unnötig, da die Kategorien unabhängig von der Häufigkeit einzelner Verhaltensweisen und ihres Auftretens in einzelnen Testsituationen bestimmt wurden.

3.8.4.3 Auswertung der C-BARQ-Fragebögen

Die dritte Referenz bildete ein weiterer Fragebogen, der von HSU und SERPELL (2003) entwickelte und evaluierte *Canine Behavioral Assessment and Research Questionnaire* (C-BARQ). Dieser Fragebogen ermittelt in 100 Fragen über das Verhalten von Hunden in verschiedenen alltäglichen Situationen unter anderem die Bereitschaft der Hunde zu "stranger-directed aggression" (Aggression gegenüber Fremden), "owner-directed aggression" (Aggression gegenüber dem Besitzer), "stranger-directed fear" (Furcht vor Fremden) und "touch sensitivity" (Berührungsempfindlichkeit). Diese und andere Kategorien werden in Punkteskalen

von 0,00 bis 4,00 angegeben, ebenso ein Vergleich der einzelnen Hunde mit den Durchschnittswerten (Punkte und Prozent) von allen mit dem C-BARQ getesteten Hunden derselben Rasse bzw. Hunden aller Rassen.
Der C-BARQ wurde in Papierform für jeden getesteten Hund vom Tierheimpersonal ausgefüllt und die Daten anschließend von der Doktorandin auf die Website http://vetapps.vet.upenn.edu/cbarq/ übertragen, wo die Auswertung der einzelnen Hunde automatisch ablief. Die Ergebnisse der Hunde in den vier oben genannten Kategorien wurden anschließend tabellarisch erfasst und die Punktzahlen der vier Kategorien des C-BARQs mittels Spearmans Korrelationskoeffizient mit den Gesamtergebnissen der Punkte- und Kategorienauswertung in Korrelation gesetzt.

3.8.5 Auswertung der Stresssignale

Wie oben erwähnt, wurden die Stresssignale der Hunde bei der Auswertung der Tests auf Futteraggression gesondert aufgenommen. Anschließend wurde ihre Anzahl pro Subtest im Test auf Futteraggression ermittelt, ebenso ihre Gesamtanzahl in den Tests.
Die gezeigten Stresssignale wurden nicht mit Punkten versehen, da bei ihnen keine Abstufung nach Intensität oder Niveau möglich war. Stattdessen wurde ihre Häufigkeit unabhängig ihrer Art für jeden Subtest ausgezählt und die Gesamtanzahl der Stresssignale pro Test und Hund ermittelt. Auf diese Weise konnte eine Aussage getroffen werden, welcher Subtest am stressreichsten für die Hunde war. Außerdem wurde eine Korrelation der Anzahl der Stresssignale mit den Punkt- und Kategorienergebnissen der einzelnen Subtests sowie den Gesamtergebnissen der Hunde vorgenommen, um zu prüfen, ob ein hohes Stresslevel mit einem hohen Testergebnis korreliert. Sowohl für den Vergleich als auch für die Korrelation wurde die erste Auswertung der Doktorandin des ersten Testdurchlaufs des Tests auf Futteraggression herangezogen.

3.9. Statistische Methoden

Die gesammelten Daten der verschiedenen Auswertungstabellen wurden im Excel-Format gespeichert. Für alle statistischen Berechnungen wurde die Statistik-Software

"Statistical Analysis System" (SAS) Enterprise Guide 6.1 verwendet. Die Graphiken wurden mithilfe von Microsoft Office Excel 2013 erstellt.

Im Zuge der deskriptiven Statistik wurden die Häufigkeiten der einzelnen Verhaltensweisen und Stresssignale in den einzelnen Tests auf Futteraggression und den Prätests ausgezählt und in Tabellen zusammengefasst.

Für die Auswertung der Prätests wurde der Wilcoxon-signed-rank-Test verwendet, um signifikante Unterschiede in den Verhaltensweisen der Hunde zwischen Prätest 1 und 2 zu ermitteln.

Für die Vergleiche der drei Gutachter untereinander und der Bestimmung der Inter-rater-Reliabilität wurde der Permutationstest für gepaarte Beobachtungen und ordinale Daten verwendet, da es sich hier um einen Drei-Stichproben-Test handelte. Mit dem Permutationstest wurde auf signifikante Unterschiede zwischen den Punkteergebnissen und den Kategorien sowohl innerhalb der einzelnen Subtests als auch im Gesamtergebnis der drei Gutachter untersucht. Außerdem wurde beim Vergleich der Kategorien der drei Gutachter zusätzlich Cohens Kappa-Koeffizient als deskriptives Maß der Übereinstimmung von Ergebnissen verwendet.

Für die Vergleiche der zwei Auswertungen der drei einzelnen Gutachter im Zuge der Bestimmung der Intra-rater-Reliabilität wurde der Wilcoxon-signed-rank-Test verwendet. Dieser wurde sowohl beim Vergleich der Punkteergebnisse als auch der Kategorien herangezogen. Cohens Kappa-Koeffizient wurde außerdem als deskriptives Maß für den Vergleich der Kategorien verwendet.

Zur Bestimmung der Test-Retest-Reliabilität wurden ebenfalls die Punkteergebnisse der ersten und zweiten Durchführung des Tests mithilfe des Wilcoxon-signed-rank-Tests und die Kategorien mit dem Wilcoxon-signed-rank-Test und dem Cohens-Kappa-Koeffizienten überprüft.

Um die Ergebnisse des Tests auf Futteraggression mit den Angaben des Datenerhebungsbogens und den Ergebnissen des Referenztests auf signifikante Unterschiede zu untersuchen, wurde beim Vergleich der Kategorien sowohl der Wilcoxon-signed-rank-Test als auch der Cohens-Kappa-Koeffizient bestimmt. Beim Vergleich der Punkteergebnisse zwischen dem Test auf Futteraggression und dem Referenztest kam erneut nur der Wilcoxon-signed-rank-Test zum Einsatz.

Eine Korrektur für multiples Testen wurde bei den statistischen Berechnungen nicht durchgeführt, da es sich um eine explorative Studie handelt.

Eine Signifikanz der Unterschiede wurde für die statistischen Vergleiche mithilfe des Wilcoxon-signed-rank-Tests und des Permutationstests für gepaarte Beobachtungen bei einer Irrtumswahrscheinlichkeit von unter 5% angenommen. So galten Werte von $p \leq 0{,}05$ als signifikant, Werte von $p \leq 0{,}01$ als hochsignifikant und Werte von $p \leq 0{,}001$ als höchstsignifikant.

Cohens-Kappa-Koeffizient dient als deskriptives Maß für die Übereinstimmung der Ergebnisse. Nach der Einteilung von LANDIS und KOCH (1977) gelten Werte von $\kappa \leq 0$ als „schlechte Übereinstimmung“, Werte von κ zwischen 0 und 0,20 als „geringe Übereinstimmung“, Werte von κ zwischen 0,21 und 0,40 als „ausreichende Übereinstimmung“, Werte von κ zwischen 0,41 und 0,60 als „mäßige Übereinstimmung“, Werte von κ zwischen 0,61 und 0,80 als „starke Übereinstimmung“ und Werte von κ zwischen 0,81 und 1,00 als fast vollständige bzw. vollständige Übereinstimmung".

Bei der Überprüfung auf mögliche Korrelationen zwischen der Häufigkeit der Stresssignale und den Punkt- und Kategorienergebnissen des Tests auf Futteraggression sowie bei der Korrelation zwischen den Ergebnissen des Tests auf Futteraggression mit den Ergebnissen des C-BARQS von HSU und SERPELL (2003) wurde Spearmans Korrelationskoeffizient berechnet. Hierbei beschreiben r-Werte kleiner 0 eine negative Korrelation, r-Werte größer 0 eine positive Korrelation und r-Werte gleich 0 keine Korrelation. P-Werte kleiner 0,05 zeigen hierbei eine Signifikanz der Korrelation an.

4 Ergebnisse

4.1 Ergebnisse des Prätests

Die Häufigkeiten der einzelnen Verhaltensweisen in den im Kapitel 3.7.1 beschriebenen Testsituationen des Prätests wurden ausgezählt und in Abbildung 4 über alle Testsituationen zusammengefasst dargestellt (N=17 Hunde). Nicht beobachtete Verhaltensweisen wurden hierbei nicht aufgeführt.

In Prätest 1 trat insgesamt 270 Mal bewertbares Verhaltensweisenauf, wobei 70 (25,9%) soziopositive Verhaltensweisen, 69 (25,5%) agonistische Verhaltensweise und 61 (22,6%) Stresssignale gezeigt wurden. Außerdem wurden 68 (25,2%) andere als agonistische oder soziopositive Verhaltensweisen gezeigt und in 2 (0,7%) Testsituationen waren Hunde außerhalb der Reichweite.

In Prätest 2 zeigten die Hunde 93 (42,7%) soziopositive Verhaltensweisen, 53 (24,3%) agonistische Verhaltensweisen, 70 (32,1%) Stresssignale und 65 (23,1%) andere als agonistische oder soziopositive Verhaltensweisen von insgesamt 281 Verhaltensweisen.

Bei der Betrachtung der einzelnen Verhaltensweisen zeigten die Hunde in Prätest 1 zu 19,6% (53 von 270) „freundliche Annäherung“, 5,2% (14 von 270) „beschnüffeln“, 0,4% (1 von 270) „belecken“, 0,7% (2 von 270) „Spielbewegungen“, 24,8% (67 von 270) „abwenden“, 0,7% (2 von 270) „knurren“, 25,2% (68 von 270) „kein agonistisches oder soziopositives Verhalten“, 0,7% (2 von 270) „außerhalb der Reichweite“, 0,4% (1 von 270) „gähnen“, 8,9% (24 von 270) „pföteln“, 12,6% (34 von 270) „maullecken“ und 0,7% (2 von 270) „schütteln.

Bei der Betrachtung der einzelnen Verhaltensweisen zeigten die Hunde in Prätest 2 zu 19,2% (54 von 281) „freundliche Annäherung“, 8,5% (24 von 281) „beschnüffeln“, 0,7% (2 von 281) „belecken“, 4,6% (13 von 281) „Spielbewegungen“, 18,9% (53 von 281) „abwenden“, 23,1% (65 von 281) „kein agonistisches oder soziopositives Verhalten“, 0,4% (1 von 281) „gähnen“, 12,5% (35 von 281) „pföteln“, 11,7% (33 von 281) „maullecken“ und 0,4% (1 von 281) „schütteln“.

Die Hunde zeigten signifikant mehr soziopositive Verhaltensweisen in Prätest 2 (p=0,0413, N=17 Hunde, Wilcoxon-signed rank Test) als in Prätest 1 (Abb. 5). Dabei wurde in Prätest 2 „freundliche Annäherung“ 1,9% häufiger gezeigt (53-mal in Prätest 1, 54-mal in Prätest 2) und „Beschnüffeln“ 71,4% häufiger (14-mal in Prätest 1, 24-mal

in Prätest 2). 100% häufiger wurde in Prätest 2 „Belecken“ beobachtet (1-mal in Prätest 1, 2-mal in Prätest 2) und „Spielbewegungen“ verzeichneten einen Zuwachs von 550% 2-mal in Prätest 1, 13-mal in Prätest 2). Die Kategorien „agonistisches Verhalten“ und „Stresssignale“ waren nicht signifikant verschieden (agonistisches Verhalten p=0,1183; Stresssignale p=0,5691; N=17 Hunde, Wilcoxon-signed rank Test) (Abb. 5).

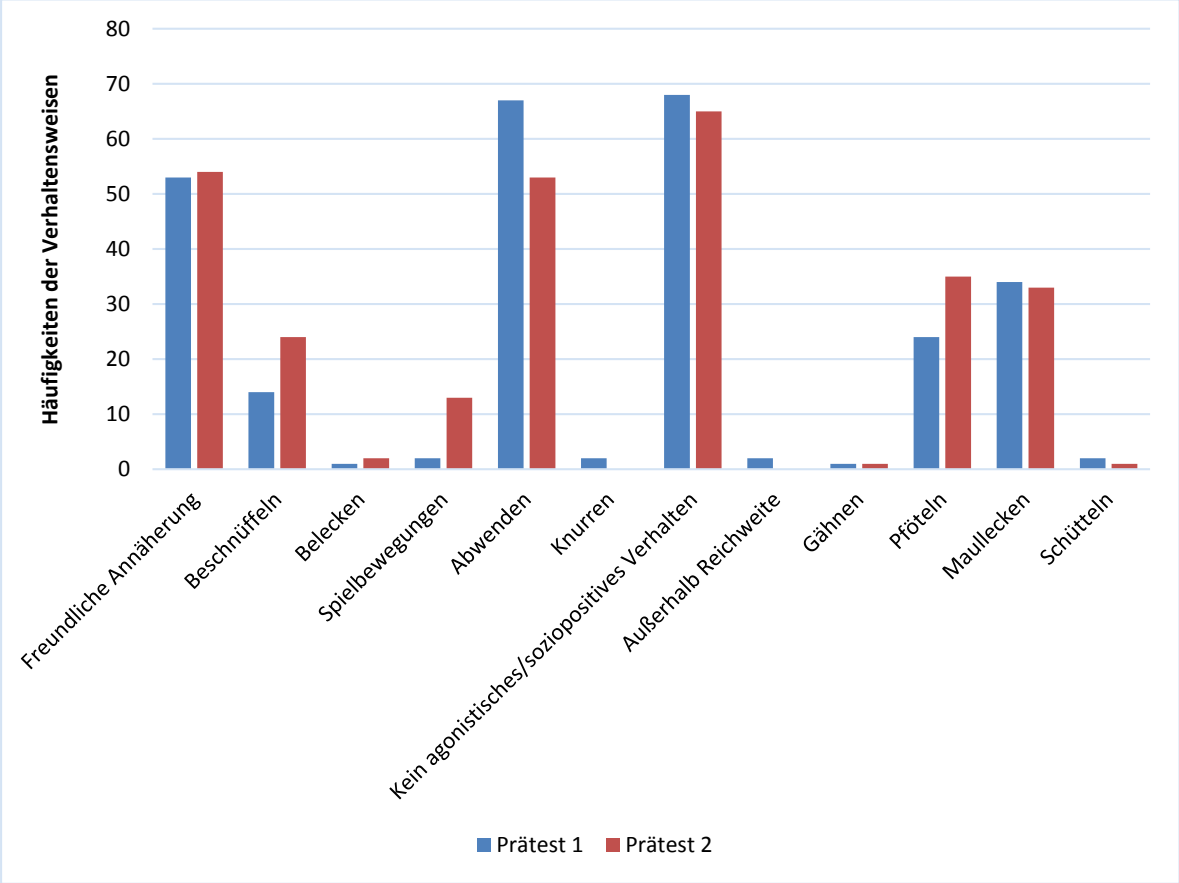

Abbildung 4: Häufigkeiten der gezeigten Verhaltensweisen in Prätest 1 (insgesamt 270 Verhaltensweisen) und Prätest 2 (insgesamt 281 Verhaltensweisen). N=17 Hunde.

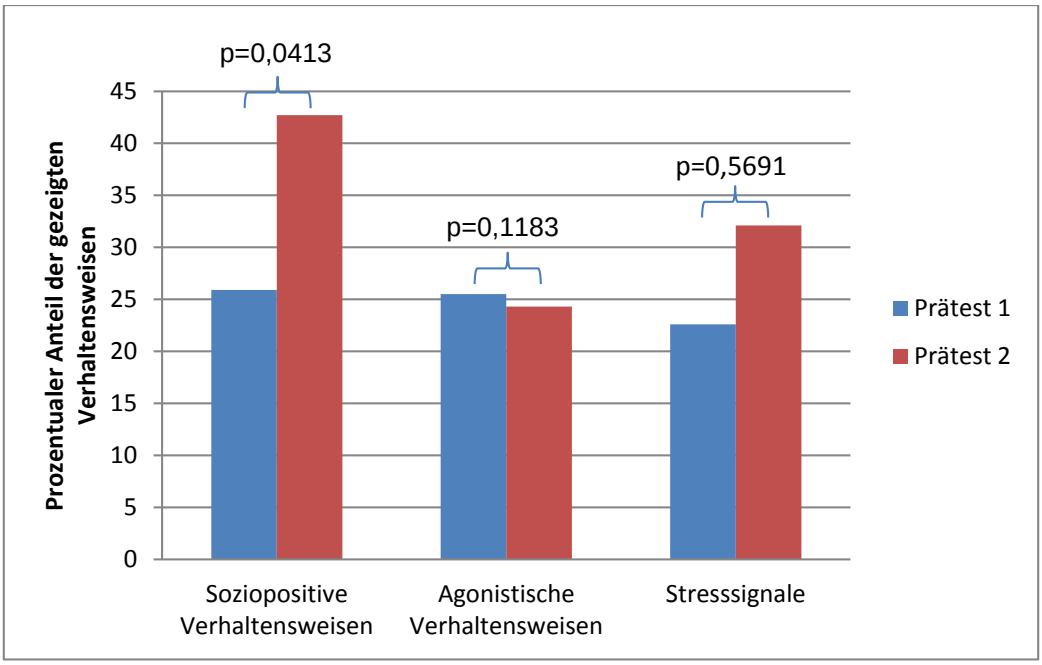

Abbildung 5: Prozentualer Anteil der gezeigten Verhaltensweisen an der Gesamtheit der gezeigten Verhaltensweisen in Prätest 1 und 2. Verhaltensweisen aller Testsituationen zusammengefasst in Kategorien. Statistischer Vergleichs zwischen den einzelnen Kategorien von Prätest 1 und 2 als p-Werte des Wilcoxon-signed-rank-Tests. N=17 Hunde.

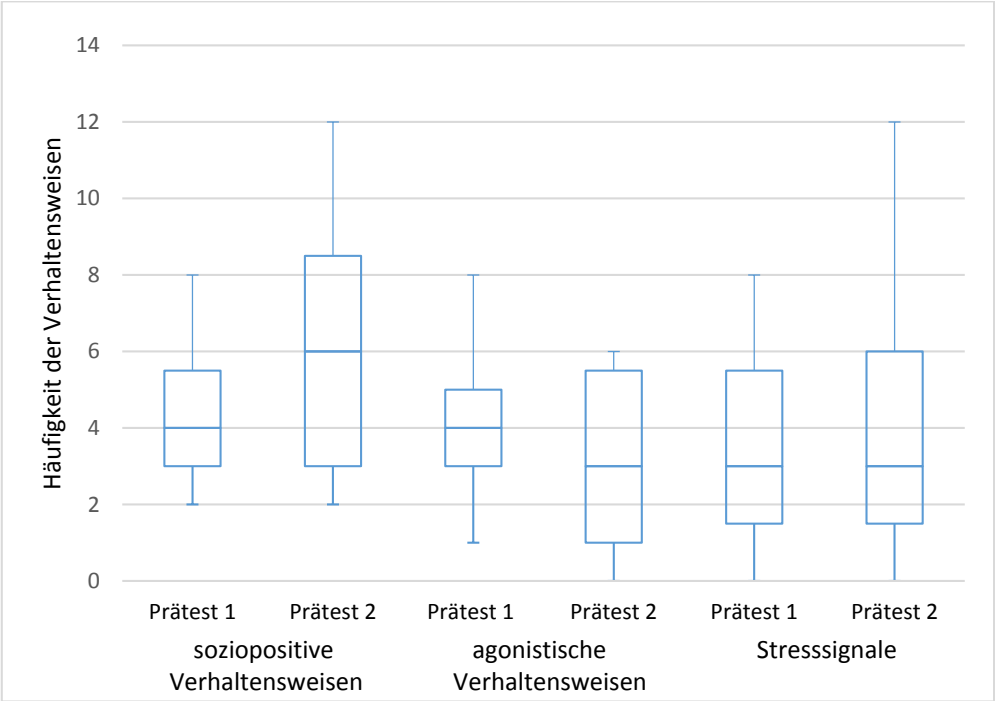

Abbildung 6: Verteilung von Häufigkeiten der Verhaltensweisen in Prätest 1 und 2, zusammengefasst in Kategorien. Boxplots mit Medianen, Interquartilsabständen sowie Minima und Maxima. N=17 Hunde.

4.2. Ergebnisse des Tests auf Futteraggression

4.2.1 Verhalten in den Subtests

Von den drei Gutachtern wurden die Häufigkeiten der einzelnen Verhaltensweisen in jedem Subtest ermittelt, wobei nicht gezeigte Verhaltensweisen nicht aufgeführt wurden.

Tabelle 12 zeigt die Anzahl sowie den prozentualen Anteil an der Gesamtzahl der vom jeweiligen Gutachter beobachteten Verhaltensweisen in Subtest 1, wobei die Stresssignale gesondert gewertet wurden. N=17 Hunde.

Tabelle 12: Anzahl und prozentualer Anteil der von den drei Gutachtern erkannten Verhaltensweisen in Subtest 1. N=17 Hunde.

Subtest 1	Gutachter 1		Gutachter 2		Gutachter 3	
	Anzahl	Prozent	Anzahl	Prozent	Anzahl	Prozent
Abwenden	49	23,2	44	20,7	3	1,4
Abschirmen	7	3,3	11	5,2	6	2,8
Erstarren	2	0,9	2	0,9	7	3,3
Drohfixieren	1	0,5	0	0	0	0
Knurren	2	0,9	2	0,9	1	0,5
Schnappen	1	0,5	1	0,5	1	0,5
Beißen	1	0,5	2	0,9	3	1,4
Kein agonistisches Verhalten	147	69,7	151	70,9	191	90,1
Außerhalb der Reichweite	1	0,5	0	0	0	0
Pfötelln	31	58,5	32	65,3	6	27,3
Maullecken	21	39,6	16	32,7	16	72,7
Schütteln	1	1,9	1	2,0	0	0

In Subtest 2 konnten nur 16 der 17 Hunde bewertet werden, da ein Hund (Studiennummer 9) auch nach wiederholten Versuchen mit attraktivem Futter nicht fraß.

Tabelle 13 zeigt die Anzahl und den prozentualen Anteil der von den drei Gutachtern erkannten Verhaltensweisen in Subtest 2. Die Prozentangaben beziehen sich erneut auf die Gesamtheit der vom jeweiligen Gutachter beobachteten Verhaltensweisen, wobei die Stresssignale gesondert gewertet wurden. N=16 Hunde.

Tabelle 13: Anzahl und prozentualer Anteil der von den drei Gutachtern erkannten Verhaltensweisen in Subtest 2. N=16 Hunde.

Subtest 2	Gutachter 1		Gutachter 2		Gutachter 3	
	Anzahl	Prozent	Anzahl	Prozent	Anzahl	Prozent
Abwenden	15	7,2	14	6,7	6	2,9
Beschleunigter Verzehr	37	17,8	44	21,2	21	10,1
Abschirmen	0	0	0	0	5	2,4
Beißen	0	0	0	0	1	0,5
Kein agonistisches Verhalten	156	75,0	150	72,1	175	84,1
Pföteln	11	68,8	6	40,0	8	100
Maullecken	5	31,2	9	60,0	0	0

Tabelle 14 zeigt die Anzahl und den prozentualen Anteil der von den drei Gutachtern erkannten Verhaltensweisen in Subtest 3. Auch hier beziehen sich die Prozentangaben auf die Gesamtheit der vom jeweiligen Gutachter beobachteten Verhaltensweisen, wobei die Stresssignale gesondert gewertet wurden. N=17 Hunde.

Tabelle 14: Anzahl und prozentualer Anteil der von den drei Gutachtern erkannten Verhaltensweisen in Subtest 3. N=17 Hunde.

Subtest 3	Gutachter 1		Gutachter 2		Gutachter 3	
	Anzahl	Prozent	Anzahl	Prozent	Anzahl	Prozent
Abwenden	3	1,4	0	0	0	0
Beschleunigter Verzehr	15	7,0	16	7,2	2	1,0
Abschirmen	54	25,4	59	26,7	32	16,2
Wegtragen	5	2,3	7	3,2	7	3,5
kein agonistisches Verhalten	122	57,3	126	57,0	143	72,2
außerhalb Reichweite	14	6,6	13	5,9	14	7,1
Pföteln	3	100	6	100	0	0

Anschließend wurde das von den Gutachtern beobachtete Verhalten der Hunde, wie im Kapitel 3.8.1 beschrieben, sowohl mit Punkten als auch mit Kategorien bewertet. Tabelle 15 zeigt die Punkteergebnisse der drei Subtests sowie des Gesamtergebnisses der drei Gutachter für die einzelnen Hunde (N=17 Hunde).

Tabelle 15: Punkteergebnisse der drei Subtests sowie des Gesamtergebnisses der drei Gutachter für die einzelnen Hunde. N=17 Hunde.

	Gutachter 1				Gutachter 2				Gutachter 3			
Hund	Sub-test 1	Sub-test 2	Sub-test 3	Ge-samt	Sub-test 1	Sub-test 2	Sub-test 3	Ge-samt	Sub-test 1	Sub-test 2	Sub-test 3	Ge-samt
1	0	0	8	8	0	0	8	8	0	0	6	6
2	2	7	11	20	4	7	16	27	0	8	4	12
3	13	7	16	36	13	6	14	33	13	8	4	25
4	4	0	4	8	6	0	4	10	2	0	4	6
5	0	0	14	14	0	0	15	15	0	1	14	15
6	13	5	8	26	13	6	10	29	14	0	4	18
7	9	0	6	15	9	3	10	22	2	6	12	20
8	0	0	0	0	0	0	0	0	0	3	0	3
9	0	.	0	0	0	.	0	0	12	.	0	12
10	6	2	10	18	6	2	10	18	6	0	6	12
11	0	0	6	6	0	0	6	6	0	9	4	13
12	2	0	4	6	2	0	4	6	2	0	4	6
13	0	0	0	0	0	0	0	0	3	1	0	4
14	0	0	6	6	0	2	4	6	0	0	2	2
15	0	7	13	20	0	8	10	18	0	2	8	10
16	2	5	12	19	0	0	10	10	0	0	8	8
17	0	4	4	8	0	5	6	11	0	0	0	0

Tabelle 16 zeigt die Kategorienergebnisse der drei Subtests sowie die Gesamtkategorien der drei Gutachter für die einzelnen Hunde (N=17 Hunde).

Tabelle 16: Kategorienergebnisse der drei Subtests sowie die Gesamtkategorien der drei Gutachter für die einzelnen Hunde. N=17 Hunde.

	Gutachter 1				Gutachter 2				Gutachter 3			
Hund	Sub-test 1	Sub-test 2	Sub-test 3	Ge-samt	Sub-test 1	Sub-test 2	Sub-test 3	Ge-samt	Sub-test 1	Sub-test 2	Sub-test 3	Ge-samt
1	1	0	2	2	1	0	2	2	0	0	2	2
2	2	2	2	2	2	2	2	2	0	2	2	2
3	4	2	2	4	4	2	2	4	4	2	2	4
4	2	0	2	2	2	0	2	2	2	0	2	2
5	1	0	2	2	1	0	2	2	0	2	2	2
6	4	2	2	4	4	2	2	4	4	0	2	4
7	3	1	2	3	3	2	2	3	2	2	2	2
8	1	0	0	1	1	0	0	1	1	2	0	2
9	1	.	0	1	1	.	0	1	3	.	0	3
10	2	2	2	2	2	2	2	2	2	0	2	2
11	0	1	2	2	1	1	2	2	0	2	2	2
12	2	0	2	2	2	1	2	2	2	1	2	2
13	0	0	0	0	0	0	0	0	3	2	0	3
14	0	1	2	2	0	2	2	2	0	0	2	2
15	0	2	2	2	0	2	2	2	0	2	2	2
16	2	2	2	2	1	2	2	2	0	0	2	2
17	1	2	2	2	1	2	2	2	0	0	0	0

Mit diesen Ergebnissen wurden anschließend die weiterführenden Berechnungen für die Bestimmung der Reliabilität und Validität des Tests auf Futteraggression durchgeführt.

4.2.2 Inter-rater-Reliabilität des Tests auf Futteraggression

Für die Auswertung mithilfe der Punktevergabe wurden folgende p-Werte für die Übereinstimmung zwischen den drei Gutachtern errechnet:

Es wurden keine signifikanten Unterschiede zwischen den Ergebnissen der drei Gutachter in Subtest 1 ermittelt (p=0,977, N=17 Hunde, Permutationstest für gepaarte Beobachtungen). Im Detail waren die Ergebnisse von Gutachter 1 und 2 nicht signifikant verschieden (p=1,000) und auch die Ergebnisse von Gutachter 1 und 3

(p=0,884) sowie von Gutachter 2 und 3 unterschieden sich nicht signifikant (p=1,000) (Tab. 17).

Tabelle 17: p-Werte für die Übereinstimmung der Punktergebnisse zwischen den Gutachtern (Inter-rater) in Subtest 1. N=17 Hunde.

Subtest 1	Gutachter 2	Gutachter 3
Gutachter 1	1,000	0,884
Gutachter 2		1,000
Gesamt	0,977	

In Subtest 2 (N=16 Hunde, Permutationstest für gepaarte Beobachtungen) konnte beim Vergleich der Ergebnisse der drei Gutachter insgesamt ebenfalls kein signifikanter Unterschied (p=0,996) festgestellt werden. Im Detail waren die Ergebnisse der Gutachter 1 und 2 nicht signifikant verschieden (p=0,883) und auch bei Gutachter 1 und 3 sowie Gutachter 2 und 3 wurde kein signifikanter Unterschied ermittelt (jeweils p=1,000) (Tab.18).

Tabelle 18: p-Werte für die Übereinstimmung der Punktergebnisse zwischen den Gutachtern (Inter-rater) in Subtest 2. N=16 Hunde.

Subtest 2	Gutachter 2	Gutachter 3
Gutachter 1	0,883	1,000
Gutachter 2		1,000
Gesamt	0,996	

Beim Vergleich der Ergebnisse der drei Gutachter von Subtest 3 (N=17 Hunde, Permutationstest für gepaarte Beobachtungen) wurde ein signifikanter Unterschied (p=0,000) errechnet. Im Detail errechnete sich für Gutachter 1 und 2 kein signifikanter Unterschied (p=0,625). Beim Verglich der Ergebnisse von Gutachter 1 und 3 wurde ein signifikanter Unterschied (p=0,004) ermittelt, ebenso beim Vergleich von Gutachter 2 und 3 (p=0,000) (Tab. 19).

Tabelle 19: p-Werte für die Übereinstimmung der Punktergebnisse zwischen den Gutachtern (Inter-rater) in Subtest 3. N=17 Hunde.

Subtest 3	Gutachter 2	Gutachter 3
Gutachter 1	0,625	0,004
Gutachter 2		0,000
Gesamt	0,000	

Beim Vergleich der Gesamtergebnisse der einzelnen Hunde (N=17 Hunde, Permutationstest für gepaarte Beobachtungen) unterschieden sich die drei Gutachter nicht signifikant voneinander (p=0,162). Im Detail erreichten hierbei Gutachter 1 und 2 einen p-Wert von 0,579, Gutachter 1 und 3 einen p-Wert von 0,173 und Gutachter 2 und 3 einen p-Wert von 0,076, welche keinen signifikanten Unterschied erkennen ließen (Tab. 20).

Tabelle 20: p-Werte für die Übereinstimmung der Punktergebnisse zwischen den Gutachtern (Inter-rater) im Gesamtergebnis. N=17 Hunde.

Gesamtergebnis	Gutachter 2	Gutachter 3
Gutachter 1	0,579	0,173
Gutachter 2		0,076
Gesamt	0,162	

Für die Übereinstimmung der Gutachter unter Verwendung der Kategorieneinteilung errechneten sich folgende p-Werte im Permutationstest für gepaarte Beobachtungen und κ -Werte für die Übereinstimmung:

Beim Vergleich der Ergebnisse von Subtest 1 (N=17 Hunde, Permutationstest für gepaarte Beobachtungen und Cohens Kappa-Koeffizient) unterschieden sich die drei Gutachter nicht signifikant voneinander (p=0,787). Im Detail erreichten die Gutachter 1 und 2 eine fast vollständige Übereinstimmung (p=1,000, κ=0,8433), Gutachter 1 und 3 eine Übereinstimmung ohne signifikanten Unterschied (p=0,693, κ=0,3982), ebenso wie Gutachter 2 und 3 (p=0,726, κ=0,3517) (Tab. 21 und 22).

Tabelle 21: p-Werte für die Übereinstimmung der Kategorien zwischen den Gutachtern (Inter-rater) in Subtest 1. N=17 Hunde.

Subtest 1	Gutachter 2	Gutachter 3
Gutachter 1	1,000	0,693
Gutachter 2		0,726
Gesamt	0,787	

Tabelle 22: κ-Werte für die Übereinstimmung der Kategorien zwischen den Gutachtern (Inter-rater) in Subtest 1. N=17 Hunde.

Subtest 1	Gutachter 2	Gutachter 3
Gutachter 1	0,8433	0,3982
Gutachter 2		0,3517

In Subtest 2 (N=16 Hunde, Permutationstest für gepaarte Beobachtungen und Cohens Kappa-Koeffizient) konnte kein signifikanter Unterschied zwischen den Gutachtern ermittelt werden (p=0,849). Im Detail erreichten die Gutachter 1 und 2 eine starke Übereinstimmung ohne signifikante Unterschiede (p=0,259, κ=0,6943). Gutachter 1 und 3 (p=1,000, κ= -0,1355) und Gutachter 2 und 3 (p=0,717, κ=0,0204) (Tab. 23 und 24) hatten dagegen eine schlechte Übereinstimmung ohne signifikante Unterschiede.

Tabelle 23: p-Werte für die Übereinstimmung der Kategorien zwischen den Gutachtern (Inter-rater) in Subtest 2. N=16 Hunde.

Subtest 2	Gutachter 2	Gutachter 3
Gutachter 1	0,259	1,000
Gutachter 2		0,171
Gesamt	0,831	

Tabelle 24: κ-Werte für die Übereinstimmung der Kategorien zwischen den Gutachtern (Inter-rater) in Subtest 2. N=16 Hunde.

Subtest 2	Gutachter 2	Gutachter 3
Gutachter 1	0,6943	-0,1355
Gutachter 2		0,0204

Beim Vergleich der Ergebnisse aller Gutachter von Subtest 3 (N=17 Hunde, Permutationstest für gepaarte Beobachtungen und Cohens Kappa-Koeffizient) wurde kein signifikanter Unterschied ermittelt (p=1,000). Im Detail stimmten die Ergebnisse aller drei Gutachter untereinander nahezu vollständig und ohne signifikanten Unterschied überein (jeweils p=1,000; Gutachter 1 und 2: κ= 1,0000, Gutachter 1 und 3: κ=0,8211, Gutachter 2 und 3 κ=0,8211) (Tab. 25 und 26).

Tabelle 25: p-Werte für die Übereinstimmung der Kategorien zwischen den Gutachtern (Inter-rater) in Subtest 3. N=17 Hunde.

Subtest 3	Gutachter 2	Gutachter 3
Gutachter 1	1,000	1,000
Gutachter 2		1,000
Gesamt	1,000	

Tabelle 26: κ-Werte für die Übereinstimmung der Kategorien zwischen den Gutachtern (Inter-rater) in Subtest 3. N=17 Hunde.

Subtest 3	Gutachter 2	Gutachter 3
Gutachter 1	1,0000	0,8211
Gutachter 2		0,8211

Der Vergleich der Gesamtergebnisse (N=17 Hunde, Permutationstest für gepaarte Beobachtungen und Cohens Kappa-Koeffizient) ergab keinen signifikanten Unterschied (p=0,783). Gutachter 1 und 2 erreichten hierbei im Detail eine vollständige Übereinstimmung ihrer Ergebnisse (p=1,000, κ=1,0000). Die Ergebnisse von Gutachter 1 und 3 stimmten ohne signifikanten Unterschiede überein (p=0,693, κ=0,4333), ebenso wie die Ergebnisse von Gutachter 2 und 3 (p=0,698, κ=0,4333) (Tab. 27 und 28).

Tabelle 27: p-Werte für die Übereinstimmung der Kategorien zwischen den Gutachtern (Inter-rater) im Gesamtergebnis. N=17 Hunde.

Gesamtergebnis	Gutachter 2	Gutachter 3
Gutachter 1	1,000	0,693
Gutachter 2		0,698
Gesamt	0,783	

Tabelle 28: κ-Werte für die Übereinstimmung der Kategorien zwischen den Gutachtern (Inter-rater) im Gesamtergebnis. N=17 Hunde.

Gesamtergebnis	Gutachter 2	Gutachter 3
Gutachter 1	1,0000	0,4333
Gutachter 2		0,4333

4.2.3 Intra-rater-Reliabilität des Tests auf Futteraggression

Für die Bestimmung der Intra-rater-Reliabilität werteten die drei Gutachter die Videoaufnahmen der Hunde des ersten Testdurchlaufs nach zwei Wochen erneut aus. Die Ergebnisse der ersten und zweiten Auswertung von jedem Gutachter wurden anschließend statistisch auf signifikante Unterschiede untersucht. Hierbei wurden erneut die Auswertungen mithilfe der Punktevergabe und der Kategorienbildung verwendet.

Tabelle 29 zeigt die Punkteergebnisse der drei Subtests sowie des Gesamtergebnisses der drei Gutachter für die einzelnen Hunde der zweiten Auswertung (N=17 Hunde).

Tabelle 29: Punkteergebnisse der drei Subtests sowie des Gesamtergebnisses der drei Gutachter für die einzelnen Hunde der zweiten Auswertung. N=17 Hunde.

	Gutachter 1				Gutachter 2				Gutachter 3			
Hund	Sub-test 1	Sub-test 2	Sub-test 3	Ge-samt	Sub-test 1	Sub-test 2	Sub-test 3	Ge-samt	Sub-test 1	Sub-test 2	Sub-test 3	Ge-samt
1	0	0	8	8	0	0	8	8	0	0	6	6
2	4	6	12	22	2	4	14	20	0	8	6	14
3	13	5	12	30	13	5	10	28	11	8	4	23
4	4	0	4	8	4	0	4	8	0	0	4	4
5	0	0	14	14	0	1	12	13	0	0	12	12
6	13	6	8	27	13	5	8	26	13	0	4	17
7	9	4	6	19	7	3	12	22	0	2	12	14
8	0	0	0	0	0	0	0	0	0	3	0	3
9	0	.	0	0	0	.	0	0	0	.	0	0
10	6	2	8	16	6	4	8	18	6	0	6	12

11	0	4	6	10	0	2	6	8	0	2	4	6
12	4	2	4	10	4	2	4	10	0	0	4	4
13	0	0	0	0	0	1	0	1	3	1	0	4
14	0	0	10	10	0	3	6	9	0	0	2	2
15	0	8	10	18	0	9	10	19	0	2	10	12
16	2	5	10	17	2	4	10	16	0	0	10	10
17	0	5	4	9	0	0	2	2	0	0	0	0

Tabelle 30 zeigt die Kategorienergebnisse der drei Subtests sowie die Gesamtkategorien der drei Gutachter für die einzelnen Hunde der zweiten Auswertung (N=17 Hunde).

Tabelle 30: Kategorienergebnisse der drei Subtests sowie die Gesamtkategorien der drei Gutachter für die einzelnen Hunde der zweiten Auswertung. N=17 Hunde.

	Gutachter 1				Gutachter 2				Gutachter 3			
Hund	Sub-test 1	Sub-test 2	Sub-test 3	Ge-samt	Sub-test 1	Sub-test 2	Sub-test 3	Ge-samt	Sub-test 1	Sub-test 2	Sub-test 3	Ge-samt
1	1	0	2	2	0	1	2	2	0	0	2	2
2	2	2	2	2	2	2	2	2	0	2	2	2
3	4	2	2	4	4	2	2	4	3	2	2	3
4	2	0	2	2	2	0	2	2	0	0	2	2
5	1	0	2	2	1	2	2	2	0	0	2	2
6	4	2	2	4	4	2	2	4	4	0	2	4
7	3	2	2	3	3	2	2	3	0	2	2	2
8	1	0	0	1	1	0	0	1	1	2	0	2
9	1	.	0	1	1	.	0	1	0	.	0	0
10	2	2	2	2	2	2	2	2	2	0	2	2
11	1	2	2	2	1	2	2	2	0	2	2	2
12	2	2	2	2	2	2	2	2	0	1	2	2
13	1	0	0	1	1	2	0	2	3	2	0	3
14	0	1	2	2	0	2	2	2	0	0	2	2
15	1	2	2	2	1	2	2	2	0	2	2	2
16	2	2	2	2	2	2	2	2	0	0	2	2
17	1	2	2	2	1	1	2	2	0	0	0	0

Die Überprüfung der Intra-rater-Reliabilität mithilfe der Punkte ergab folgende Ergebnisse (N=17 Hunde, Wilcoxon-signed-rank-Test):
Für Gutachter 1 wurden keine signifikanten Unterschiede in der Auswertung mittels Punktevergabe zwischen der ersten und zweiten Auswertung gefunden (N=17 Hunde, Wilcoxon-signed-rank-Test, Subtest 1: p=0,5000, Subtest 2: p=0,2109, Subtest 3: p=0,5000, Gesamtergebnis: p=0,4766) (Tab. 31).

Tabelle 31: p-Werte für die Übereinstimmung der Punkteergebnisse zwischen Auswertung 1 und 2 (Intra-rater) von Gutachter 1. N=17 Hunde.

Gutachter 1	Subtest 1	Subtest 2	Subtest 3	Gesamt
p-Wert	0,5000	0,2109	0,5000	0,4766

Auch bei Gutachter 2 fanden sich keine signifikanten Unterschiede zwischen der ersten und zweiten Auswertung mithilfe der Punktevergabe (N=17 Hunde, Wilcoxon-signed-rank-Test, Subtest 1: p=1,0000, Subtest 2: p=0,4360, Subtest 3: p=0,1328, Gesamtergebnis: p=0,5806) (Tab. 32).

Tabelle 32: p-Werte für die Übereinstimmung der Punkteergebnisse zwischen Auswertung 1 und 2 (Intra-rater) von Gutachter 2. N=17 Hunde.

Gutachter 2	Subtest 1	Subtest 2	Subtest 3	Gesamt
p-Wert	1,0000	0,4360	0,1328	0,5806

Gutachter 3 (N=17 Hunde, Wilcoxon-signed-rank-Test) erreichte in Subtest 1 einen p-Wert von 0,0313, weshalb die Ergebnisse dieses Subtests in erster und zweiter Auswertung als signifikant verschieden anzusehen sind. In Subtest 2 und 3 sowie im Gesamtergebnis waren die Ergebnisse nicht signifikant verschieden (N=17 Hunde, Wilcoxon-signed-rank-Test,Subtest 2: p=0,2500, Subtest 3: p=0,6250, Gesamtergebnis: p=0,0879) (Tab. 33).

Tabelle 33: p-Werte für die Übereinstimmung der Punkteergebnisse zwischen Auswertung 1 und 2 (Intra-rater) von Gutachter 3. N=17 Hunde.

Gutachter 3	Subtest 1	Subtest 2	Subtest 3	Gesamt
p-Wert	0,0313	0,2500	0,6250	0,0879

Für die Auswertung mithilfe der Kategorien ergaben sich im Vergleich zwischen erster und zweiter Auswertung der Gutachter folgende Ergebnisse:

Gutachter 1 erreichte in allen drei Subtests sowie im Gesamtergebnis eine starke bis vollständige Übereinstimmung der Ergebnisse ohne signifikante Unterschiede (N=17 Hunde, Wilcoxon-signed-rank-Test und Cohens Kappa-Koeffizient, Subtest 1: p=0,25001, κ=0,7628; Subtest 2: p=0,2500, κ=0,6863; Subtest 3: p=1,0000, κ=1,0000; Gesamtergebnis: p=1,0000, κ=0,8917) (Tab. 34).

Tabelle 34: p- und κ-Werte für die Übereinstimmung der Kategorienergebnisse zwischen Auswertung 1 und 2 (Intra-rater) von Gutachter 1. N=17 Hunde.

Gutachter 1	Subtest 1	Subtest 2	Subtest 3	Gesamt
p-Wert	0,2500	0,2500	1,0000	1,0000
κ-Wert	0,7628	0,6863	1,0000	0,8917

Auch Gutachter 2 erreichte in allen drei Subtests eine ausreichende bis vollständige Übereinstimmung der Ergebnisse ohne signifikanten Unterschiede (N=17 Hunde, Wilcoxon-signed-rank-Test und Cohens Kappa-Koeffizient, Subtest 1: p=0,6250, κ=0,6746; Subtest 2: p=0,1563, κ=0,2836; Subtest 3: p=1,0000, κ=1,0000; Gesamtergebnis: p=1,0000, κ=0,8851) (Tab. 35).

Tabelle 35: p- und κ-Werte für die Übereinstimmung der Kategorienergebnisse zwischen Auswertung 1 und 2 (Intra-rater) von Gutachter 2. N=17 Hunde.

Gutachter 2	Subtest 1	Subtest 2	Subtest 3	Gesamt
p-Wert	0,6250	0,1563	1,0000	1,0000
κ-Wert	0,6746	0,2836	1,0000	0,8851

Die Ergebnisse von Gutachter 3 stimmten ebenfalls in allen drei Subtests sowie im Gesamtergebnis mäßig bis vollständig und ohne signifikante Unterschiede überein (N=17 Hunde, Wilcoxon-signed-rank-Test und Cohens Kappa-Koeffizient, Subtest 1:

p=0,0625, κ=0,5330; Subtest 2: p=1,0000, κ=0,8881; Subtest 3: p=1,0000, κ=1,0000; Gesamtergebnis: p=0,5000, κ=0,7518) (Tab. 36).

Tabelle 36: p- und κ-Werte für die Übereinstimmung der Kategorienergebnisse zwischen Auswertung 1 und 2 (Intra-rater) von Gutachter 3. N=17 Hunde.

Gutachter 3	Subtest 1	Subtest 2	Subtest 3	Gesamt
p-Wert	0,0625	1,0000	1,0000	0,5000
κ-Wert	0,5330	0,8881	1,0000	0,7518

4.2.4 Test-Retest-Reliabilität des Tests auf Futteraggression

Die Auswertung des acht Wochen später durchgeführten Re-Tests (zweiter Testdurchlauf) durch den Gutachter 1 kam zu folgenden Ergebnissen (Tab. 37):

Tabelle 37: Punkt- und Kategorienergebnisse des zweiten Testdurchlaufs (Re-Test), ausgewertet von Gutachter 1.

	Punkte				Kategorien			
Retest Hund	Subtest 1	Subtest 2	Subtest 3	Gesamt	Subtest 1	Subtest 2	Subtest 3	Gesamt
1	0	1	10	11	1	2	2	2
2	0	7	10	17	1	2	2	2
3	13	7	4	24	4	2	0	4
4	0	0	4	4	1	1	2	2
5	0	0	15	15	1	0	2	2
6	8	6	8	22	4	2	2	4
7	2	3	11	16	2	2	2	2
8	0	0	0	0	1	0	0	1
9	0	0	2	2	1	0	2	2
10	4	5	10	19	2	2	2	2
11	0	4	4	8	1	2	2	2
12	4	1	2	7	2	2	2	2
13	0	0	2	2	1	0	2	2
14	0	4	4	8	1	2	2	2
15	0	11	4	15	1	2	2	2
16	0	4	15	19	1	2	2	2

17	0	3	8	11	1	2	2	2

Beim Vergleich der Punkte- und Kategorienergebnisse des ersten Testdurchlaufs (Auswertung von Gutachter 1) mit denen des zweiten Testdurchlaufs acht Wochen später (Auswertung Gutachter 1) konnte folgenden Übereinstimmungen errechnet werden:

Die Punkte- und Kategorienergebnisse der Hunde in Subtest 1 unterschieden sich in beiden Tests nicht signifikant voneinander (Punkte: N=17 Hunde, Wilcoxon-signed-rank-Test, p=0,0781; Kategorien: N=17 Hunde, Wilcoxon-signed-rank-Test und Cohens Kappa-Koeffizient, p=1,0000, κ=0,3524) (Tab. 38).

Tabelle 38: p- und κ-Werte für die Übereinstimmung der Punkte- bzw. Kategorienergebnisse des ersten und zweiten Testdurchlaufs (Test-Retest) in Subtest 1. N=17 Hunde.

Subtest 1	Punkte	Kategorien
p-Wert	0,0781	1,0000
κ-Wert		0,3524

In Subtest 2 errechnete sich ein signifikanter Unterschied zwischen den Punkteergebnissen beider Testdurchläufe (N=16 Hunde, Wilcoxon-signed-rank-Test, p=0,0313) Auch bei der Auswertung mit Kategorien konnte ein signifikanter Unterschied festgestellt werden (N=16 Hunde, Wilcoxon-signed-rank-Test und Cohens Kappa-Koeffizient, p=0,0313, κ=0,3642) (Tab. 39).

Tabelle 39: p- und κ-Werte für die Übereinstimmung der Punkte- bzw. Kategorienergebnisse des ersten und zweiten Testdurchlaufs (Test-Retest) in Subtest 2. N=16 Hunde.

Subtest 2	Punkte	Kategorien
p-Wert	0,0313	0,0313
κ-Wert		0,3642

Der Vergleich der Punkte- und Kategorienergebnisse beider Testdurchläufe zeigte in Subtest 3 keine signifikanten Unterschiede (Punkte: N=17 Hunde, Wilcoxon-signed-rank-Test, p=0,8794; Kategorien: N=17 Hunde, Wilcoxon-signed-rank-Test und Cohens Kappa-Koeffizient, p=1,0000, κ=0,3014) (Tab. 40).

Tabelle 40: p- und κ-Werte für die Übereinstimmung der Punkte- bzw. Kategorienergebnisse des ersten und zweiten Testdurchlaufs (Test-Retest) in Subtest 3. N=17 Hunde.

Subtest 3	Punkte	Kategorien
p-Wert	0,8794	1,0000
κ-Wert		0,3014

Auch beim Vergleich der Punkte und Kategorien der Gesamtpunktergebnisse konnten keine signifikanten Unterschiede errechnet werden (Punkte: N=17 Hunde, Wilcoxon-signed-rank-Test, p=0,8315; Kategorien: N=17 Hunde, Wilcoxon-signed-rank-Test und Cohens Kappa-Koeffizient, p=0,7500, κ=0,6047) (Tab. 41).

Tabelle 41: p- und κ-Werte für die Übereinstimmung der Punkte- bzw. Kategorienergebnisse des ersten und zweiten Testdurchlaufs (Test-Retest) im Gesamtergebnis. N=17 Hunde.

Gesamt	Punkte	Kategorien
p-Wert	0,8315	0,7500
κ-Wert		0,6047

4.2.5 Kriteriumsvalidität des Tests auf Futteraggression

Im Folgenden werden die Ergebnisse der Kriteriumsvaliditätsermittlung dargestellt, welche die Vergleiche der Ergebnisse der Hunde des ersten Testdurchlaufs (Auswertung von Gutachter 1) mit den Ergebnissen des Datenerhebungsbogens, des Referenztests und des C-BARQs von HSU und SERPELL (2003) beinhalten.

4.2.5.1 Datenerhebungsbogen

Tabelle 42 beschreibt die Ergebnisse des Datenerhebungsbogens in Kategorien (N=17 Hunde).

Tabelle 42: Kategorienergebnisse des Datenerhebungsbogens laut Angaben der Tierheimleitung. N=17 Hunde.

Hund	leere Futterschüssel	normales Futter	besonderes Futter	Gesamtkategorie
1	0	2	2	2
2	0	2	2	2

3	0	0	2	2
4	0	0	0	0
5	0	2	2	2
6	0	2	2	2
7	0	2	2	2
8	0	2	0	2
9	0	0	2	2
10	0	2	2	2
11	0	2	2	2
12	0	2	2	2
13	0	2	2	2
14	0	3	2	3
15	0	2	2	2
16	0	2	2	2
17	0	2	2	2

Beim Vergleich dieser Kategorien mit denen des ersten Testdurchlaufs (Auswertung von Gutachter 1) des Tests auf Futteraggression wurden folgende p-Werte mit dem Wilcoxon-signed-rank-Test und κ-Werte mit Cohens-Kappa-Koeffizienten ermittelt:

Beim Vergleich der Kategorien der Fragen „leere Futterschüssel“ und „normales Futter“ mit den Kategorien der Subtests 1 und 2 konnten signifikante Unterschiede ermittelt werden (N=17 Hunde, Wilcoxon-signed-rank-Test und Cohens-Kappa-Koeffizient, Vergleich „leere Futterschüssel“ mit Subtest 1: p=0,0002, κ=0,000; Vergleich „normales Futter“ mit Subtest 2: p=0,043, κ=0,0588)

Beim Vergleich der Kategorien der Frage „besonderes Futter“ mit den Kategorien von Subtest 3 wurden keine signifikanten Unterschiede ermittelt (N=17 Hunde, Wilcoxon-signed-rank-Test und Cohens-Kappa-Koeffizient, p=1,0000, κ=0,3014).

Auch der Vergleich der Gesamtkategorien der Hunde zeigte keine signifikanten Unterschiede (N=17 Hunde, Wilcoxon-signed-rank-Test und Cohens-Kappa-Koeffizient, p=0,6406, κ=0,1148) (Tab. 43).

Tabelle 43: p- und κ-Werte für die Übereinstimmung der Gesamtkategorien mit dem Datenerhebungsbogen. N=17 Hunde.

Daten-erhebungs-bogen	Leere Futterschüssel (Subtest 1)	Normales Futter (Subtest 2)	Besonderes Futter (Subtest 3)	Gesamt-kategorie (Gesamt)
p-Wert	0,0002	0,0430	1,0000	0,6406
κ-Wert	0,0000	0,0588	0,3014	0,1148

4.2.5.2 Referenztest

Der Referenztest wurde in den Zwingern der Hunde und von der Tierheimleitung durchgeführt. Da ein Hund vor den Referenztests aufgrund von Krankheit und hohen Alters euthanasiert werden musste, konnten zum Referenztest nur 16 Hunde getestet werden.

Die Häufigkeiten der einzelnen gezeigten Verhaltensweisen in den drei Teilen des Referenztests sind in Tabelle 44 dargestellt (N=16 Hunde).

Tabelle 44: Häufigkeiten der gezeigten Verhaltensweisen in den drei Teilen des Referenztests. N=16 Hunde.

Verhaltensweisen	Teil 1	Teil 2	Teil 3
Abwenden	29	13	.
beschleunigter Verzehr	.	33	4
Festhalten	.	.	6
Abschirmen	3	.	44
Knurren	2	.	.
kein agonistisches Verhalten	105	99	80
Pfötelln	22	9	4
Maullecken	20	.	1
Schütteln	3	.	.

Die Punkte-und Kategorienergebnisse der einzelnen Hunde in den drei Teilen des Referenztests sind in Tabelle 45 dargestellt.

Tabelle 45: Punkte- und Kategorienergebnisse des Referenztests. N=16 Hunde.

	Punkte				Kategorien			
Hund	Teil 1	Teil 2	Teil 3	Gesamt	Teil 1	Teil 2	Teil 3	Gesamt
1	0	0	6	6	1	1	2	2
2	0	6	12	18	0	2	2	2
3	10	6	2	18	3	2	2	3
4	.	.	.	.	.	.	.	.
5	0	0	10	10	1	0	2	2
6	0	5	8	13	1	2	2	2
7	4	0	6	10	2	1	2	2
8	0	0	0	0	1	0	0	1
9	0	0	0	0	1	0	0	1
10	0	3	10	13	1	2	2	2
11	0	0	6	6	0	0	2	2
12	0	0	6	6	0	1	2	2
13	0	0	0	0	1	0	0	1
14	0	0	6	6	1	0	2	2
15	0	6	12	18	1	2	2	2
16	2	4	10	16	2	2	2	2
17	0	3	4	7	0	2	2	2

Beim Vergleich der Punkteergebnisse der einzelnen Subtests des Tests auf Futteraggression (erster Testdurchlauf, Auswertung von Gutachter 1) mit den einzelnen Teilen des Referenztests sowie der Gesamtergebnisse beider Tests konnte folgende Übereinstimmung errechnet werden:

Signifikante Unterschiede der Anzahl errechneter Punkte fanden sich zwischen Teil 1 des Referenztests und Subtest 1 des Tests auf Futteraggression (N=16 Hunde, Wilcoxon-signed-rank-Test, p=0,0313) sowie zwischen den Gesamtpunkten beider Tests (N=16 Hunde, Wilcoxon-signed-rank-Test, p=0,0020). Die Punkteergebnisse von Teil 2 des Referenztests und Subtest 2 des Tests auf Futteraggression unterschieden sich nicht signifikant (N=16 Hunde, Wilcoxon-signed-rank-Test, p=0,2188), ebenso die Punkteergebnisse von Teil 3 des Referenztests und Subtest 3 des Tests auf Futteraggression (N=16 Hunde, Wilcoxon-signed-rank-Test, p=0,5313) (Tabelle 46).

Tabelle 46: p-Werte für die Übereinstimmung der Punktergebnisse mit dem Referenztest. N=16 Hunde.

Referenztest Punkte	Teil 1 (Subtest 1)	Teil 2 (Subtest 2)	Teil 3 (Subtest 3)	Gesamtergebnis
p-Werte	0,0313	0,2188	0,5313	0,0020

Beim Vergleich der Kategorienergebnisse der einzelnen Subtests des Tests auf Futteraggression mit den Teilen des Referenztests sowie der Gesamtkategorien beider Tests konnten folgende Übereinstimmungen ermittelt werden:
Die erreichten Kategorien zwischen den Teilen des Referenztests und den Subtests des Tests auf Futteraggression waren nicht signifikant verschieden (N=16 Hunde, Wilcoxon-signed-rank-Test und Cohens Kappa-Koeffizient, Vergleich Teil 1 und Subtest 1: p=0,1309, κ=0,1398; Vergleich Teil 2 und Subtest 2: p=1,0000, κ=0,5775; Vergleich Teil 3 und Subtest 3: p=1,0000, κ=1,0000; Vergleich Gesamtkategorien: p=0,5000, κ=0,5039) (Tab. 47).

Tabelle 47: p- und κ-Werte für die Übereinstimmung der Kategorien mit dem Referenztest. N=16 Hunde.

Referenztest Kategorien	Teil 1 (Subtest 1)	Teil 2 (Subtest 2)	Teil 3 (Subtest 3)	Gesamtergebnis
p-Werte	0,1309	1,0000	1,0000	0,5000
κ-Werte	0,1398	0,5775	1,0000	0,5039

4.2.5.3 C-BARQ

Tabelle 48 zeigt die Ergebnisse der vier ausgewählten Kategorien des C-BARQs (N=17 Hunde).

Tabelle 48: Ergebnisse des C-BARQs in vier ausgewählten Kategorien. N=17 Hunde.

Hund	Stranger-directed aggression	Owner-directed aggression	Stranger-directed fear	Touch sensitivity
1	0,11	0	2	0,5
2	0,5	0	0	0

3	0	0,17	0	1
4	0,2	0	0,25	1,25
5	0,3	0,14	1	0,67
6	0,2	0	0	0
7	0	0	0	0
8	0	0	0	0,25
9	0	0	0	1,33
10	0	0	0,5	1
11	0,3	0	0,5	0,67
12	0,3	0	0	0,25
13	0	0	0,75	1,25
14	0	0	0,75	1
15	0,4	0	0,3	1,25
16	0	0	1	1
17	0	0	0,75	0,75

Die Überprüfung auf Korrelation von vier ausgewählten Kategorien des C-BARGs ("owner-directed aggression", "stranger-directed aggression", "stranger-directed fear" und "touch-sensitivity") mit den Gesamtergebnissen der Punkte- und Kategorienauswertung des Tests auf Futteraggression mithilfe des Spearmans Korrelationskoeffizienten (r inkl. p-Werte) ergab keine signifikante Korrelation. Im Einzelnen ergaben sich folgende Ergebnisse:

Die Kategorie "owner-directed aggression" korrelierte leicht positiv aber nicht signifikant mit den Punkteergebnissen (N=17 Hunde, Spearmans Korrelationskoeffizient, r=0,353, p=0,1647) und mit den Kategorienergebnissen (N=17 Hunde, Spearmans Korrelationskoeffizient, r=0,346, p=0,1734) des Tests auf Futteraggression. Ebenso korrelierte die Kategorie "stranger-directed aggression" leicht positiv aber nicht signifikant mit den Ergebnissen des Tests auf Futteraggression (N=17 Hunde, Spearmans Korrelationskoeffizient, Punkte: r=0,262, p=0,3103; Kategorien r=0,152, p=0,5598).

Dagegen korrelierte die Kategorie "stranger-directed fear" leicht negativ aber ebenfalls nicht signifikant mit den Punkte- (N=17 Hunde, Spearmans Korrelationskoeffizient, r= -0,120, p=0,6477) und Kategorienergebnissen (N=17 Hunde, Spearmans Korrelationskoeffizient, r= -0,215, p=0,4069) des Tests auf Futteraggression. Auch die

Kategorie "touch-sensitivity" korrelierte leicht negativ und ohne Signifikanz mit den Ergebnissen des Tests auf Futteraggression (N=17 Hunde, Spearmans Korrelationskoeffizient, Punkte: r= -0,220, p=0,3959; Kategorien: r= -0,426, p=0,0881) (Tab. 49).

Tabelle 49: Spearmans-Korrelationskoeffizient (r) und p-Werte zwischen den vier C-BARQ-Kategorien und den Gesamt-Punkteergebnissen und Gesamt-Kategorienergebnissen des Tests auf Futteraggression. N=17 Hunde.

C-BARQ	Owner-directed aggression		Stranger-directed aggression		Stranger-directed fear		Touch sensitivity	
	r	p	r	p	r	p	r	p
Punkte	0,353	0,1647	0,262	0,3103	-0,120	0,6477	-0,220	0,3959
Kategorien	0,346	0,1734	0,152	0,5598	-0,215	0,4069	-0,426	0,0881

4.3 Ergebnisse der Auswertung der Stresssignale

Die von den einzelnen Gutachtern erkannten Stresssignale wurden unabhängig ihrer Art für jeden Subtest ausgezählt und in Tabelle 50 zusammengefasst (N=17 Hunde).

Tabelle 50: Anzahl der von den drei Gutachtern beobachteten Stresssignale je Subtest. N=17 Hunde.

Stresssignale	Subtest 1	Subtest 2	Subtest 3	Gesamt
Gutachter 1	53	16	3	72
Gutachter 2	49	15	6	70
Gutachter 3	22	8	0	30

Anschließend wurden die einzelnen Häufigkeiten der Stresssignale der Hunde in jedem Subtest mit den Punkte- bzw. Kategorienergebnissen der Hunde in jedem Subtest mithilfe des Spearman-Korrelationskoeffizienten (r inkl. p-Werte) auf Korrelationen überprüft.

Die Stresssignale korrelierten in Subtest 1 schwach positiv und nicht signifikant mit den Punkteergebnissen (N=17 Hunde, Spearmans Korrelationskoeffizient, r=0,100, p=0,7013) bzw. schwach negativ und nicht signifikant mit den Kategorienergebnissen (N=17 Hunde, Spearmans Korrelationskoeffizient, r= -0,024, p=0,9261).

Die Ergebnisse der Subtests 2 und 3 sowie die Gesamtergebnisse korrelierten positiv mit der Anzahl der gezeigten Stresssignale (N=17 Hunde, Spearmans

Korrelationskoeffizient, Subtest 2: Punkte: r=0,260, p=0,3304, Kategorien: r=0,510, p=0,0434; Subtest 3: Punkte: r=0,371, p=0,1432, Kategorien: r=0,169, p=0,5175; Gesamtergebnisse: Punkte: r=0,538, p=0,0259, Kategorien: r=0,547, p=0,0232), wobei die Kategorienergebnisse von Subtest 2 sowie die Gesamtpunkte und -kategorien signifikant mit der Anzahl der Stresssignale korrelierten (Tab. 51).

Tabelle 51: Spearmans-Korrelationskoeffizient (r) und p-Werte zwischen der Anzahl der Stresssignale und den Punkte- und Kategorienergebnissen des Tests auf Futteraggression. N=17 Hunde (Subtest 2 N=16 Hunde).

Stresssignale	Subtest 1		Subtest 2		Subtest 3		Gesamtergebnis	
	r	p	r	p	r	p	r	p
Punkte	0,100	0,7013	0,260	0,3304	0,371	0,1432	0,538	0,0259
Kategorien	-0,024	0,9261	0,510	0,0434	0,169	0,5175	0,547	0,0232

5 Diskussion

5.1 Der Prätest

Bei der Verwendung der künstlichen Hand stellte sich die Frage, ob diese echt genug auf die Hunde wirkte, um ähnliche Reaktionen wie in vergleichbaren realen Situationen hervorzurufen. Da Modelle, wie in Kapitel 2.2.2 erwähnt, nie zur Gänze das reale Körperteil simulieren können und sowohl anders aussehen, als auch anders riechen und sich bewegen, kann die künstliche Hand Furcht und Angstaggression auslösen (BARNARD et al., 2012). Auch kann sie von den Hunden als Spielzeug betrachtet werden und nicht als Bedrohung, die sie in der Testsituation darstellen soll (KROLL et al., 2004).

Die signifikant häufigeren soziopositiven Verhaltensweisen gegenüber der künstlichen Hand (Prätest 2) verglichen mit der echten Hand (Prätest 1) lag darin begründet, dass die Hunde in Prätest 2 signifikant häufiger Erkundungs- und Spielverhalten gegenüber der künstlichen Hand zeigten. Das für die Hunde unbekannte Objekt wurde demnach vermehrt exploriert oder vermutlich aufgrund des Gummigeruchs als Spielzeug wahrgenommen. Dies deutet auf eine Unterscheidungsfähigkeit von künstlicher und echter Hand hin, was bedeutend im Hinblick des Einsatzes von künstlichen Händen bei Verhaltenstests ist. Denn sind die Hunde trotz der Bemühungen um eine möglichst lebensechte Benutzung der künstlichen Hand doch zu einer Unterscheidung fähig, ist es wahrscheinlich, dass das Verhalten der Hunde in allen Verhaltenstests, in denen solche Modelle als unnatürlich verlängerter Arm genutzt werden, stark beeinflusst wird. Darum sollte nicht nur der Einsatz selbst, sondern auch die Art der Verwendung einer künstlichen Hand wissenschaftlich erprobt und für jeden Verhaltenstest evaluiert werden. Einen Anfang machten KROLL et al. 2004, welche zu vergleichbaren Ergebnissen kamen wie der Prätest dieser Studie. Hier wurden die Reaktionen von Hunden auf Modelle (Hand und Puppe) mit der Aggressionshistorie der Hunde vergleichen. Dabei kam es im Test mit den Modellen häufig zu falsch positiven und falsch negativen Ergebnissen, wobei nicht nur offensive Aggression, sondern auch Angst und Angstaggression ausgelöst wurde. Eine aggressive Reaktion bei den Modellen heißt demnach nicht, dass die Hunde auch in vergleichbaren realen Umweltsituationen aggressiv reagieren. Darum sollten die Ergebnisse von Tests, welche mit Modellen arbeiten, nicht überbewertet werden. Verhaltenstests, welche mit

Modellen arbeiten, lassen stattdessen lediglich einen Tendenz im Verhalten der Hunde erkennen (KROLL et al., 2004). Sie sollten nur als ein Kriterium betrachtet werden und nie allein ausschlaggebend für die Zukunft eines Hundes sein (KROLL et al., 2004; VAN DER BORG et al., 2010; VALSECCHI et al. 2011; BENNETT et al., 2012; MARDER et al., 2013).

Die Entscheidung, in dieser Studie trotzdem auf das Modell der künstlichen Hand zurückzugreifen, lag einerseits in der Notwendigkeit von Sicherheitsmaßnahmen bei der Provokation von aggressivem Verhalten begründet. Andererseits wurde davon ausgegangen, dass die möglichst realitätsnahe Verwendung der künstlichen Hand die Hunde nicht negativ beeinflussen und weder vermehrtes agonistisches Verhalten noch ein erhöhtes Maß an Stress auslösen würde. Beim Vergleich dieser beiden Kategorien („agonistisches Verhalten“ und „Stresssignale“) konnte tatsächlich kein signifikanter Unterschied zwischen Prätest 1 und 2 errechnet werden, weshalb davon ausgegangen werden kann, dass die künstliche Hand von den Hunden nicht als Bedrohung wahrgenommen wurde. Somit konnte eine Beeinflussung der Ergebnisse des Tests auf Futteraggression, welcher lediglich agonistisches Verhalten messen soll bzw. Stresssignale dokumentiert, durch die künstliche Hand ausgeschlossen werden.

5.2 Das Verhalten in den Subtests des Tests auf Futteraggression

Die in dieser Pilotstudie verwendeten Hunde verhielten sich im Test verhältnismäßig wenig futteraggressiv gegenüber der Testperson. Im Durchschnitt erreichten sie eine Gesamtpunktzahl von 12,35 (Minimum 0 Punkte, Maximum 36 Punkte) und eine Gesamtkategorie von 2 (Minimum 0, Maximum 4), was einer leichten Futteraggression mit Verhaltensweisen wie „beschleunigtem Verzehr“, „abschirmen“, „festhalten“ oder „wegtragen“ gemäß dem in dieser Studie verwendetem Ethogramm (Anhang 9.2) entspricht. "Kein agonistisches Verhalten" wurde am häufigsten beobachtet, gefolgt von "abwenden" in Subtest 1, "beschleunigtem Verzehr" in Subtest 2 und "abschirmen" in Subtest 3. Ob diese drei Verhaltensweisen neben nicht agonistischem Verhalten nur zufällig häufiger vorkamen oder ob Hunde bei den dargebotenen Futtermitteln dieses Verhalten präferieren, könnte in weiterführenden Studien an einer größeren Hundepopulation unterschiedlicher Rassen untersucht werden. Interessant wäre zu untersuchen, ob futteraggressive Hunde generell dazu neigen, eine bestimmte

Verhaltensweise zu zeigen und ob eine solche Verhaltensweise als ein Marker für Futteraggression dienen könnte.

In Subtest 1, in dem den Hunden eine leere Futterschüssel gereicht wurde, zeigten drei der Hunde (Nr. 3, 6 und 7) jedoch auch „erstarren“, „drohfixieren“, „knurren“, „schnappen“ und „beißen“. Da diese drei Hunde in den beiden folgenden Subtests mit normalem Futter und Rinderpansen kein oder nur gering ausgeprägtes Drohverhalten zeigten, kann davon ausgegangen werden, dass es sich bei diesem Aggressionsverhalten um Frustrationsaggression aufgrund der leeren Futterschüssel und den für die Hunde stressreichen Testbedingungen handelte. Einer dieser Hunde (Nr. 3) zeigte während dieses Subtests auch "schütteln" als starkes Stresssignal, während er in den anderen Subtests kein Stresssignal zeigte. Eine niedrige Frustrationstoleranz bestätigte die Tierheimleitung bei diesem und einem weiteren Hund (Nr. 7), wogegen beim dritten Hund (Nr. 6) eine Abrichtung auf einen Hetzärmel nachgewiesen werden konnte und dieser noch immer mit leichtem Beißen in den Arm Aufmerksamkeit einforderte.
Gutachter 3 bewertete eine Testsituation in Subtest 2 bei einen Hund mit "beißen", bei dem die anderen beiden Gutachter sich für "Spielverhalten" entschieden haben. Dieser Hund mit der Studiennummer 11 zeigte auch in Subtest 1 starkes Spielverhalten, sodass dieser Subtest aufgrund von wildem Beißspiel mit der künstlichen Hand abgebrochen werden musste. Auch in seinem Gesamtbild zeigte dieser Hund ein stark ausgeprägtes Spielverhalten ohne Anzeichen von Aggressivität, weshalb davon ausgegangen werden kann, dass es sich in seinem Fall beim Beißen tatsächlich um Spielverhalten handelte.
Da in Subtest 1 wie gezeigt auch andere Arten von Aggressionen für agonistisches Verhalten verantwortlich sein können bzw. Spielbeißen als agonistisches Verhalten fehlinterpretiert werden kann und dadurch der Anteil an falschpositiven Ergebnissen für den Test auf Futteraggression steigt, ist der Einsatz dieses Subtests als fraglich anzusehen. Ein vollständiges Ausschließen falschpositiver und falschnegativer Ergebnisse für einen Test gleich jeder Art ist jedoch nicht möglich (KROLL et al., 2004; TAYLOR und MILLS, 2006) und auch die Vorselektion von Hunden bezüglich anderer Aggressivitätsursachen kann nie zu einhundert Prozent genau sein. Darum sollte der Test auf Futteraggression in weiterführenden Arbeiten mit weniger engen Einschlusskriterien der Hunde bzgl. ihrer Aggressivität auf seine Gültigkeit geprüft

werden. Da Subtest 1 entwickelt wurde, um stark futteraggressive Hunde, die bereits die leere Futterschüssel aggressiv gegen eine Bedrohung verteidigen, und eine von ihnen ausgehende Gefahr zu erkennen, könnte dieser Subtest eventuell auch am Ende des Tests auf Futteraggression durchgeführt werden. Ob er notwendig ist und überhaupt durchgeführt wird, könnte abhängig vom Verhalten der Hunde in Subtest 2 und 3 entschieden werden. Denn Hunde, die in Subtest 2 und 3 nicht oder nur wenig futteraggressives Verhalten zeigen, werden bei einer leeren Futterschüssel höchstwahrscheinlich kein oder nur geringes futterverteidigendes Verhalten zeigen. Um hier falschpositive Ergebnisse aufgrund anderer Aggressivitätsursachen (z.B. Frustrationsaggression) oder Fehlinterpretation von Spielverhalten zu vermeiden, könnte dann auf diesen Subtest verzichtet werden. Dies würde außerdem eine Zeitersparnis bedeuten, welche in einem Tierheim von großem Vorteil wäre. Subtest 1 sollte aber mindestens dann durchgeführt werden, wenn Hunde in Subtest 2 und 3 mäßig bis stark futteraggressiv reagieren, da dann vermutlich die Wahrscheinlichkeit steigt, dass diese Hunde auch bei leerer Futterschüssel mit futterverteidigendem Verhalten reagieren. Ob jedoch wirklich die Bereitschaft der Hunde zur Verteidigung der leeren Futterschüssel proportional zur Futteraggressivität bei normalem bzw. besonderem Futter steigt, kann dabei nur angenommen und sollte in weiterführenden Studien untersucht werden.

5.3 Die Evaluierung des Tests auf Futteraggression

5.3.1 Zur Standardisierung und Objektivität des Tests

Die in dieser Studie verwendeten Hunde stammten aus dem Tierheim "Bullterrier in Not e.V." in Herford und gehörten ausschließlich den Rassen Bullterrier und Miniatur-Bullterrier an. Die Beschränkung auf nur eine einzige Rasse war notwendig, um im Zuge der Evaluierung Unterschiede im Verhalten aufgrund verschiedener Rassedispositionen zu vermeiden (DIEDERICH und GIFFROY, 2006). Es wurde weiterhin nur dieses Tierheim gewählt, um Abweichungen bei der Durchführung des Tests (v.a. Testgelände) und bei den Haltungsbedingungen der Tiere auszuschließen. So konnten eine weitestgehend einheitliche Hundepopulation und ein hohes Maß an Standardisierung sichergestellt werden, welche für die Evaluierung eines Verhaltenstests laut TAYLOR und MILLS (2006) sowie DIEDERICH und GIFFROY (2006) zwingend notwendig sind. Durch die Beschränkung auf eine Rasse und ein

Tierheim kann aus den Ergebnissen dieser Studie jedoch kein Rückschluss auf Hunde aller Rassen oder Tierheime im Allgemeinen geschlossen werden. Die durchgeführte Studie ist vielmehr als Pilotstudie zu betrachten, weshalb der Test auf Futteraggression in weiterführenden Arbeiten mit Hunden anderer Rassen durchgeführt und überprüft werden sollte (DIEDERICH und GIFFROY, 2006), um zu prüfen, ob der Test auch dann zuverlässig Futteraggression misst.

Um weiterhin ein hohes Maß an Standardisierung zu erreichen, wurde die Methodik des Tests auf Futteraggression sowie die Durchführung der Prätests detailliert beschrieben und bei jedem Hund streng nach Anleitung durchgeführt. So wurde jeder Hund denselben Testbedingungen ausgesetzt, bei denen das Verhalten der Hunde die einzige variable Messgröße war (DIEDERICH und GIFFROY, 2006).

Da das Testgelände im Freien und nicht überdacht war, konnte der Test nur dann stattfinden, wenn das Wetter entsprechend stabil war. Es wurde nicht bei starkem Wind, Regen oder Schneefall getestet, um auch hier keine wetterbedingten Abweichungen im Verhalten der Hunde auszulösen. Eine weitere Standardisierung der Wetterbedingungen war nicht möglich, sodass sich die Temperaturen an den einzelnen Testtagen zum Teil um mehrere Grad Celsius unterschieden. Erstrebenswert wäre daher ein Testbereich im Inneren eines Gebäudes, beispielsweise ein weitestgehend leerer und ausreichend großer Raum. In vielen Tierheimen ist ein solcher Raum jedoch oftmals nicht vorhanden, weshalb dann, wie in dieser Studie, nur auf ein eingezäuntes Auslaufgelände zurückgegriffen werden kann.

Eine allgemeine Untersuchung wurde bei jedem getesteten Hund durchgeführt, da einige der Tiere vorberichtlich an diversen chronischen Erkrankungen wie Futtermittelallergien, Herz- oder Nierenleiden erkrankt waren, welche einen Ausschluss vom Test jedoch nicht nötig machten. Dabei fielen keine vom Test ausschließenden Krankheiten oder Schmerzzustände auf. Diese Feststellung war wichtig, da laut BREUER (2000) rund 30% der Verhaltensauffälligkeiten auf organische Ursachen zurückzuführen sind.

Zudem wurden die Tests von derselben Testperson durchgeführt, die dabei immer gleich gekleidet war und sich gegenüber den Hunden neutral verhielt. Außerdem wurde die Auswertung des gezeigten Verhaltens mittels genau definierten Ethogramme, einheitlicher Auswertungstabelle sowie einheitlicher Punkteskalierung und Kategorisierung durchgeführt, welche im Vorfeld der Versuche entwickelt und

standardisiert wurden. So wurde versucht sicherzustellen, dass das Verhalten jedes Hundes auch von verschiedenen Gutachtern unter Verwendung der gleichen Maßstäbe bewertet wurde.

Eine große Schwierigkeit beim Beurteilen von Hundeverhalten besteht darin, möglichst nicht anthropomorph, also aus menschlicher Sichtweise, und subjektiv zu bewerten, sondern objektiv das gezeigte Verhalten auf seinen biologischen Sinn zu prüfen (FEDDERSEN-PETERSEN, 2008). Denn eine subjektive Datenaufnahme ist meist schwer zu vereinheitlichen, vor allem wenn andere Personen auswerten als der Entwickler des Tests (DIEDERICH und GIFFROY, 2006). Durch die Verwendung von Ethogrammen, in denen das relevante Verhalten genau definiert ist, wird die Aufnahme des Hundeverhaltens in Verhaltenstests objektiviert und eine direkte Interpretation des Verhaltens verhindert (FEDDERSEN-PETERSEN, 2008). Eine genaue Beschreibung des Verhaltens sowie eine einheitliche Auswertungs- und Bewertungsmethode sind dazu unumgänglich. Hier existieren jedoch bisher keine Vorgaben oder Richtlinien (DIEDERICH und GIFFROY, 2006). Die in dieser Studie genutzten Ethogramme für die Prätests (Anhang 9.4) und den Test auf Futteraggression (Anhang 9.2) wurden auf Grundlage von EISFELD (1966), ZIMEN (1971 und 1990), FEDDERSEN (1978), RUGAAS (2001), DONALDSON (2002), VAN DEN BERG et al. (2003), FEDDERSEN-PETERSEN (2004), TALCEK (2005), VAN DER BORG et al. (2010) und VALSECCHI et al. (2011) entwickelt. Es wurden nur Verhaltensweisen aus den Bereichen der Agonistik, der soziopositiven Verhaltensweisen und der Stresssignale aufgeführt, welche für die entwickelten Tests relevant waren.
Da die Mimik und Gestik von Hunden jedoch sehr dezent sein und sich die Signale sehr schnell ändern können, ist die menschliche Wahrnehmungsfähigkeit schnell überschritten. Gerade aggressive Verhaltensweisen sind mit schnellen Bewegungsabläufen in Mimik, Gestik und Körpersprache verbunden (JONES, 2009). Deswegen sind Videoaufnahmen für die Beurteilung von Verhalten bei Hunden sinnvoll, da durch wiederholte und gegebenfalls verlangsamte Wiedergabe auch kleine Details der Körpersprache berücksichtigt werden können, die andernfalls übersehen worden wären. Eine Beurteilung des Gesamtverhaltens ist damit genauer und sicherer als durch direkte Beobachtung (NETTO und PLANTA, 1997; HIRSCHFELD, 2005; OTT et al., 2008; MAZUR 2012).

Aufgrund dieser Maßnahmen ist davon auszugehen, dass der Test auf Futteraggression standardisiert und objektiv ist, was als Voraussetzung für Reliabilität und Validität eines Tests gilt (TAYLOR und MILLS, 2006; DIEDERICH und GIFFROY, 2006).

5.3.2 Die Reliabilität des Tests auf Futteraggression

5.3.2.1 Die Inter-rater-Reliabilität des Tests auf Futteraggression

Als Gutachter wurden drei Personen mit unterschiedlichen Wissensgrad und Vorerfahrung in Bezug auf Hundeverhalten gewählt, um eine Verzerrung (Bias) der Ergebnisse aufgrund von Erfahrung und Übung in der Auswertung solcher Videos zu reduzieren (KROLL et al., 2004).

Bei der Überprüfung der Inter-rater-Reliabilität fiel auf, dass die Übereinstimmung der Ergebnisse der drei Gutachter bei der Auswertung mithilfe der Kategorieneinteilung insgesamt besser war als die Übereinstimmung der Ergebnisse bei der Auswertung mithilfe der Punktevergabe, weshalb diese Methode der Auswertung bei weiterer Benutzung des Tests auf Futteraggression vorzuziehen ist. So erreichte im direkten Vergleich der p-Werte für die einzelnen Subtests und der Gesamtergebnisse der Auswertung mit Punkten und Kategorien die Auswertung mithilfe der Kategorien in Subtest 1 und 2 nur leicht niedrigere p-Werte (p=0,787 und p=0,832), also nur eine etwas schlechtere Übereinstimmung der Ergebnisse der drei Gutachter, als die Auswertung mit Punkten (p=0,977 und p=0,996). Dagegen erreichte die Auswertung mit Kategorien in Subtest 3 eine komplette Übereinstimmung der Ergebnisse der Gutachter mit einem p-Wert von 1,000, wogegen die Auswertung mit Punkten höchstsignifikante Unterschiede zwischen den Ergebnissen der Gutachter mit einem p-Wert von 0,000 wiederspiegelte. Im Gesamtergebnis erreichte die Auswertung mithilfe der Kategorien die deutlich bessere Übereinstimmung der Ergebnisse der Gutachter (p=0,783) als die Auswertung mithilfe der Punktevergabe (p=0,162).

Als Grund für die zum Teil niedrigeren p-Werte bei Verwendung der Punkteauswertung kann deren hohe Empfindlichkeit in Bezug auf die Verhaltensweisen und deren Häufigkeiten in Betracht gezogen werden. So spielen bei der Punktevergabe vor allem die Häufigkeiten der gezeigten bzw. erkannten Verhaltensweisen eine entscheidende Rolle für die Endergebnisse der einzelnen Hunde, da hier die Punkte für gezeigtes Verhalten in den jeweiligen Testsituationen der Subtests summiert werden, wogegen

bei der kategorialen Auswertung nur das Auftreten oder Nichtauftreten bzw. das Erkennen oder Nichterkennen von Verhaltensweisen relevant ist. Erkennt ein Gutachter bei der Auswertung eines Hundes also nicht so häufig die Verhaltensweisen aus dem Ethogramm wie ein anderer Gutachter, erreicht der erste Gutachter automatisch einen geringeren Punktwert als der Zweite, auch wenn die gleichen Verhaltensweisen erkannt wurden. Dagegen erreichen die Gutachter dieselben Kategorien, wenn sie dieselben Verhaltensweisen erkennen, unabhängig von ihrer Häufigkeit.

Dieses Phänomen wurde deutlich beim Vergleich der ersten beiden Gutachter mit Gutachter 3, welcher, generell gesehen, weniger Verhaltensweisen dokumentiert hat als die anderen beiden Gutachter. So erkannte der erste Gutachter 126, der zweite Gutachter 144 und der dritte Gutachter nur 91 punktpflichtige Verhaltensweisen[9] bei der ersten Auswertung. Dies würde zum Teil auch die schlechteren Übereinstimmungen der Gutachter 2 und 3 erklären (vor allem bei Subtest 3, p=0,000), da die Differenz der erkannten punktpflichtigen Verhaltensweisen bei diesen beiden Gutachtern besonders hoch ist (z.B. in Subtest 3, 82 bzw. 41 erkannte punktpflichtige Verhaltensweisen). Auch die Punktergebnisse von Gutachter 1 und Gutachter 3 sind in Subtest 3 hochsignifikant verschieden, weshalb in der Gesamtübereinstimmung für diesen Subtest der p-Wert bei 0,000 liegt.

Eine mögliche Erklärung für den Umstand, dass Gutachter 3 zahlenmäßig weniger Verhaltensweisen erkannt hat als die anderen beiden Gutachter, könnte dessen langjährige Erfahrung im Bereich Hundeverhalten und Wesenstestauswertung sein. Auf eine mündliche Einweisung in die Auswertung der Testvideos und des Ethogramms wurde aus diesem Grund von Seiten der Doktorandin bewusst verzichtet und nur eine schriftliche Anleitung zur Verwendung der Auswertungstabellen und des Ethogramms inklusive einer E-Mail-Adresse für Rückfragen an Gutachter 3 übermittelt. Dagegen erhielt Gutachter 2, welcher keinerlei Vorerfahrung bezüglich der Auswertung von Wesens- oder Verhaltenstests aufwies, sowohl dieselbe schriftliche Anleitung zur Auswertung, als auch mündliche Erläuterungen zu einem Probevideo, welches nicht für diese Studie genutzt wurde. Außerdem nahm Gutachter 2 die Möglichkeit zu Rückfragen bezüglich des Ethogramms wahr, die Gutachter 3 zwar ebenfalls hatte, aufgrund der langjährigen Vorerfahrungen aber nicht nutzte. Auf diese Weise kann die relativ gute Übereinstimmung der Ergebnisse von Gutachter 1 und 2

[9] =Verhaltensweisen, die mit mindestens einem Punkt bewertet werden

und die etwas schlechtere Übereinstimmung der Ergebnisse dieser beiden Gutachter mit Gutachter 3 möglicherweise erklärt werden.

Bei der weiteren Verwendung des Tests auf Futteraggression sollte daher darauf geachtet werden, dass alle beteiligten Gutachter dieselbe ausführliche Einweisung zur korrekten Auswertung erhalten, um eine möglichst hohe Inter-rater-Reliabilität zu erhalten. Außerdem könnte der statistische Vergleich noch einmal mit den Rohdaten der einzelnen erkannten Verhaltensweisen der drei Gutachter durchgeführt werden, um zu prüfen, ob die Methode der Punkteauswertung für die schlechte Übereinstimmung der Inter-rater-Reliabilität verantwortlich sein könnte, da die Einteilung in Kategorien eine gute Übereinstimmung der Ergebnisse der drei Gutachter ohne signifikante Unterschiede erzielte. In diesem Fall würden die Häufigkeiten der einzelnen erkannten Verhaltensweisen der Gutachter miteinander verglichen werden, wobei dann keine signifikanten Unterschiede gefunden werden dürften. Fänden sich jedoch Unterschiede, würde dies eine Abweichung zwischen den Auswertungen der Gutachter beweisen.

Für die Bestimmung der Inter-rater-Reliabilität wurden aber deshalb die Bewertungssysteme mit Punkte- und Kategorienauswertung verwendet, da diese bei der Evaluierung eines Tests immer mitevaluiert werden müssen (TAYLOR und MILLS, 2006). Würden nur die Rohdaten, in diesem Fall also die Häufigkeiten der einzelnen Verhaltensweisen, für die Evaluierung miteinander verglichen werden, könnte eine eventuelle Verzerrung der Ergebnisse durch Punkte bzw. Kategorien nicht erkannt werden. Die Umwandlung von Verhaltensweisen in auswertbare Daten mithilfe eines Bewertungssystems ist aber dahingehend notwendig, als dass ein Verhaltenstest bestimmte Verhaltensweisen gezielt messen soll, wodurch eine Beurteilung und Wertung des Hundeverhaltens möglich ist. Dies könnte allein durch das Aufnehmen und Auszählen von Verhaltensweisen nicht erreicht werden. Auch in der Literatur finden sich verschiedenste Bewertungssysteme zur Beurteilung von Verhalten: Auszählen von Frequenz und Dauer des Verhaltens (VAN DER BORG et al., 1991; LEDGER und BAXTER, 1997; VAN DEN BERG et al., 2003), Einteilung mit abgestuften Punkten nach Intensität des Verhaltens (NETTO und PLANTA, 1997; VAN DER BORG et al., 2010; VALSECCHI et al., 2011, BENNETT et al., 2012) oder Einteilung der Verhaltensweisen in Kategorien (BOLLEN und HOROWITZ, 2008; DOWLING-GUYER et al., 2011). Die beiden gängigsten Methoden, das Bewerten von Verhalten mit Punkten sowie die Einteilung in Kategorien, sollten in dieser Studie für

die Auswertung des Tests auf Futteraggression verwendet und im direkten Vergleich gegenüber gestellt werden.

Das für diese Studie entwickelte und in Kapitel 3.8.2 beschriebene Bewertungssystem mithilfe einer Punkteskalierung hat sich im Verlauf der Studie als weniger geeignet zum Zwecke der reliablen und validen Beurteilung der Hunde erwiesen. So ergaben sich für die Bestimmungen der Inter- und Intra-rater-Reliabilität sowie für die Test-Retest-Reliabilität zum Teil schlechte Übereinstimmungen der Ergebnisse mit signifikanten Unterschieden (p=0,0000 in Subtest 3 bei der Inter-rater-Reliabilität aller Gutachter; p=0,0313 in Subtest 1 bei der Intra-rater-Reliabilität von Gutachter 3 und p=0,0313 in Subtest 2 bei der Test-Retest-Reliabilität). Auch bei der Überprüfung der Validität mithilfe des Referenztests ergaben sich bei der Verwendung der Punktebewertung schlechte Übereinstimmungen mit signifikanten Unterschieden (p=0,0313 bei Subtest 1 und p=0,0020 beim Gesamtergebnis im Vergleich mit dem Referenztest).

Stattdessen hat sich die Bewertung mithilfe einer im selben Kapitel erläuterten Kategorieneinteilung als zuverlässiger und praktikabler herausgestellt. Die Übereinstimmungen bei der Berechnung der Reliabilität waren besser unter Verwendung der Kategorieneinteilung als mithilfe der Punktevergabe. So ergaben sich im Vergleich zu den oben genannten p-Werten für die Punktevergabe bei der Bestimmung der Inter-rater-Reliabilität in Subtest 3 bei der Verwendung der Kategorien ein p-Wert von 1,0000 und bei der Intra-rater-Reliabilität von Gutachter 3 in Subtest 1 ein p-Wert von 0,0625. Nur der p-Wert in Subtest 2 bei der Bestimmung der Test-Retest-Reliabilität nahm auch bei der Bewertung mithilfe der Kategorien einen Wert von p=0,0313 an, was ebenfalls auf einen signifikanten Unterschied der Ergebnisse hindeutet. Dagegen erreichte die Auswertung mit Kategorien im Vergleich zu der Auswertung mit Punkten bessere p-Werte bei der Überprüfung der Validität mithilfe des Referenztests (p= 0,1309 bei Subtest 1 und p=0,5000 beim Gesamtergebnis).

Diese Ergebnisse sind nicht verwunderlich, da eine Einteilung von 18 Verhaltensweisen in fünf Kategorien das gezeigte Verhalten der Hunde in Art und Häufigkeit deutlich reduziert und damit ungenauer ist als das Bewerten der Verhaltensweisen mit Punkten, was sensibler ist, weil die einzelnen Verhaltensweisen gewertet werden und auch ihre Häufigkeit Einfluss auf das Ergebnis nimmt.

Eine Auswertung des Verhaltens der Hunde mithilfe einer kategorialen Einteilung des Verhaltens ist jedoch dahingehend sinnvoll, als das es eine schnelle und einfache Interpretation der Ergebnisse zulässt, vor allem, wenn eine große Zahl an Hunden getestet werden muss (TAYLOR und MILLS, 2006). Bei der Vergabe von Punkten müssen dagegen zunächst experimentell und mit einer hohen Anzahl an Hunden Normbereiche erstellt werden, um eine Interpretation der Punkteergebnisse eines einzelnen Hundes zu ermöglichen. Diese liegen für den hier entwickelten Test auf Futteraggression noch nicht vor. Bei diesen Normbereichen sind jedoch auch immer rassespezifische Abweichungen zu berücksichtigen. Die Auswertung mithilfe der Punktevergabe ist sensibler im Hinblick auf die einzelnen Verhaltensweisen und genauer aufgrund der Berücksichtigung ihrer Häufigkeit als die Auswertung mithilfe der Kategorien. Eine Beurteilung einzelner Hunde oder der Vergleich einzelner Hunde mit sich selbst, z.B. im Zuge einer Verhaltenstherapie und der Überprüfung von Therapieerfolgen, ist mithilfe der Punkteauswertung besser möglich. Zum Zweck der ungefähren Einschätzung des Verhaltens vieler Hunde in einem Verhaltenstest, etwa wie es in Tierheimen nötig und üblich ist, ist die Bewertung mithilfe der Kategorieneinteilung jedoch besser geeignet, da sie einfacher auszuwerten, reliabler und valider ist sowie eine schnelle Interpretation zulässt. Dennoch gilt auch hier, dass eine höhere Anzahl an Kategorien (z.B. fünf Kategorien, 0= kein agonistisches Verhalten, 1= Meide- und Fluchtverhalten, 2= schwache Futteraggression, 3= moderate Futteraggression, 4= starke Futteraggression) das Verhalten genauer erfasst als wenn komplexes Verhalten in wenigen Kategorien (z.B. zwei Kategorien, keine Futteraggression und Futteraggression) zusammengefasst und zu sehr reduziert wird, wodurch eine Validierung kaum möglich wird (TAYLOR und MILLS, 2006). Denn dann würden Rohdaten verloren gehen und komplexes Verhalten zu groben Begriffen zusammengefasst, wodurch voreilige Entscheidungen über die Zukunft des Hundes gefällt werden könnten.

Verhaltenstests werden entwickelt, um in verschiedenen Tierheimen oder Auffangstationen und damit von vielen verschiedenen Personen durchgeführt und beurteilt zu werden. Dementsprechend ist es wichtig, die Übereinstimmung der Beurteilungen zwischen verschiedenen Bewertern zu überprüfen (MURPHY, 1998; TAYLOR und MILLS, 2006). Die Inter-rater-Reliabilität ist eine der wichtigsten Qualitätsmerkmale für Verhaltenstests, da sie die Unabhängigkeit der Ergebnisse vom

Gutachter beweist (DIEDERICH und GIFFROY, 2006). Ein Verhaltenstest, dessen Ergebnis vom bewertenden Gutachter abhängt und mit jedem anderen Gutachter signifikant variieren kann, ist nicht geeignet, Hundeverhalten vorherzusagen und auf dessen Grundlage Entscheidungen bezüglich der Zukunft des Hundes zu treffen. Darum ist die Inter-rater-Reliabilität auch unter dem Aspekt des Tierschutzes eines der entscheidendsten Qualitätsmerkmale für einen Verhaltenstest.
Abgesehen von den zum Teil schlechten Übereinstimmungen mit Gutachter 3, für welche vor allem die ungenügende Einweisung durch den Doktoranden und das sensible Bewertungssystem mit Punkten verantwortlich ist, hat sich der Test auf Futteraggression als reliabel in Bezug auf die Inter-rater-Reliabilität erwiesen. Er ist damit geeignet, von verschiedenen Bewertern ausgewertet zu werden, vorausgesetzt, dass sie alle dieselbe Einweisung und Anleitung zum Auswerten erhalten haben.

5.3.2.2 Die Intra-rater-Reliabilität des Tests auf Futteraggression

Die Ergebnisse zeigen, dass Gutachter 1 und 2 bei allen Subtests und im Gesamtergebnis sowohl bei der Auswertung mithilfe der Punkte als auch mittels der Kategorien gute Übereinstimmungen ohne signifikante Unterschiede erreicht haben. Dabei ergab die Punktevergabe bessere Ergebnisse in Subtests 1, bei Gutachter 2 auch in Subtest 2, wogegen die Ergebnisse des dritten Subtests sowie die Gesamtergebnisse bei der Auswertung mittels Kategorien vollständige Übereinstimmungen (p=1,0000) zeigten. Bei Gutachter 3 konnte lediglich bei der Auswertung mithilfe der Punktevergabe beim ersten Subtest ein signifikanter Unterschied festgestellt werden (p=0,0313). Bei der Auswertung dieses Subtests mithilfe der Kategorien lag der p-Wert dagegen knapp oberhalb eines signifikanten Unterschieds (p=0,0625). Bei Gutachter 3 zeigten weiterhin alle Kategorienergebnisse bessere Übereinstimmung als die Auswertung mithilfe der Punktevergabe.
Diese Unterschiede bei Subtest 1 von Gutachter 3 können erneut darauf zurückzuführen sein, dass die Häufigkeiten der einzelnen erkannten Verhaltensweisen eine große Rolle bei der Punkteauswertung spielten, wodurch diese sehr sensibel wird. Um diese Annahme zu bestätigen, müsste auch hier anhand der Häufigkeiten der einzelnen erkannten Verhaltensweisen ein Vergleich der Rohdaten durchgeführt werden, welche dann ohne signifikante Unterschiede übereinstimmen müssten. Ein weiterer Grund für die Abweichungen von Gutachter 3 in Subtest 1 könnte jedoch auch eine fehlerhafte Auswertung an sich sein. Hierbei muss vor allem das zum Teil nicht

eindeutige Verhalten der Hunde in Subtest 1 erwogen werden, welches, wie im Kapitel 5.1 bereits erläutert, Spielraum für Fehlinterpretationen ließ. So erkannte Gutachter 3 bei der ersten Auswertung des Subtests 3-mal „abwenden“ und in der zweiten Auswertung 4-mal. Weiterhin beobachtete Gutachter 3 in der ersten Auswertung dieses Subtests 6 mal „abschirmen“, bei der zweiten Auswertung jedoch nur noch 3 mal. Ebenso beobachtete Gutachter 3 in der ersten Auswertung 7-mal „erstarren“ und in der zweiten Auswertung nur noch 3-mal. Von 3-mal „beißen“ in der ersten Auswertung erkannte Gutachter 3 in Subtest 1 bei der zweiten Auswertung nur noch 1 mal wieder. Hier wird deutlich, dass Gutachter 3 vor allem bei den aggressiven Verhaltensweisen wie „abschirmen“, „erstarren“ und „beißen“ unsicher war, was zu Fehlinterpretationen der Verhaltensweisen und einer mangelhaften Intra-rater-Reliabilität geführt haben könnte.

Dagegen ist die Auswertung der Kategorien unabhängig von der Häufigkeit des gezeigten bzw. von den Gutachtern erkannten Verhaltensweisen, was sich in besseren Übereinstimmungen ihrer Ergebnisse, vor allem in Subtest 3 sowie der Gesamtergebnisse der Hunde, wiederspiegelt. Diese Auswertungsmethode ist darum im folgenden Gebrauch des Tests auf Futteraggression der Auswertung mithilfe der Punktevergabe vorzuziehen, auch wenn hierbei das komplexe Verhalten der Hunde auf fünf Kategorien reduziert wird.

Die Bestimmung der Intra-rater-Reliabilität ist eine grundlegende Messung der Reliabilität und ist ein wichtiges Qualitätsmerkmal, da sie Voraussetzung für eine hohe Test-Retest-Reliabilität ist. Sie ist wichtig, um Zufallsbefunde des Gutachters auszuschließen und sicher zu stellen, dass derselbe Gutachter zu jeder Zeit das Verhalten objektiv bewertet (TAYLOR und MILLS, 2006; DIEDERICH und GIFFROY, 2006). Erst wenn die Ergebnisse eines Gutachters unabhängig sind vom Zeitpunkt oder der Häufigkeit der Auswertung und erst wenn die Beurteilung von Verhaltensweisen einen gewissen Grad an Stabilität erreicht hat, ist eine faire und objektive Beurteilung von Hunden möglich.

Abgesehen von der mangelhaften Übereinstimmung im Subtest 1 von Gutachter 3 überwiegen die Übereinstimmungen der Gutachter mit sich selbst sowohl bei der Punkte- als auch bei der Kategorieneinteilung des Verhaltens. Darum ist der Test auf Futteraggression als reliabel in Bezug auf die Intra-rater-Reliabilität zu betrachten. Um dennoch Fehlinterpretationen und Unsicherheiten beim Auswerten der Hunde zu vermeiden, sollte Subtest 1, wenn er aufgrund von mäßiger bis starker

Futteraggression in den Subtests 2 und 3 nötig ist, am Ende des Tests durchgeführt werden. Außerdem sollte jeder Gutachter, welcher ihm unbekannte Hunde bewertet, zusätzliche Informationen zu den Hunden erhalten und in seine Bewertung mit einfließen lassen. Denn nur so kann ein komplexes Bild über das Verhalten und den Charakter eines Hundes entstehen, was ein einziger Verhaltenstest nicht ermöglicht.

5.3.2.3 Die Test-Retest-Reliabilität des Tests auf Futteraggression

Sowohl bei Subtest 1 und 3, als auch bei den Gesamtergebnissen beider Testdurchläufe wurde kein signifikanter Unterschied zwischen den Ergebnissen festgestellt, weshalb der Test als wiederholbar und reliabel einzustufen ist. Dabei fiel auf, dass in den Subtests 1 und 3 die Auswertung mithilfe der Kategorien erneut die höheren p-Werte für die Übereinstimmung erreichte (jeweils p=1,0000) als die Auswertung mithilfe der Punktevergabe (p=0,0781 und p=0,8794). Auch hier kann als Ursache die bereits ausgeführte hohe Empfindlichkeit der Punkteauswertung hinsichtlich der Häufigkeit von Verhaltensweisen herangezogen werden. Zeigte ein Hund also im zweiten Test zwar dieselben Verhaltensweisen aus dem Ethogramm, aber davon weniger als im ersten Test, erreichte er im zweiten Test eine geringere Punktzahl. Dagegen ist die kategoriale Auswertung unabhängig von der Häufigkeit des gezeigten Verhaltens, sodass in beiden Tests dieselben Verhaltensweisen immer zu denselben Kategorienergebnissen führen. Aus diesem Grund ist die kategoriale Auswertung des Verhaltens der Hunde im folgenden Gebrauch des Tests auf Futteraggression vorzuziehen.

Im Gesamtergebnis unterschieden sich die p-Werte für die verschiedenen Auswertungsmethoden nur gering (p=0,8315 bei der Punktevergabe und p=0,7500 bei der Kategorieneinteilung).

Weiterhin fiel auf, dass in Subtest 2 sowohl die Auswertung mithilfe von Punkten als auch die Einteilung in Kategorien zu signifikanten Unterschieden zwischen beiden Testdurchläufen geführt hat (jeweils p=0,0313), wobei die Ergebnisse in Subtest 2 des zweiten Durchlaufs höher ausfielen als im ersten Testdurchlauf (durchschnittliche Abweichung der Punkteauswertung +1,19 bzw. durchschnittliche Abweichung der Kategorien +0,5). Dies kann darauf zurückzuführen sein, dass ein gewisser Lern- und Sensibilisierungseffekt beim ersten Testdurchlauf der Hunde stattgefunden hat, weshalb sie im zweiten Testdurchlauf eher defensiv auf die Testsituation reagierten.

In der Literatur sind Lerneffekte bei der Bestimmung der Test-Retest-Reliabilität bereits bekannt und von FEDDERSEN-PETERSEN (2004), DIEDERICH und GIFFROY (2006), TAYLOR und MILLS (2006) sowie MAZUR (2012) beschrieben worden. Sowohl Habituation[10] als auch Sensibilisierung[11] können hierbei als Ursachen für geringe Übereinstimmung bzw. signifikante Unterschiede zwischen erstem und zweitem Testdurchlauf verantwortlich sein und lassen sich in keinem Verhaltenstest gänzlich vermeiden, da die im ersten Test unbekannten Reize im zweiten Test bereits bekannt sind (SVARTBERG et al., 2005; TAYLOR und MILLS, 2006). Auch können viele andere Faktoren auf das Hundeverhalten Einfluss nehmen, z.B. Stress, Hunger oder Krankheit. Ebenso kann erfolgreiches Verhalten im ersten Test zu einem veränderten erlernten Verhalten im zweiten Test führen (TAYLOR und MILLS, 2006). Dennoch ist es wichtig, in einem Verhaltenstest wiederholbare Ergebnisse zu produzieren, da Beständigkeit und Wiederholbarkeit von Verhalten dem Sinn eines Verhaltens- oder Wesenstests zugrunde liegen. Nur so lässt sich eine Prognose oder Vorhersage treffen, wie sich ein Hund bei einem bestimmten Reiz in Zukunft verhalten wird. Damit nimmt die Test-Retest-Reliabilität eine besondere Rolle unter den Qualitätsmerkmalen von Verhaltenstests ein. Sie spiegelt in hohem Maße den Sinn solcher Tests wieder und ist eines der entscheidendsten Kriterien bei der Beurteilung von Verhaltenstests (SVARTBERG et al., 2005; TAYLOR und MILLS, 2006). Im Falle des Tests auf Futteraggression stimmten bis auf Subtest 2 alle erhobenen Werte überein. Weil mit Abweichungen aufgrund von Lernerfahrungen zu rechnen war, ist dies ein gutes Ergebnis für einen Verhaltenstest, sodass der Test als reliabel im Hinblick auf die Retest-Reliabilität zu betrachten ist.

5.3.3 Die Validität des Tests auf Futteraggression

5.3.3.1 Die Inhaltsvalidität des Tests auf Futteraggression

Die Inhaltsvalidität nimmt keinen Zahlwert an, sondern wird von unabhängigen Fachleuten bewertet (DIEDERICH und GIFFROY, 2006). Diese entscheiden, ob die Methodik des Tests logisch sinnvoll erscheint, um die gegebene Fragestellung zu beantworten. Aufgrund des Reliabilitäts-Validitäts-Dilemmas, welches in Kapitel 2.3.3 bereits angesprochen wurde, kann es jedoch unter Umständen schwierig sein, einen

[10] =Reaktionen auf einen wiederholten Reiz nehmen ab
[11] = Reaktionen auf einen wiederholten Reiz nehmen zu

inhaltsvaliden Test zu entwickeln. Demnach geht eine Optimierung der Reliabilität eines Tests immer zu Lasten der Validität, insbesondere der Inhaltsvalidität, da durch strenge Standardisierung von Vorgehensweisen der Bezug zur Realität mehr und mehr verloren geht. So kann eine Steigerung der Messgenauigkeit dazu führen, dass ein Test etwas völlig anderes misst als das, was er messen soll (ROST, 2004).

Im Rahmen des Tests auf Futteraggression bedeutet dies, dass die Fütterungssituation für die Hunde im Test, angebunden an einer zwei Meter langen Leine, in einem Auslauf und von einer fremden Person mit stockähnlichem Arm, ungewöhnlich und nicht alltäglich ist. Deshalb kann sich ihr Verhalten im Test unter Umständen von ihrer Reaktion in komplexeren Umweltsituationen unterscheiden.

Um dennoch ein möglichst hohes Maß an Inhaltsvalidität zu gewährleisten, wurden mehrere Maßnahmen ergriffen:

Zum einen wurden jene Hunde mithilfe des Datenerhebungsbogens vom Test auf Futteraggression ausgeschlossen, welche laut Vorerfahrungen der Tierheimmitarbeiter auch ohne Anwesenheit von Futter aggressiv oder ängstlich gegenüber der Testperson reagieren könnten. So wurde ausgeschlossen, dass andere als durch die Verteidigung von Futter motivierte Aggressionen oder Furcht für das Verhalten der Hunde im Test verantwortlich war.

Denselben Zweck erfüllten zum anderen die beiden Prätests, deren Hauptziel es war, die Unterscheidungsfähigkeit der Hunde zwischen echter und künstlicher Hand zu untersuchen. Sie stellten aber zusätzlich auch sicher, dass die getesteten Hunde nicht schon allein aufgrund der für sie ungewöhnlichen Testbedingungen aggressiv oder ängstlich reagierten. In diesem Fall wären die entsprechenden Hunde ebenfalls vom anschließenden Test auf Futteraggression ausgeschlossen worden.

Eine dritte Maßnahme zur Wahrung der Inhaltsvalidität war die Allgemeinuntersuchung der getesteten Hunde am Ende des Tests, welche Schmerzzustände der Tiere als Ursache für gesteigerte Reaktionsbereitschaft und Aggressivität ausschließen sollte. Sie wurde bewusst am Ende des Tests durchgeführt, um den Stress der Tiere vor dem Test nicht durch körperliche Manipulation zu erhöhen und eventuell Angstreaktionen vor der Testperson zu provozieren.

Aus diesen Gründen wird für den Test auf Futteraggression eine ausreichende Inhaltsvalidität angenommen.

5.3.3.2 Die Kriteriumsvalidität des Tests auf Futteraggression

Da es zur Ermittlung von Futteraggression bis zum Zeitpunkt dieser Studie keinen gesonderten und evaluierten Verhaltenstest gab und somit ein Goldstandard nicht zur Verfügung stand, wurden für die Ermittlung der Kriteriumsvalidität die Ergebnisse des Tests auf Futteraggression mit mehreren externen Messungen verglichen: dem Datenerhebungsbogen, einem Referenztest und dem C-BARQ-Fragebogen von HSU und SERPELL (2003).

- **Vergleich mit dem Datenerhebungsbogen**

Beim Verglich der Ergebnisse des Tests auf Futteraggression mit den Angaben der Tierheimleitung im Datenerhebungsbogen (Anhang 9.1) über das Verhalten der Hunde stellte sich heraus, dass der Vergleich der Kategorien von Subtest 1 mit Frage 1 (Verhalten bei leerer Futterschüssel) hochsignifikante Unterschiede (p=0,0002) aufwies. Auch der Vergleich von Subtest 2 mit Frage 2 (Verhalten bei normalem Futter) ergab signifikante Unterschiede (p=0,0430). Dagegen ergab der Vergleich von Subtest 3 mit Frage 3 (Verhalten bei besonderem Futter) eine vollständige Übereinstimmung (p=1,0000) und der Vergleich der Gesamtkategorien eine Übereinstimmung (p=0,6406) der Einschätzung der Tierheimleitung mit den Ergebnissen des Tests auf Futteraggression. Als Grund für die schlechte Übereinstimmung von Subtest 1 und 2 mit der Einschätzung der Tierheimleitung kann in Betracht gezogen werden, dass in einem Tierheim nicht regelmäßig das Maß an futterverteidigendem Verhalten bei leerer Futterschüssel oder normalem Futter überprüft wird. So konnte die Tierheimleitung trotz langer Aufenthaltsdauer der Hunde und engem Verhältnis zu diesen nur schwer einschätzen, wie sich die Hunde in diesen beiden Situationen verhalten würden. Da sie außerdem das Verhalten der Hunde in diesen Situationen nicht für die Beantwortung des Datenerhebungsbogens austesten sollte, schätzte die Tierheimleitung vielmehr die Reaktion der Hunde anhand ihrer Vorerfahrungen mit den Tieren ein. Die Frage nach dem Verhalten in Anwesenheit von besonderem Futter (Subtest 3) wurde mit guter Übereinstimmung zu den Ergebnissen des Tests beantwortet, was auch zu einer guten Übereinstimmung der Gesamtkategorien führt. Diese gute Übereinstimmung ist darauf zurück zu führen, dass die Tierheimleitung das Verhalten der Hunde in Anwesenheit von besonderem Futter gut vorhersagen konnte, da sie diese Situation regelmäßig im Alltag mit ihren Hunden testete und demnach Erfahrungswerte einbeziehen konnte.

Generell betrachtet, werden solche Besitzerfragebögen in der Literatur relativ häufig eingesetzt, um das Verhalten von Hunden zu ermitteln (VAN DER BORG et al., 1991; NETTO und PLANTA, 1997; LEDGER und BAXTER, 1997; VAN DEN BERG et al., 2003; PLANTA und DE MEESTER, 2007; BOLLEN und HOROWITZ, 2008; VAN DER BORG et al. 2010; VALSECCHI et al., 2011; BENNETT et al., 2012). Hierbei wird meistens nach aggressivem Verhalten der Hunde in der Vergangenheit gefragt oder aber nach Problemverhalten, sobald der Hund von einem neuen Besitzer adoptiert wurde. LEDGER und BAXTER (1997) fanden Übereinstimmungen zwischen den Angaben der Besitzer und ihrem Verhaltenstest, ebenso SVARTBERG (2005). Jedoch ist die Validität von Besitzerfragebögen anzuzweifeln, da auch diese im Vorfeld evaluiert werden müssten (VAN DER BORG et al., 1991; TAYLOR und MILLS, 2006). VAN DER BORG et al. (1991) merkten außerdem an, dass Besitzer gerade bei der Frage nach Problemverhalten wenig aussagekräftige Angaben machen, da sie das Problemverhalten nicht als solches erkennen oder es selbst nicht als Problem einschätzen und deshalb nicht im Fragebogen angeben. Zu einem ähnlichen Ergebnis gelangten MARDER et al. (2013), welche die Ergebnisse eines Futteraggressionstest in einem Tierheim mit der Einschätzung der neuen Besitzer über dieses Problemverhalten verglichen. So hängt die Übereinstimmung von Verhaltenstest und Besitzerfragebögen erheblich vom Besitzer selbst ab, von seinem Wissenstand und seinen Erfahrungen über und mit dem Hund und von seinem eigenen Charakter (TAYLOR und MILLS, 2006).

Auch beim vorliegenden Test auf Futteraggression zeigte sich, dass die Erfahrungswerte mit den Hunden bei der Beantwortung eines Fragebogens eine entscheidende Rolle spielen. Obwohl die Kooperationsbereitschaft der Tierheimleitung bei der Mithilfe in dieser Studie sehr groß, die Hunde schon relativ lange im Tierheim und das Verhältnis zu den Hunden sehr gut waren, wurde nur eine von drei Fragen in Übereinstimmung zu den Testergebnissen beantwortet. Ratsam wäre deshalb, im alltäglichen Umgang mit den Hunden auch zumindest die Situation einer Annäherung an das normale Futter der Hunde bei einer Mahlzeit auszutesten, um ein breiteres Bild vom Verhalten der Hunde zu erhalten.

Als weiterer Grund für die Abweichungen kann auch die Kategorieneinteilung in Betracht gezogen werden, in die das Verhalten der Hunde eingeteilt wurde. Bei einer Überprüfung der Validität mithilfe der beobachteten bzw. vermuteten Verhaltensweisen sollte in diesem Fall eine Übereinstimmung zu errechnen sein.

- **Vergleich mit dem Referenztest**

Beim Vergleich der Ergebnisse des Tests auf Futteraggression mit dem von der Tierheimleitung durchgeführten Referenztest ergaben sich für die Auswertung mit Punkten signifikante Unterschiede zwischen Subtest 1 und Teil 1 des Referenztests (p=0,0313) sowie hochsignifikante Unterschiede zwischen den Gesamtpunkt-ergebnissen (p=0,0020). Dabei nahmen die Werte des Referenztests insgesamt niedrigere Punktwerte an als im Test auf Futteraggression, (Subtest 1 durchschnittlich -2,0 Punkte, Gesamtpunkte durchschnittlich -2,7 Punkte). Dagegen erzielte die Auswertung mithilfe der Punkte gute Übereinstimmungen zwischen Subtest 2 und 3 und den Teilen 2 und 3 des Referenztests (p=0,2188 bzw. p=0,5313). Als Ursache für die signifikanten Unterschiede kann erneut die hohe Empfindlichkeit der Punkteauswertung bezüglich der Häufigkeiten der gezeigten Verhaltensweisen und deren Einfluss auf die Punktzahlen in Betracht gezogen werden. Denn bei der Auswertung mithilfe der Kategorieneinteilung erreichte der Vergleich zwischen dem Test auf Futteraggression und dem Referenztest deutlich bessere Übereinstimmungen ohne signifikante Unterschiede (p-Werte zwischen 0,1309 und 1,0000), weshalb diese Art der Bewertung des Verhaltens als valide anzusehen und im folgenden Gebrauch des Tests auf Futteraggression zu bevorzugen ist. Ein Vergleich der Häufigkeiten der einzelnen Verhaltensweisen könnte darüber Aufschluss geben, ob die Punktevergabe für die Abweichungen verantwortlich ist, wobei berücksichtigt werden müsste, dass der Referenztest vier Testsituationen weniger umfasst als der Test auf Futteraggression.

Als weitere Ursachen für signifikante Unterschiede zwischen den Ergebnissen der Tests können aber auch die unterschiedlichen Testbedingungen in Betracht gezogen werden. So wurde der Referenztest nicht auf einem standardisierten Testgelände durchgeführt, sondern in den Zwingern der Hunde, welche den Hunden sehr vertraut waren, ebenso wie die Testperson selbst, welche beim Referenztest die Tierheimleitung selbst war. Aus diesem Grund könnten die Hunde im ersten Teil des Referenztest weniger reaktiv auf die Bedrohung der sich annähernden Person reagiert haben. Gegen diese Theorie spricht die kaum veränderte Anzahl der Stresssignale in diesem Teil des Tests, welche im Referenztest um 0,28 Signale zunahm.

- **Vergleich mit dem C-BARQ**

Es zeigte sich eine positive doch nicht signifikante Korrelation zwischen den Ergebnissen des Tests auf Futteraggression und den Kategorien "stranger-directed aggression" (Punkte: r=0,353, p=0,1647; Kategorien: r=0,346, p=0,1734) und "owner-directed aggression" (Punkte: r=0,262, p=0,3103; Kategorien: r=0,152, p=0,5598). Dem gegenüber steht die negative aber ebenfalls nicht signifikante Korrelation der Ergebnisse des Tests auf Futteraggression mit den C-BARQ-Werten für die Kategorien "stranger-directed fear" (Punkte: r= -0,120, p=0,6477; Kategorien: r= -0,215, p=0,4069) und "touch sensitivity" (Punkte: r= -0,220, p=0,3959; Kategorien: r= -0,426, p=0,0881).

Als Grund für die nicht signifikanten Korrelationen zwischen dem Test auf Futteraggression und den C-BARQ-Ergebnissen der getesteten Hunde könnte der geringe Stichprobenumfang der Studie in Betracht gezogen werden, wonach 17 Hunde nicht ausreichten, um eine signifikante Korrelation zu berechnen. Hier wäre eine Poweranalyse nötig gewesen, um zu prüfen, ob der Stichprobenumfang von 17 Hunden ausreicht, um eine Korrelation nachzuweisen. Deshalb sollte in weiterführenden Studien zu diesem Thema der nötige Stichprobenumfang mittels Poweranalyse ermittelt und dann die Anzahl der getesteten Hunde darauf ausgerichtet werden, was möglich wird, wenn dann nicht mehr nur Hunde derselben Rasse getestet werden.

Ein weiterer Grund für die nicht signifikanten Korrelationen mit den C-BARQ-Ergebnissen könnten fehlerhafte Angaben und Einschätzungen der Tierheimleitung im Fragebogen sein, welche zu verzerrten Punktergebnissen des C-BARQs führen würde. Wie oben schon erläutert, sind die Angaben in Fragebögen immer abhängig von den Erfahrungen und der Einschätzungsfähigkeit der ausfüllenden Person, in diesem Fall von der Fähigkeit der Tierheimleitung, das Verhalten der Hunde richtig zu beurteilen und vorherzusagen. Des Weiteren muss in Betracht gezogen werden, dass der C-BARQ nicht geeignet ist, das Verhalten von Tierheimhunden zu beurteilen. Er wurde von HSU und SERPELL (2003) nur für Hunde aus Privathaltungen evaluiert, da es in Tierheimen in der Regel nicht möglich ist, die große Anzahl an Fragen über das Verhalten der Hunde korrekt zu beantworten. Da jedoch die Beziehung der Tierheimleitung zu den getesteten Hunden und damit auch die Einschätzungsfähigkeit der Tierheimleitung zum Verhalten der Hunde als sehr gut und mit einem privaten Mehrhundehaushalt vergleichbar anzusehen war, wurde diese Tatsache in der

vorliegenden Studie bewusst nicht berücksichtigt. Es gilt damit in weiterführenden Studien zu prüfen, ob ein zu geringer Stichprobenumfang oder die Ergebnisse des C-BARQs selbst für die nicht signifikanten Korrelationen mit den Ergebnissen des Tests auf Futteraggression verantwortlich sind.

5.4 Die Stresssignale

Bei der Betrachtung der Stresssignale während des Tests auf Futteraggression fiel auf, dass die Hunde in Subtest 1 am häufigsten Stresssignale zeigten (53 Stresssignale bei Gutachter 1), gefolgt von Subtest 2 (16 Stresssignale bei Gutachter 1) und Subtest 3 (3 Stresssignale bei Gutachter 1). Dies kann darauf zurückzuführen sein, dass die Hunde in Subtest 1 wenig Interesse an der leeren Futterschüssel zeigten und somit mehr Zeit hatten für die Auseinandersetzung mit den Testbedingungen. Diese waren allen Hunden unbekannt, sodass es zu vermehrten Stresssignalen kam. In Subtest 2 und 3 dagegen waren die Hunde mit einer möglichst schnellen Futteraufnahme beschäftigt, sodass eventuell wenig Zeit und Möglichkeit für Stresssignale blieb. Zudem könnte auch hier bereits eine Habituation an die Testbedingungen und die Testperson stattgefunden haben, weshalb die Hunde weniger Stresssignale zeigten.

Bei der Korrelation der Anzahl der Stresssignale mit den Punkte- und Kategorienergebnissen des Tests auf Futteraggression lag keine Signifikanz vor. Das heißt, dass die Häufigkeit der Stresssignale in Subtest 1 nicht mit den Ergebnissen der Punkte- und Kategorienauswertung von Subtest 1 korreliert.

In Subtest 2 korrelierte die Anzahl an Stresssignalen ebenfalls schwach positiv und ohne Signifikanz mit den Punkteergebnissen ($r=0,260$, $p=0,3304$), aber signifikant positiv mit den Kategorienergebnissen ($r=0,510$, $p=0,0434$), was vermuten lässt, dass ein erhöhtes Maß an Stress mitverantwortlich ist für die Kategorienergebnisse.

In Subtest 3 korrelierte die Anzahl der Stresssignale schwach positiv und ohne Signifikanz mit den Punkteergebnissen ($r=0,371$, $p=0,1432$) und den Kategorienergebnissen ($r=0,169$, $p=0,5175$). Hier ist ein erhöhter Stresslevel nicht mit einem hohen Testergebnis verbunden.

Das Gesamtergebnis der Punkte- und Kategorienauswertung der einzelnen Hunde korreliert dagegen signifikant positiv mit der Gesamtanzahl der Stresssignale der Hunde (Punkte: $r=0,538$, $p=0,0259$; Kategorien $r=0,547$, $p=0,0232$). Diese Ergebnisse

lassen mutmaßen, dass ein erhöhtes Maß an Stress bei den Hunden zu höheren Gesamttestergebnissen führt.

Diese signifikanten Korrelationsergebnisse von Subtest 2 (Kategorien) sowie der Gesamtergebnisse sind nicht verwunderlich, da in Verhaltenstests Hunde absichtlich verschiedenen optischen, akustischen und olfaktorischen Stimuli und Stressoren ausgesetzt werden, um (aggressives) Verhalten zu provozieren (FEDDERSEN-PETERSEN, 2004). Eine positive Korrelation von Stress mit den Testergebnissen ist demnach zu erwarten und in gewissem Maß auch erwünscht, solange nicht der Stress allein ohne bestimmte Stimuli aggressives Verhalten und damit falsch-positive Testergebnisse verursacht (MARDER et al., 2013). Unter anderem, um diesen Sachverhalt auszuschließen, wurde der Prätest durchgeführt. Mit seiner Hilfe wurden Hunde erkannt und vom Test ausgeschlossen, die bereits im Vorfeld ohne Anwesenheit von Futter und allein aufgrund der Testbedingungen aggressives oder ängstliches Verhalten zeigten. Da keiner der getesteten Hunde im Prätest ein solches Verhalten zeigte, kann davon ausgegangen werden, dass bei der Entwicklung des Tests auf Futteraggression für Subtest 2 und 3 die passenden Stimuli bzw. Stressoren gewählt wurden, um futterverteidigendes Verhalten bei den Hunden auszulösen.

Die gefundenen positiven Korrelationen zwischen Testergebnissen und Stresssignalen sind jedoch nur dann von relevanter Bedeutung, wenn vorausgesetzt wird, dass das Zeigen von Stresssignalen vom Hund tatsächlich mit einem erhöhten Stressempfinden einhergeht. Dies wurde in dieser Studie angenommen, aber nicht bewiesen. In der Literatur werden zunehmend Messungen des sogenannten Stresshormons Cortisol im Speichel der Tiere durchgeführt, um das Maß an Stress genau beurteilen zu können (EL-GHOUL und HOFMANN, 2002; SCHULZ et al., 2005; ERLER, 2010 und STUCKE, 2012). Aufgrund des hohen Aufwands einer solchen Messung und des relativ geringen erwarteten Erkenntniszuwachses für diese Studie wurde hier darauf verzichtet und stattdessen die körperlichen Signale der Hunde als Maß für Stressempfinden herangezogen. Ob jedoch wirklich das Zeigen von Stresssignalen mit einem erhöhten Stressempfinden der Tiere einhergeht, gilt in weiterführenden Studien zu beweisen.

Da in dieser Studie nur in geringem Maße Korrelationen zwischen Testergebnissen und Stresssignalen gefunden wurden, kann angenommen werden, dass diese Korrelationen zufällig und nicht beweisend dafür sind, dass Stress mit einem erhöhten Testergebnis einhergeht. Möglicherweise wurde aber auch mit den Stresssignalen

nicht das passende Maß gewählt, um das Stressempfinden der Hunde zu beurteilen. Darum sollten die gefundenen Korrelationen zwischen den Kategorienergebnissen in Subtest 2 und den Gesamtergebnissen mit der Anzahl der Stresssignale der Hunde nicht überbewertet werden.

5.5 Praktikabilität und Tierschutz

Aus Gründen der Praktikabilität wurde der Test auf Futteraggression so entwickelt, dass er ohne großen Aufwand und mit möglichst wenigen Hilfsmitteln von einer einzigen Person durchzuführen war. So konnten Kosten und Mühen von Hilfspersonen und teurem Equipment eingespart werden, was den Test für Tierheime leicht durchführbar machen sollte (TAYLOR und MILLS, 2006). Der Test wurde außerdem kurz gehalten, um so wenig wie möglich im normalen Tierheimalltag zu interferieren und um möglichst stressarm für die Tiere zu sein, da Stress die Reaktionsbereitschaft von Hunden zu aggressivem Verhalten heraufsetzen kann (VAN DER BORG et al., 2010). Durch eine umfassende Vorselektion wurden nur Hunde zum Test zugelassen, die physisch und psychisch stabil genug waren, den Test vollständig zu absolvieren und genügend auswertbare Daten zu liefern. So wurden Leiden und Schäden der Hunde durch den Test auf ein Minimum reduziert und konnten sowohl während als auch nach dem Test nicht festgestellt werden.

Im Verlauf des Tests stellte sich heraus, dass dieser gut von nur einer Testperson durchzuführen war, solange die Hunde mit ihnen unbekannten Personen umgänglich waren. Sollten jedoch trotz der Gefahr einer Verzerrung der Ergebnisse Hunde getestet werden müssen, welche aggressiv oder ängstlich gegenüber fremden Personen sind, könnte eine Hilfsperson von Seiten des Tierheims notwendig werden, welche mit den Hunden vertraut ist. Diese müsste dann zumindest die Vorbereitungsphase des Tests inkl. dem Ausführen und Festbinden der Hunde sowie die Nachbereitungsphase mit abermaligem Ausführen und Zurückbringen in die Zwinger übernehmen, um den Stress der Hunde im Umgang mit einer fremden Person und um das Verletzungsrisiko der Testperson durch die Hunde zu reduzieren. Die emotionale Situation der Hunde sowie die Managementmöglichkeiten des Tierheims sollten hierbei aber immer individuell bewertet und vor dem Test abgewogen werden.

5.6 Schlussfolgerung

Der in dieser Studie entwickelte Test auf Futteraggression hat sich als geeignet erwiesen, um futterverteidigendes Verhalten von Tierheimhunden gegenüber Menschen zu erkennen. Der Test ist in dieser Pilotstudie mit 17 Bullterriern eines Tierheims bis auf wenige Ausnahmen reliabel und valide.

Die zur Studie verwendeten Hunde zeigten aufgrund der Vorselektion auf Furcht und Aggressivität gegenüber der Testperson sowie Gesundheit ein weitestgehend geringes Maß an Futteraggression während des Tests. Diese Vorselektion sowie die Beschränkung auf nur eine Rasse von Hunden waren zwar für diese Pilotstudie notwendig zum Zwecke der Evaluierung, sind jedoch im Alltag eines Tierheims mit gemischter Hundepopulation unüblich und nicht praktikabel. So sollte in weiterführenden Arbeiten mit einer größeren Anzahl von Hunden unterschiedlicher Rasse und Herkunft getestet werden, um den Test auf Futteraggression auch innerhalb einer breiteren Hundepopulation und mit einem größeren Stichprobenumfang auf seine Zuverlässigkeit und Gültigkeit zu prüfen (DIEDERICH und GIFFROY, 2006). Eine Poweranalyse sollte dabei durchgeführt werden, um den notwendigen Stichprobenumfang zu ermitteln.

Die im Test verwendeten Hilfsmittel sowie die gewählte Kameraposition erwiesen sich als gut geeignet. Vor allem die künstliche Hand zeigte sich als nützlich und ihre modifizierte Handhabung verhinderte ein verzerrtes Ergebnis aufgrund von Furcht, Aggressivität oder Stress. Dennoch stellte sich im durchgeführten Prätest heraus, dass einige der Hunde durchaus in der Lage waren, die künstliche Hand von der echten Hand der Testperson zu unterscheiden, was sich in einer signifikanten Zunahme von soziopositiven Verhaltensweisen ausdrückte. Dieses Ergebnis stellt die Benutzung der künstlichen Hand in Verhaltenstest, vor allem mit ihrer üblichen Handhabung als verlängerten Arm, in Frage. Daher sollten weiterführende Untersuchungen bezüglich der Verwendung von künstlichen Händen oder anderen Gegenständen in Verhaltenstests und ihren Einfluss auf das Verhalten der Hunde durchgeführt werden.

Die Aufnahme der Stresssignale im Prätest war in dieser Pilotstudie hilfreich, um festzustellen, wie stressreich die Testbedingungen auf die Hunde wirkten. Im Test auf Futteraggression zeigte sich der Subtest 1 als der stressreichste Subtest für die Hunde. Zwar korrelierte die hohe Anzahl an Stresssignalen nicht signifikant mit einem hohen Punkte- bzw. Kategorienergebnis in Subtest 1, dennoch löste dieser Subtest

auch andere Aggressionsarten aus bzw. wurde Hundeverhalten zum Teil fehlinterpretiert. Seine Ergebnisse sind daher vorsichtig zu bewerten. Vorstellbar wäre, die Durchführung von Subtest 1 abhängig von den Signalen der Hunde in den beiden anderen Subtests zu machen und ihn damit ans Ende des Testablaufes zu verschieben. Zwar konnte weiterhin eine positive signifikante Korrelation zwischen der Anzahl der Stresssignale und den Gesamtergebnissen der Hunde errechnet werden, dennoch gab es keine weiteren Korrelationen zwischen den Stresssignalen und den Testergebnissen. Darum wird angenommen, dass nicht ein erhöhtes Maß an Stress für hohe Testergebnisse verantwortlich ist. Inwiefern jedoch das Zeigen von Stresssignalen mit von den Hunden empfundenem Stress einhergeht, sollte in einer weiterführenden Studie untersucht werden und wurde in dieser Studie nur angenommen.

Beim Vergleich der Bewertungssystematiken erwies sich die Einteilung des Verhaltens in fünf Kategorien als ungenauer, aber deutlich reliabler und valider als die Bewertung des Verhaltens mithilfe von Punkten. So erzielte die Punktevergabe bei der Bestimmung der Inter- und Intra-rater-Reliabilität zum Teil deutlich schlechtere Übereinstimmungen als die Auswertung mithilfe der Kategorien. Die Verwendung der Kategorien ist damit geeigneter, ein ungefähres Bild über die Futteraggression vieler zu testenden Hunde zu gewinnen, z.B. in einem Tierheim, wo viele Hunde zu testen sind, wenig Zeit für die Auswertung ist und verschiedene Personen das Verhalten bewerten. Um im Rahmen einer Verhaltenstherapie ein genaues Maß zu erhalten und Fortschritte desselben Hundes durch denselben Gutachter zu erkennen, ist dagegen die Auswertungsmethode mit Punkten besser geeignet, um auch kleine Tendenzen und Veränderungen des Verhaltens genau erfassen zu können.

Es stellte sich heraus, dass eine genaue Einweisung aller Gutachter in die Auswertungsmethodik und das Ethogramm von Vorteil ist und die Reliabilität des Tests möglicherweise erhöhen könnte. Die zum Teil schlechteren Ergebnisse der Test-Retest-Reliabilität sind dagegen auf Lerneffekte von Seiten der Hunde zurückzuführen, welche auch bei anderen Studien zu beobachten sind (DIEDERICH und GIFFROY (2006), TAYLOR und MILLS (2006); MAZUR (2012).

Der Test stellte sich weiterhin als inhaltlich valide heraus. Im Vergleich mit einem Referenztest erreichte die kategoriale Auswertung deutlich bessere Übereinstimmungen ohne signifikante Unterschiede als die Auswertung mithilfe von Punkten und ist daher als valide anzusehen. Signifikante Unterschiede der Ergebnisse

zur Einschätzung des Verhaltens von Seiten der Tierheimmitarbeiter im Datenerhebungsbogen ergaben sich lediglich in Subtest 1 und 2. Dies kann darauf zurückzuführen sein, dass die Mitarbeiter des Tierheims nur Situationen mit besonders schmackhaftem Futter im Alltag mit den Hunden prüfen, jedoch nicht mit normalem Futter oder leerer Futterschüssel, was zu ungenauen Einschätzungen des Verhaltens der Hunde in diesen Situationen führte. Ebenso könnte ein zu geringer Stichprobenumfang verantwortlich sein für die nicht signifikanten Korrelationen mit dem C-BARQ-Fragebogen von HSU und SERPELL (2003), bzw. könnte der C-BARQ für Tierheimhunde ungeeignet sein.

So erwies sich der Test auf Futteraggression mit kategorialer Auswertung des Hundeverhaltens in dieser Pilotstudie als ein relativ praktikabler, reliabler und valider Verhaltenstest, um futterverteidigendes Verhalten von Tierheimhunden gegenüber sich annähernden Menschen tierschutzkonform zu messen, während sich die Auswertungsmethode mit Punkten als weniger reliabel und valide herausstellte. Die Evaluierung von Subtest 1 zeigte einige Schwachpunkte dieses Subtests auf, weshalb seine Verwendung abhängig vom Verhalten der Hunde in Subtest 2 und 3 gemacht werden sollte und damit ans Ende des Tests auf Futteraggression verschoben werden sollte.

6 Zusammenfassung

Im Rahmen dieser Studie sollte ein Verhaltenstest entwickelt und evaluiert werden, welcher es ermöglicht, agonistisches Verhalten von Hunden gegenüber sich annähernden Personen während der Futteraufnahme praxisnah zu messen. Hierfür wurden 17 Hunde aus dem Tierheim "Bullterrier in Not e.V." ausgewählt und ihr Verhalten während der Futteraufnahme gegenüber einer sich annähernden Testperson sowie der Berührung der Hunde und des Futters mit einer künstlichen Hand mittels einer Videokamera aufgezeichnet. Anschließend wurde das gezeigte Verhalten mit zwei unterschiedlichen Bewertungssystemen bewertet und diese wurden auf Reliabilität und Validität hin evaluiert.

Durch einen Prätest wurden Furcht und Aggressivität der Hunde gegenüber der Testperson und den Testbedingungen, inklusive der verwendeten künstlichen Hand, ausgeschlossen. Des Weiteren wurde in diesem Prätest überprüft, ob die Hunde in der Lage waren, die verwendete künstliche Hand von der echten Hand der Testperson zu unterscheiden. Die Ergebnisse lassen den Schluss zu, dass einige der Hunde in der Lage waren, die künstliche Hand von der Echten zu unterscheiden. So wurde gegenüber der künstlichen Hand signifikant häufiger ($p=0,0413$) soziopositives Verhalten (Erkundungs- und Spielverhalten) gezeigt. Da jedoch keine signifikanten Unterschiede bei den agonistischen Verhaltensweisen ($p=0,1183$) und den Stresssignalen ($p=0,5691$) ermittelt werden konnten, war eine Beeinflussung der Testergebnisse auf Futteraggression, wo agonistisches Verhalten und Stresssignale gemessen wurden, durch die künstliche Hand ausgeschlossen.

Bei der Durchführung des entwickelten Tests auf Futteraggression wurde den Hunden unterschiedliches Futter in drei Subtests (Subtest 1: leere Futterschüssel, Subtest 2: normales Futter, Subtest 3: Rinderpansen) gereicht, während sich die Testperson an die Hunde annähert und die Hunde und das Futter mit der künstlichen Hand berührt. Das Verhalten der Hunde wurde mit einer Videokamera aufgezeichnet und von drei Personen (Testperson und zwei weitere Gutachter) mit Hilfe eines Ethogramms auf Stresssignale und agonistisches Verhalten hin untersucht. Anschließend wurde das beobachtete Verhalten der Hunde mittels zweier Bewertungssysteme bewertet. Zum

einen wurde eine Punktevergabe verwendet, bei welcher die Punktwerte mit steigendem Eskalationsniveau des aggressiven Verhaltens zunahmen. Zum anderen wurde das Verhalten in fünf Kategorien eingeteilt. Mit diesen zwei Bewertungssystemen wurde der Test anschließend auf Reliabilität (Inter-rater-Reliabilität, Intra-rater-Reliabilität, Test-Retest-Reliabilität) und Validität (Inhalts- und Kriteriumsvalidität) evaluiert.

Die Inter-rater-Reliabilität des entwickelten Tests wurde anhand der Auswertungen der drei Gutachter mittels des Permutationstests für gepaarte Beobachtungen bestimmt. Hier zeigte sich mit einer Ausnahme (Subtest 2, Gutachter 1-3: $\kappa = -0{,}1355$; Gutachter 2-3: $\kappa = 0{,}0204$) eine gute Übereinstimmung der Gutachter bei der Verwendung der kategorialen Auswertung des Verhaltens ($p > 0{,}05$) und ebenso mit einer Ausnahme (Subtest 3, Gutachter 1-3: $p = 0{,}004$; Gutachter 2-3: $p = 0{,}000$) bei der Verwendung der Auswertung mit Punkten ($p > 0{,}05$).

Mit den Daten einer wiederholten Auswertung derselben Videos konnte mittels des Wilcoxon-signed-rank-Tests die Intra-rater-Reliabilität für die Auswertung mit Punkten ermittelt werden. Für die Auswertung mithilfe der Kategorien wurde zusätzlich der Cohens-Kappa-Koeffizient berechnet. Hier zeigte sich eine gute Übereinstimmung bei den Gutachtern 1 und 2 sowohl bei der Punkte- als auch bei der Kategorienauswertung (Gutachter 1 Punkte gesamt: $p = 0{,}4766$; Kategorien gesamt: $p = 1{,}0000$, $\kappa = 0{,}8917$; Gutachter 2 Punkte gesamt: $p = 0{,}5806$; Kategorien gesamt: $p = 1{,}0000$, $\kappa = 0{,}8851$). Die Auswertungen des dritten Gutachters wiesen bei der Verwendung der Punkte zum Teil signifikante Unterschiede (Subtest 1, $p = 0{,}0313$) auf, während die Auswertungen unter Verwendung der Kategorien nicht signifikant voneinander abwichen ($p > 0{,}05$).

Der entwickelte Futteraggressionstest wurde nach acht Wochen mit derselben Hundepopulation erneut durchgeführt (Retest). Die Test-Retest-Reliabilität wurde erneut mithilfe des Wilcoxons-signed-rank-Tests für beide Methoden, und für die Kategorienauswertung zusätzlich mit dem Cohens-Kappa-Koeffizienten, durchgeführt. Hier zeigten sich keine signifikanten Unterschiede in den Subtests 1 und 3 ($p > 0{,}05$), wobei die Auswertung mithilfe der Kategorien bessere Übereinstimmungen aufwies als die Auswertung mithilfe der Punkte. Auch die Gesamtergebnisse der Hunde wiesen keine signifikanten Unterschiede auf ($p > 0{,}05$). Lediglich in Subtest 2 waren signifikante Unterschiede (Punkte und Kategorien, jeweils $p = 0{,}0313$) bei beiden Auswertungsmethoden festzustellen.

Die Inhaltvalidität wurde gewährleistet, indem eine Allgemeinuntersuchung durchgeführt wurde und eine Vorauswahl der Hunde bezüglich Furcht und Aggressivität gegenüber der Testperson mithilfe eines Datenerhebungsbogens getroffen wurde.

Um die Kriteriumsvalidität zu bestimmen, wurde das Verhalten der Hunde im durchgeführten Test mit den Angaben des Tierheims über das erwartete Verhalten mittels des Wilcoxons-signed-rank-Tests (bei der Kategorieneinteilung zusätzlich mit dem Cohens-Kappa-Koeffizienten) verglichen. Hierbei zeigten sich signifikante Abweichungen bei den Subtests 1 ($p=0{,}0002$, $\kappa=0{,}0000$) und 2 ($p=0{,}0430$, $\kappa=0{,}0588$) und eine gute Übereinstimmung ($p=1{,}0000$, $\kappa=0{,}3014$) mit dem Subtest 3.

Sieben Wochen nach dem Retest wurde ein Referenztest, welcher weniger standardisiert war und unter alltäglicheren Bedingungen in den Zwingern der Hunde stattfand, durch die Tierheimleitung durchgeführt. Auch bei diesem Vergleich wurden die Berechnungen mithilfe des Wilcoxons-signed-rank-Tests bzw. Cohens-Kappa-Koeffizienten durchgeführt. Dabei zeigten sich Übereinstimmungen bei der Auswertung mithilfe der Kategorien ($p>0{,}05$), wogegen bei der Auswertung mithilfe der Punkte zum Teil signifikante Unterschiede zu ermitteln waren (Vergleich Teil 1 mit Subtest 1: $p=0{,}0313$; Vergleich Gesamtergebnisse: $p=0{,}0020$). Insgesamt zeigte sich, dass die Bewertungsmethode mit Punkten sehr sensibel auf Abweichungen im Verhalten reagiert, da sie im Gegensatz zu den Kategorien abhängig ist von den Häufigkeiten der gezeigten Verhaltensweisen.

Bei dem Versuch einer Korrelation der Ergebnisse mit dem bereits evaluierten Fragebogen C-BARQ von HSU und SERPELL (2003) mithilfe von Spearmans Korrelationskoeffizienten ergaben sich keine nachweisbaren Zusammenhänge zwischen den Angaben des Tierheims und den ermittelten Testergebnissen ($r<0{,}353$, $p>0{,}05$).

Bei jedem Testdurchgang wurde die Art und Anzahl von Stresssignalen der Hunde aufgezeichnet und ebenfalls mit dem Spearmans-Korrelationskoeffizienten in Korrelation mit den vorherigen Testergebnissen gesetzt. Hierbei fiel auf, dass mit steigender Anzahl an Stresssignalen nur das Testergebnis der Hunde in Subtest 3 bei den Kategorien ($r=0{,}510$, $p=0{,}0434$) und im Gesamtergebnis beider Auswertungsmethoden (Punkte: $r=0{,}538$, $p=0{,}0259$, Kategorien: $r=0{,}547$, $p=0{,}0232$) stieg und damit positiv korrelierte.

Die Ergebnisse der Studie lassen den Schluss zu, dass der entwickelte Test geeignet ist, um Futteraggression von Hunden gegenüber Menschen zu erkennen. Er hat sich bei der Evaluierung, trotz weniger Ausnahmen, als reliabel und valide erwiesen und erfüllt damit zum großen Teil die Ansprüche für die Qualität eines Verhaltenstests. Außerdem hat sich der entwickelte Test in der Durchführung als praktikabel sowie als tierschutzkonform erwiesen. Die Testauswertung mit Punkten erscheint besser geeignet für einen Einsatz im Diagnose- und Therapiebereich, wo einzelne Hunde getestet und im Zuge der Therapie mit sich selbst verglichen werden. Dagegen scheint die Auswertung mit fünf Kategorien besser geeignet zu sein für den Einsatz in einem Tierheim, wo viele Hunde getestet werden und unterschiedliche Gutachter das Verhalten schnell bewerten und interpretieren müssen.

7 Summary

In the context of this study a behavior test should be developed and evaluated, which enables the practice measuring of agonistic behavior of dogs against approaching humans while feeding.

For this 17 Bullterrier from the shelter “Bullterrier in Not e.V.” were selected and their behavior against the approaching testing person and the touching of the dogs and their food with an artificial hand while feeding was recorded with a camera. The shown behavior was rated with two rating systems, which were evaluated by reliability and validity.

Fear and aggressiveness due to the test conditions inclusive the used artificial hand were excluded by the pretest. Further with this pretest was investigated, if the tested dogs were able to distinguish the used artificial hand from the real hand of the testing person. The results prove the ability of some dogs to distinguish the artificial hand from the real one. So they showed significantly more (p=0.0414) socio-positive behavior (behavior of exploration and play) against the artificial hand. However, since there wasn’t a significant difference in the agonistic behavior (0=0.1183) and the signs of stress (p=0.5691), an influence on the results of the test for food aggression, which measures agonistic behavior and stress, could be excluded.

In carrying out of the developed test there were handed out different food to the dogs in three subtests (subtest 1: empty food bowl, subtest 2: normal food, subtest 3: beef rumen), while the testing person approached to the dogs and touched them and the food. The behavior of the dogs was recorded by a video camera and examined by three persons (the testing person and two further observers) for agonistic behavior and signs of stress with the help of an ethogramm. Then the observed behavior of the dogs was rated with two rating systems. On one hand there was used a scoring system, thus the scores raised with increasing escalation level of aggressive behavior. On the other hand the behavior was split up in a five category system. With the help of these rating systems the test for food aggression was evaluated concerning reliability (inter-rater reliability, intra-rater reliability, test-retest reliability) and validity (content validity and criterion validity).

The inter-rater reliability of the test was determined by analysis of three observers with the help of the permutation test for paired observations. It shows, except from one calculation (subtest 2, observer 1-3: κ= -0.1355; observer 2-3: κ=0.0204), a good congruence of the observers under usage of the categorial analysis (p>0,05) and also with one exception (subtest 3, observer 1-3: p=0.004; observer 2-3: p=0.000) under usage of the scores (p>0.05).

With the data of a repeated analysis of the same videos the intra-rater reliability was calculated. For the calculation of the scores the Wilcoxon signed rank test was used. For the categories the same test, in addition to the Cohen's kappa coefficient, was calculated. It showed a good agreement of observer 1 and 2, as well in the categories as in the scores (observer 1 total scores: p=0.4766; total categories: p=1.0000, κ=0.8917; observer 2 total scores: p=0.5806; total categories: p=1.0000, κ=0.8851). The analysis of the third observer showed partly significant differences under usage of the scores (subtest 1, p=0.0313), while the analysis under usage of the categories did not differ significantly (p>0.05).

The developed test for food related aggression was repeated after eight weeks with the same population of dogs (retest). The test-retest reliability was calculated again with the usage of the Wilcoxon signed rank test for both rating systems, and for the categories in addition of Cohen's kappa coefficient. It showed no significant differences in subtest 1 and 3 (p>0.05), in which the categorical analysis showed better congruence than the analysis with the scores. Also the total results showed no significant differences (p>0.05). Only subtest 2 showed significant differences in both rating systems (scores and categories, respectively p=0.0313).

The content validity was guaranteed by a general examination and a preselection of the dogs with a questionnaire concerning fear and aggressiveness against the testing person.

To determine the criterion validity, the behavior of the dogs in the used test was compared with the data of the expected behavior, recorded by the shelter management, with the Wilcoxon signed rank test (by the rating with categories in addition of Cohen's kappa coefficient). This comparison showed significant differences in subtest 1 (p=0.0002, κ=0.0000) and 2 (p=0.0430, κ=0.0588) and a good agreement in subtest 3 (p=1.0000, κ=0.3014).

Seven weeks after the retest a reference test was accomplished by the shelter management, which was less standardized and conducted under daily conditions in

the kennels of the dogs. For this comparison the calculations were made with the Wilcoxon signed rank test and respectively Cohen's kappa coefficient. It showed an agreement in the analysis with the category system ($p>0.05$), but the analysis with the score system determined partly significant differences (part 1-subtest 1: $p=0.0313$; total score: $p=0.0020$). Overall it showed that the rating systems with scores reacted very sensitive to deviations in the behavior, because in comparison to the categories it depends on the frequency of the showed behavior.
The attempt to correlate the results with the evaluated questionnaire C-BARQ of HSU and SERPELL (2003) with Spearman's correlation coefficient showed no detectable correlations between the data of the shelter and the determined results of the test ($r<0.353$, $p>0.05$).
With each test run the quantity and the kind of stress signs of the dogs was recorded and correlated with the previous results of the test, as well as with Spearman's correlation coefficient. It was shown, that only the results of the dogs in subtest 3 with categorical rating ($r=0.510$, $p=0.0434$) and the total results of both ratings are rising with the increased number of signs of stress and therefore were positively correlated with each other.

The results of the study implies that the developed test is suitable to discover food related aggression of dogs against humans. It has proven reliable and valid despite of few exceptions and thereby it fulfills a great part of the claims of quality of a behavior test. Also the test has proven that it is practicable in his implementation and compliant with animal welfare. The scoring seems to be more suitable for usage in areas of diagnostic and therapy, where single dogs were tested and in context of a therapy compared to themselves. In contrast to that, the analysis with the categories seems to be more suitable for a usage in a shelter, where many dogs has to be tested and different observers have to assess and interpret the behavior quickly.

8 Literaturverzeichnis

ARCHER, J. (1988)
The behavioural biology of aggression. Cambridge studies in behavioural biology.
Cambridge University Press, Cambridge.

BARNARD, S., C. SIRACUSA, I. REISNER, P.VALSECCHI u. J. A. SERPELL (2012)
Validity of model devices used to assess canine temperament in behavioral tests.
Applied Animal Behaviour Science 138, 79–87.

BEAVER, B. V. (1983)
Clinical classification of canine aggression.
Applied Animal Ethology 10, 35–43.

BENNETT, S. L. (2010)
Temperament Tests: What we do and do not know about them.
Hrsg. v. Purdue Animal Behavior, Clinic Maddie's Animal Behavior Resident.

BENNETT, S. L., A. LITSTER, H.-Y. WENG, S. L. WALKER u. A. U. LUESCHER (2012)
Investigating behavior assessment instuments to predict aggression in dogs.
Applied Animal Behaviour Science 141, 139–148.

BERGLER, R. (2000)
Gesund durch Heimtiere: Beiträge zur Prävention und Therapie gesundheitlicher und seelischer Risikofaktoren.
Deutscher Instituts-Verlag, Köln.

BERNAUER-MÜNZ, H. u. C. QUANDT (1995)
Problemverhalten beim Hund: Lösungswege für den Tierarzt.
VET spezial-Reihe.
VEB Gustav Fischer Verlag, Jena, Stuttgart.

BLACKSHAW, J. K. (1985)
Human and Animal Interrelationships.4. Behavioral-Problems of Dogs. 1.
Australian Veterinary Practitioner 15 (3), 114-118.

BLACKSHAW, J. K. (1987)
Behavioural problems in dogs - some case studies.
Australian Veterinary Practitioner 17, 132-135.

BOLLEN, K. S. u. J. HOROWITZ (2008)
Behavioral evaluation and demographic information in the assessment of aggressiveness in shelter dogs.
Applied Animal Behaviour Science 112, 120–135.

BORCHELT, P. L. (1983)
Aggressive behavior of dogs kept as companion animals: classification and influence of sex, reproductive status and breed.
Applied Animal Ethology 10, 45–61.

BREUER, U. (2000)
Somatische Ursachen als Auslöser für Verhaltensprobleme und Verhaltensstörungen bei Hund und Katze.
Tierärztliche Umschau 55, 69–72.

BÜHNER, M. (2011)
Einführung in die Test- und Fragebogenkonstruktion. 5. Auflage.
Naturbuch Verlag, Augsburg.

DE MEESTER, R. H., D. DE BAQUER, K. PEREMANS, S. VERMEIRE, D. J. PLANTA, F. COOPMAN u. K. AUDENAERT (2008)
A preliminary study on the use of the Socially Acceptable Behavior test as a test for shyness/confidence in the temperament of dogs.
Journal of Veterinary Behavior 3, 161–170.

DEMEESTER, R. H. u. D. J. U. PLANTA (2007)
Validity of the Socially Acceptable Behavior (SAB) test as a measure of aggression in dogs towards non-familiar humans.
Vlaams Diergeneeskundig Tijdschrift, 76 (5), 359-368.

DIEDERICH, C. u. J. M. GIFFROY (2006)
Behavioural testing in dogs: A review of methodology in search for standardisation.
Applied Animal Behaviour Science 97, 51–72.

DIESEL, G., D. BRODBELT u. D. U. PFEIFFER (2008)
Reliability of assessment of dogs behavioural responses by staff working at a welfare charity in the UK.
Applied Animal Behaviour Science 115, 171–181.

DONALDSON, J. (2002)
Mine! A Practical Guide to Resource Guarding in Dogs.
Dogwise Publishing, Wenatchee.

DOWLING-GUYER, S., A. MARDER u. S. D'ARPINO (2011)
Behavioural traits detected in shelter dogs by a behavior evaluation.
Applied Animal Behaviour Science 130, 107–114.

EISFELD, D. (1966)
Verhaltensbeobachtungen an einigen Wildcaniden.
Zeitschrift für wissenschaftliche Zoologie 174, 227-289.

EL-GHOUL, W. u. W. HOFFMANN (2002)
Einfluss von Klauenkrankheiten verschiedenen Grades auf die Höhe der messbaren Stressreaktionen unter besonderer Berücksichtigung von Cortisol und Laktat im Blutserum beim Rind.
Praktischer Tierarzt 83 (4), 354-361.

ERLER, R. (2010)
Untersuchungen zur Stressbelastung von Wildschweinen bei der Ausbildung von Hunden zur Verhaltensanpassung im Schwarzwildgatter.
Tierärztliche Hochschule, Hannover, Diss.

FATJO, J., J. L. RUIZ-DE-LA-TORRE u. X. MANTECA (2006)
The epidemiology of behavioural problems in dogs and cats: a survey of veterinary practitioners.
Animal Welfare 15 (2), 179-185

FEDDERSEN, D. (1978)
Ausdrucksverhalten und soziale Organisation von Goldschakalen, Zwergpudeln und deren Gefangenschaftsbastarden.
Tierärztliche Hochschule, Hannover, Diss.

FEDDERSEN-PETERSEN, D. (1993)
Genesen des Aggressionsverhaltens von Hunden.
Collegium veterinarium XXIV, 104-108.

FEDDERSEN-PETERSEN, D. u. F. OHL (1995)
Ausdrucksverhalten beim Hund.
VEB Gustav Fischer Verlag, Jena, Stuttgart.

FEDDERSEN-PETERSEN, D. (2004)
Hundepsychologie. Sozialverhalten und Wesen, Emotionen und Individualität. 1. Aufl.
Franckh-Kosmos Verlag, Stuttgart.

FEDDERSEN-PETERSEN, D. (2008)
Ausdrucksverhalten beim Hund. Mimik und Körpersprache, Kommunikation und Verständigung.
Franckh-Kosmos Verlag, Stuttgart.

GATTERMANN, R. (Hrsg.) (1993)
Verhaltensbiologie.
VEB Gustav Fischer Verlag, Jena.

GUY, N. C., U. A. LUESCHER, S. E. DOHOO, E. SPANGLER, J. B. MILLER, I. R. DOHOO u. L. A. BATE (2001)
Demographic and aggressive characteristics of dogs in a general veterinary caseload.
Applied Animal Behaviour Science 74, 15–28.

HART, B. L. (1980)
Canine behavior: a practitioner monograph.
Veterinary Practice Publishing Company, Santa Barbara.

HART, B. L. u. L. A. HART (1985)
Selecting pet dogs on the basis of cluster analysis of breed behavior profiles and gender.
Journal of the American Veterinary Medical Association 186 (11), 1181-1185.

HIRSCHFELD, J. (2005)
Untersuchung einer Bullterrier-Zuchtlinie auf Hypertrophie des Aggressionsverhaltens.
Tierärztliche Hochschule, Hannover, Diss.

HOUPT, K. A. (1979)
Aggression in dogs.
The compendium on continuing education for the small animal practitioner 1, 123-128.

HSU, Y. u. J. A. SERPELL (2003)
Development and validation of a questionnaire for measuring behavior and temperament traits in pet dogs.
Journal of the American Veterinary Medical Association 123 (9), 1293-1300.

HUNSBERGER, R. W. (1956)
Affektreaktionen auf elektrische Reizung im Hirnstamm der Katze.
Helvetica Physiol Pharmacol Acta 14, 70-92

IMMELMANN, K. (1982)
Wörterbuch der Verhaltensforschung.
Verlag Paul Parey, Berlin, Hamburg.

JONES, R. (2009)
Aggression bei Hunden. Von Besitzanspruch bis Drohverhalten.
Franckh-Kosmos-Verlag, Stuttgart.

JONES, A. C. u. S. D. GOSLING (2005)
Temperament and personality in dogs (Canis familiaris): A review and evaluation of past research.
Applied Animal Behaviour Science 95, 1–53.

KAHN, A., P. BAUCHE u. J. LAMOUREUX (2003)
Child victims of dog bites treated in emergency departments: a prospective survey.
European Journal of Pediatrics 162, 254–258.

KROLL, T. L., K. A. HOUPT u. H. N. ERB (2004)
The Use of Novel Stimuli as Indicators of Aggressive Behavior in Dogs.
Journal of the American Animal Hospital Association 40, 13–19.

LANDIS, J. R. u. G. G. KOCH (1977)
The measurement of observer agreement for categorical data.
Biometrics 33, 159–174.

LANDSBERG, G. M., W. HUNTHAUSEN u. L. ACKERMANN (2003)
Handbook of behaviour problems of the dog and cat.
Sauders. Edinburgh, London.

LEDGER, R. A. u. M. R. BAXTER (1997)
The Development Of A Validated Test To Assess The Temperament Of Dogs In A Rescue Shelter.
In: Proceedings of the First International Conference on Veterinary Behavioral Medicine.

LINDSAY, S. R. (2000)
Handbook of applied dog behaviour and training, Volume one: Adaption and learning.
Iowa State University Press, Ames.

LINDSAY, S. R. (2001)
Handbook of applied dog behaviour and training, Volume two: Etiology and assessment of behaviour problems.
Iowa State University Press, Ames.

LINDSAY, S. R. (2005)
Handbook of applied dog behaviour and training, Volume three: Procedures and protocols.
Blackwell Publishing, Ames.

MARDER, A. R., A. SHABELANSKY, G. J. PATRONEK, S. DOWLING-GUYER u. S. S. D'ARPINO (2013)
Food-related aggression in shelter dogs: A comparison of behavior identified by a behavior evaluation in the shelter and owner reports after adoption.
Applied Animal Behaviour Science 184, 150–156.

MARTIN, P. u. P. BATESON (2007)
Measuring Behaviour. Third Edition. 6. Aufl.
Cambridge University Press, Cambridge.

MAZUR, M. (2012)
Entwicklung und Evaluierung eines speziellen Verhaltenstests zur Untersuchung der rassetypischen Verhaltensweisen von Hunden der Rasse Australian Shepherd.
Tierärztliche Hochschule, Hannover, Diss.

MCGREEVY, P. D., A. M. MASTERS (2008)
Risk factors for seperation-related distress and feed-related aggression in dogs: Additional findings from a survey of Australian dog owners.
Applied Animal Behaviour Science 109, 320–328.

MECH, L. D. (1970)
The Wolf: Ecology and Behavior of an Endangered Species.
University of Minnesota Press, Minneapolis.

MIKUS, R. (2006)
Statistische Auswertung von Sachverständigengutachten über Hunde mit Beißvorfällen in Bayern.
Ludwig-Maximillians-Universität, München, Diss.

MOHAN-GIBBONS, H., E. WEISS u. M. SLATER (2012)
Preliminary Investigation of Food Guarding Behavior in Shelter Dogs in the United States.
Animals 2, 331.

MURPHY, J. A. (1998)
Describing categories of temperament in potential guide dogs for the blind.
Applied Animal Behaviour Science 58, 163–178.

NEILSON, J. C. (2007)
How I treat food-related aggression in dogs.
Advanstar Communications Inc. Veterinary Medicine.
[http://veterinarymedicine.dvm360.com/vetmed/article/articleDetail.jsp?id=418393]

NETTO, W. J. u. D. J. U. PLANTA (1997)
Behavioural testing for aggression in the domestic dog.
Applied Animal Behaviour Science 52, 243–263.

NIEDERSÄCHSICHES MINISTERIUM FÜR ERNÄHRUNG, LANDWIRTSCHAFT UND FORSTEN (NMELF) (2000)
Niedersächsischer Wesenstest für Hunde
[http://www.ml.niedersachsen.de/download/2815]

OTT, S. (2010)
Entwicklung eines speziellen Verhaltenstests zur Evaluierung der Auswirkungen der Hund-Halter-Beziehung und Hund-Halter-Bindung auf das Verhalten des Hundes.
Tierärztliche Hochschule, Hannover, Diss.

PAROZ, C., S. G. GEBHARDT-HENRICH u. A. STEIGER (2008)
Reliability and validity of behaviour tests in Hovawart-dogs.
Applied Animal Behaviour Science 115, 67–81.

PLANTA, J. u. R. DE MEESTER (2007)
Validity of the Social Acceptable Behavior (SAB) test as a measure of aggression in dogs towards non-familiar humans.
Vlaams Diergeneeskundig Tijdschrift 76, 359–368.

ROST, J. (2004)
Lehrbuch Testtheorie-Testkonstruktion. 2. Auflage.
Huber Verlag, Bern.

RUGAAS, T. (2001)
Calming signals. Die Beschwichtigungssignale der Hunde.
Animal Learn Verlag, Grassau.

SCHENKEL, R. (1967)
Submission: Its Features and Functions in the Wolf and Dog.
Am. Zool. 7, 319-329.

SCHÖNING, B. u. K. RÖHRS (2013)
Hundesprache. Mimik und Körpersprache richtig deuten. 1. Aufl.
Franckh-Kosmos Verlag, Stuttgart.

SCHULZ, P., L.J. JANSEN u. W. SCHLOTZ (2005)
Stressreaktivität: Theoretisches Konzept und Messung.
Diagnostica 51 (3), 124-133.

SCOTT, J. P. u. E. FREDERICSON (1951)
The cause of fighting in mice and rats.
Psychological Zoology 24, 273-309.

SERPELL, J. A. (1995)
The domestic dog: its evolution, behavior, and interactions with people. 2. Origins of the dog: domestication and early history. 6. Aufl.
The Press Syndicate of the University of Cambridge, Cambridge.

SERPELL, J. A. u. Y. HSU (2001)
Development and validation of a novel method for evaluating behavior and temperament in guide dogs.
Applied Animal Behaviour Science 72, 347–364.

SLABBERT, J. M. u. J. S. J. ODENDAAL (1998)
Early prediction of adult police dog efficiency - a longitudinal study.
Applied Animal Behaviour Science 64, 269-288.

STERNBERG, S. (2002)
Great dog adoptions: A guide for shelters.
Latham Foundation, Alameda.

STUCKE, D. (2012)
Überprüfung der Anwendbarkeit der „Chronopsychobiologischen Regulationsdiagnostik“ (CRD) zur Beurteilung von Belastungssituationen und Bestimmung von Stressreaktionstypen bei Pferden.
Tierärztliche Hochschule, Hannover, Diss.

SVARTBERG, K., I. TAPPER, H. TEMRIN, T. RADESÄTER u. S. THORMAN (2005)
Consistency of personality traits in dogs.
Animal Behaviour 69, 283–291.

TAYLOR, K. D. u. D. S. MILLS (2006)
The development and assessment of temperament tests for adult companion dogs.
Journal of Veterinary Behavior 1, 94–108.

TIEFENBACH, P. (2001)
Untersuchungen über die Häufigkeitsverteilung von Verhaltensproblemen bei Hunden und Katzen.
Ludwig-Maximillians-Universität, München, Diss.

TEMBROCK, G. (1983)
Spezielle Verhaltensbiologie der Tiere, Band 2.
VEB Gustav Fischer Verlag, Jena.

TEMBROCK, G. (1987)
Verhaltensbiologie.
Wissenschaftlicher Verlag Harri Deutsch GmbH, Frankfurt am Main.

TEMBROCK, G. (1992)
Verhaltensbiologie.
VEB Gustav Fischer Verlag, Jena.

VALSECCHI, P., S. BARNARD, C. STEFANINI u. S. NORMANDO (2011)
Temperament test for re-homed dogs validated through direct behavioral observation in shelter and home environment.
Journal of Veterinary Behavior 6, 161–177.

VAN DEN BERG, L., M. B. H. SCHILDER u. B. W. KNOL (2003)
Behavior Genetics of Canine Aggression: Behavioral Phenotyping of Golden Retrievers by Means of an Aggression Test.

Behavior Genetics 33 (5), 469–483.

VAN DER BORG, J. A. M., W. J. NETTO u. D. J. U. PLANTA (1991)
Behavioural testing of dogs in animal shelters to predict problem behaviour.
Applied Animal Behaviour Science 32, 237–251.

VAN DER BORG, J., B. BEERDA, M. OOMS, A. S. DE SOUZA, M. VAN HAGEN u. B. KEMP (2010)
Evaluation of behaviour testing for human directed aggression in dogs.
Applied Animal Behaviour Science 128, 78–90.

VERBAND FÜR DAS DEUTSCHE HUNDEWESEN (VDH) e.V. (2013)
Geschäftsbericht zum Jahr 2012.
Dortmund.

VOITH, V. L. u. P. L. BORCHELT (1982)
Diagnosis and treatment of dominance aggression in dogs.
Veterinary Clinics of North America: Small Animal Practice 12, 655–663.

VON HOLST, E. u. U. VON SAINT PAUL (1960)
Vom Wirkungsgefüge der Triebe.
Naturwissenschaften 47, 409-422.

WEISS, E. (2007)
Meet Your Match: SAFER Manual and Training Guide.
ASPCA.

WEISS, E. u. G. GREENBERG (1997)
Service dog selection tests: Effectiveness for dogs from animal shelters.
Applied Animal Behaviour Science 53, 297–308.

WILSSON, E. u. P.-E. SUNDGREN (1997a)
The use of a behaviour test for the selection of dogs for service and breeding, I: Method of testing and evaluating test results in the adult dog, demands on different kinds of service dogs, sex and breed differences.
Applied Animal Behaviour Science 53, 279–295.

WILSSON, E. u. P.-E. SUNDGREN (1997b)
The use of a behaviour test for selection of dogs for service and breeding. II. Heritability for tested parameters and effect of selection based on service dog characteristics.
Applied Animal Behaviour Science 54, 235–241.

WRIGHT, J. C. (1996)
Canine Aggression: Dog Bites to People.
In: VOITH, V. L. u. P. L. BORCHELT (Hrsg.): Readings in Companion Animal Behavior.
Veterinary Learning Systems, Trenton.

WRIGHT, J. C. (1999)
The Dog who would be King. Chapter 6: Dog (Canis familiaris) Behavior.
Rodale Press, Emmaus

ZIEMEN, E. (1971)
Wölfe und Königspudel.
Piper Verlag, München.

ZIMEN, E. (1990)
Der Wolf : Verhalten, Ökologie und Mythos.
Knesebeck GmbH & Co. Verlags KG, München.

9 Anhang

9.1 Datenerhebungsbogen

Angaben zum Hund

Name: _____ Studien-Nr.: _____

Zwinger-Nr.: _____ Chip-Nr.: _____

Aufnahmedatum ins Tierheim: _____(TT.MM.JJ) Rasse: _____

Alter: _____Wochen/Monate/Jahre Geschlecht: O männlich O weiblich

Kastriert: O ja O nein im Alter von: _____ Wochen/Monate/Jahre

Vorbericht des Besitzers (soweit bekannt)

Anzahl Vorbesitzer: _____

Art der Haltung beim Vorbesitzer: O im Haus O im Garten O im Zwinger O _____

O einzeln O mit anderen Hunden

Abgabegrund: _____

Vorbericht des Tierheims *(Einschätzung durch zuständige Tierpflegerin)*

zuständige(r) Tierpfleger(in): _____

Fütterungsintervall: O 1x am Tag O 2x am Tag O 3x am Tag O > 3x am Tag

Fütterungszeiten: _____Uhr, _____ Uhr, _____Uhr, _____Uhr, _____Uhr, _____Uhr

Futtermittel: O Trockenfutter, der Marke(n) _____

O Nassfutter, der Marke(n) _____

O Futterbelohnungen, und zwar _____

Appetit: O gut (frisst alles sofort hektisch auf)

O durchschnittlich (frisst alles auf, weder hektisch noch mit Pausen)

O nicht so gut (frisst nur Lieblingsfutter, mit Pausen oder lässt Reste im Napf)

bekannte Krankheiten: _____

Medikamente: _____

bekannte Verhaltensprobleme (außer Aggression): _____

_____

Beschreibung der Verhaltensprobleme: _____

_____

Häufigkeit und Dauer des Problemverhaltens: _____

Futteraggression: O nein

O ja, Situation: O Futter/Napf bereitstellen O Futter/Napf wegnehmen

Futteraggression bei:	Futteraggression gegenüber:	Beobachtete Verhaltensweisen:
O leerem Futternapf	O bekannter Person Geschlecht: O männlich O weiblich Alter: _____Jahre	*AGONISTIK* O Abwenden O Festhalten O Abschirmen O Wegtragen O Erstarren O Drohfixieren O Knurren O Zähneblecken O Schnappen O Beißen O _____ *STRESSSIGNALE* O Gähnen O Pföteln O Maullecken O Winseln O Schütteln O _____
	O fremder Person Geschlecht: O männlich O weiblich Alter: _____Jahre	*AGONISTIK* O Abwenden O Festhalten O Abschirmen O Wegtragen O Erstarren O Drohfixieren O Knurren O Zähneblecken O Schnappen O Beißen O _____ *STRESSSIGNALE* O Gähnen O Pföteln O Maullecken O Winseln O Schütteln O _____
O normalem Futter	O bekannter Person Geschlecht: O männlich O weiblich Alter: _____Jahre	*AGONISTIK* O Abwenden O Beschl. Verzehr O Festhalten O Abschirmen O Wegtragen O Erstarren O Drohfixieren O Knurren O Zähneblecken O Schnappen O Beißen O _____ *STRESSSIGNALE* O Gähnen O Pföteln O Maullecken O Winseln O Schütteln O _____
	O fremder Person	*AGONISTIK* O Abwenden O Beschl. Verzehr O Festhalten O Abschirmen O Wegtragen O Erstarren

	Geschlecht: O männlich O weiblich Alter: _____Jahre	O Drohfixieren O Knurren O Zähneblecken O Schnappen O Beißen O _____ *STRESSSIGNALE* O Gähnen O Pföteln O Maullecken O Winseln O Schütteln O _____
O besonderem Futter (z.B. Kauartikel)	O bekannter Person Geschlecht: O männlich O weiblich Alter: _____Jahre	*AGONISTIK* O Abwenden O Beschl. Verzehr O Festhalten O Abschirmen O Wegtragen O Erstarren O Drohfixieren O Knurren O Zähneblecken O Schnappen O Beißen O _____ *STRESSSIGNALE* O Gähnen O Pföteln O Maullecken O Winseln O Schütteln O _____
	O fremder Person Geschlecht: O männlich O weiblich Alter: _____Jahre	*AGONISTIK* O Abwenden O Beschl. Verzehr O Festhalten O Abschirmen O Wegtragen O Erstarren O Drohfixieren O Knurren O Zähneblecken O Schnappen O Beißen O _____ *STRESSSIGNALE* O Gähnen O Pföteln O Maullecken O Winseln O Schütteln O _____

Andere Aggressionen (außer Futteraggression): O nein O ja

Aggression gegenüber:	Beobachtete Verhaltensweisen:	Situation und gesetzte Verletzungen:
O bekannter Person Geschlecht: O männlich O weiblich Alter: _____Jahre	O Erstarren O Drohfixieren O Fellsträuben O Knurren O Zähneblecken O Schnappen O Beißen O _____	Situation: Gesetzte Verletzungen:
O fremder Person Geschlecht:	O Erstarren O Drohfixieren O Fellsträuben O Knurren	Situation:

O männlich O weiblich Alter: _____Jahre	O Zähneblecken O Schnappen O Beißen O _____	Gesetzte Verletzungen:

Furcht vor Personen:

O ja, der Hund weicht mit geduckter Körperhaltung oder eingeklemmter Rute vor einer fremden oder bekannten Person aus und nähert sich nach freundlicher Ansprache (hinhocken, wegdrehen, ruhige Stimme) durch die Testperson nicht innerhalb von 30 Sekunden an

O nein

Anmerkungen:

Auswertung

Testzulassung: O ja

O nein, wegen: O Aggression O Furcht O relevante Krankheit/Verhaltensstörung

9.2 Ethogramm (Test auf Futteraggression)

Agonistisches Verhalten	
Abwenden (mod. nach FEDDERSEN, 1978)	Der Hund bewegt seinen Kopf oder seinen Vorderkörper *ohne das Objekt* (z.B. Futter, Rinderpansen) von einem auslösenden Reiz (z.B. Testperson, künstliche Hand) weg, befindet sich aber (ab Testsituation 6.) noch in Reichweite der Testperson bzw. der künstlichen Hand.
Beschleunigter Verzehr (mod. nach DONALDSON, 2002; VALSECCHI et al., 2011)	Der Hund nimmt in 3 sec öfter, größere und/oder hektischer Futterhappen zu sich als in 3 sec ungestörtem Fressen.
Festhalten (mod. nach VALSECCHI et al., 2011)	Der Hund klemmt ein Objekt (z.B. Futtermittel) zwischen den Vordergliedmaßen oder zwischen den Kiefern ein oder stellt eine oder beide Vordergliedmaßen auf oder in das Objekt, sodass die Testperson das Objekt nicht mit der künstlichen Hand aufnehmen oder anderweitig entfernen kann; dabei bewegt sich der Hund *nicht* von der Testperson weg.
Abschirmen (mod. nach VALSECCHI et al., 2011)	Der Hund bewegt mindestens seinen Kopf zwischen ein Objekt (z.B. Futtermittel) und einen auslösenden Reiz (z.B. Testperson, künstliche Hand) oder hält mindestens den Kopf über das Objekt oder *nimmt das Objekt ins Maul* und dreht seinen Kopf oder seinen Körper vom auslösenden Reiz weg, bleibt dabei aber *in Reichweite* der Testperson, bzw. der künstlichen Hand; außer bei soziopositiver Kontaktaufnahme zur Testperson (der Hund wendet sich mit entspannter Körperhaltung direkt der Testperson zu, blickt diese an, verringert die Distanz zu ihr und wedelt eventuell leicht (mod. nach Umlauf, 1993)).

Wegtragen (mod. nach ZIMEN, 1971, 1990)	Der Hund nimmt ein Objekt (z.B. Futtermittel) ins Maul und bewegt sich *damit* bis *außerhalb der Reichweite* von einem Reiz (z.B. Testperson, künstliche Hand) weg.
Erstarren (mod. nach DONALDSON, 2002; VAN DEN BERG et al., 2003)	Der Hund hält in der Bewegung inne und verharrt regungslos.
Drohfixieren (nach EISFELD, 1966; VAN DEN BERG et al., 2003, VALSECCHI et al., 2011)	Der Hund richtet seinen Blick starr in die Augen eines Gegners (z.B. Testperson), oft in Verbindung mit Erstarren.
Knurren (nach FEDDERSEN-PETERSEN 2004; VAN DEN BERG et al., 2003)	Der Hund gibt einen atonalen, tief-frequenten Kehllaut von sich.
Zähneblecken (nach ZIMEN, 1971, 1990; VAN DEN BERG et al., 2003)	Der Hund zieht die Ober- bzw. die Unterlippe nach oben bzw. nach unten, die Zähne und das Zahnfleisch sind mehr oder weniger sichtbar, der Nasenrücken ist gerunzelt.
Schnappen (mod. nach ZIMEN, 1971, 1990; VAN DEN BERG et al., 2003)	Der Hund führt Beißbewegungen in der Luft aus, die gegen einen auslösenden Reiz (z.B. Testperon, künstliche Hand) gerichtet sind, dabei berühren die Zähne des Hundes nicht den auslösenden Reiz und das Aufeinanderschlagen der Zähne kann hörbar sein.
Beißen (mod. nach VAN DER BORG et al., 2010)	Der Hund führt Beißbewegungen gegen ein Objekt (z.B. künstliche Hand) aus, wobei es zum Kontakt des Objekts mit den Zähnen des Ober- und Unterkiefer des Hundes kommt; außer bei Spielbewegungen (der Hund führt übertriebene Bewegungen aus, um den Reiz (z.B. die Testperson, die

	künstliche Hand) mit der Schnauze anzustoßen, mit ausholenden Vorderbeinbewegungen zu packen, mit schnellen und ausholenden Bewegungen des Kopfes zu verfolgen oder mit den Zähnen zu beknabbern oder zu fangen und festzuhalten; der Schwanz ist aufgerichtet und wedelt eventuell, die Ohren sind aufgestellt und das Gesicht entspannt; vorher und dabei werden keine agonistischen Verhaltensweisen gezeigt (nach Feddersen-Petersen, 2008, S. 223)).
Stresssignale	
Gähnen (nach EISFELF, 1966; RUGAAS, 2001)	Der Hund öffnet das Maul weit und atmet dabei tief ein und kurz aus.
Pföteln (mod. nach VAN DEN BERG et al., 2003)	Der Hund hebt eine Vorderpfote und/oder legt sie einem Objekt (z.B. künstliche Hand) auf.
Maullecken (nach FEDDERSEN-PETERSEN, 2004; VAN DEN BERG et al., 2003)	Der Hund leckt über seine eigene Schnauze oder führt Intentionsbewegungen zum Lecken aus.
Winseln (nach FEDDERSEN-PETERSEN, 2008; VAN DEN BERG et al., 2003)	Der Hund gibt aneinandergereihte, hohe, langgezogene tonale Laute von sich.
Schütteln (mod. nach VAN DER BORG et al., 2010)	Der Hund führt schnelle rotierende Seitwärtsbewegungen mit dem Kopf und/oder dem Körper aus.
Anderes Verhalten	

Außerhalb Reichweite (mod. nach FEDDERSEN, 1978)	Der Hund *bewegt sich oder befindet sich* außerhalb der Reichweite der Testperson oder der künstlichen Hand.
Kein agonistisches Verhalten	Keine der oben beschriebenen, agonistischen Verhaltensweisen beobachtet.

9.3 Auswertungstabellen (Test auf Futteraggression)

9.3.1 Subtest 1

Hund: SUBTEST 1 Datum:

Verhalten/Testsituation	1	2	3	4	5	6	7	8	9	10	11	12	13
Agonist. Verhalten													
0 Abwenden													
2 Festhalten													
2 Abschirmen													
2 Wegtragen													
3 Erstarren													
4 Drohfixieren													
5 Knurren													
5 Zähneblecken													
6 Schnappen													
7 Beißen													
Kein agonist. Verhalt.													
Außerhalb Reichweite													
Stresssignale													
Gähnen													
Pföteln													
Maullecken													
Winseln													
Schütteln													
PUNKTE													
Gesamt Subtest 1													

9.3.2 Subtest 2

Hund: SUBTEST 2 Datum:

Verhalten/Testsituation	1	2	3	4	5	6	7	8	9	10	11	12	13
Agonist. Verhalten													
0 Abwenden													
1 Beschleun. Verzehr													
2 Festhalten													
2 Abschirmen													
2 Wegtragen													
3 Erstarren													
4 Drohfixieren													
5 Knurren													
5 Zähneblecken													
6 Schnappen													
7 Beißen													
Kein agonist. Verhalt.													
Außerhalb Reichweite													
Stresssignale													
Gähnen													
Pföteln													
Maullecken													
Winseln													
Schütteln													
PUNKTE													
Gesamt Subtest 2													

9.3.3 Subtest 3

Hund: SUBTEST 3 Datum:

Verhalten/Testsituation	1	2	3	4	5	6	7	8	9	10	11	12
Agonist. Verhalten												
0 Abwenden												
1 Beschleun. Verzehr												
2 Festhalten												
2 Abschirmen												
2 Wegtragen												
3 Erstarren												
4 Drohfixieren												
5 Knurren												
5 Zähneblecken												
6 Schnappen												
7 Beißen												

Kein agonist. Verhalt.												
Außerhalb Reichweite												
Stresssignale												
Gähnen												
Pföteln												
Maullecken												
Winseln												
Schütteln												
PUNKTE												
Gesamt Subtest 3												

9.4 Ethogramm (Prätest)

Soziopositive Verhaltensweisen	
Freundliche Annäherung (nach Umlauf, 1993; Eisfeld, 1966)	Der Hund wendet sich mit entspannter Körperhaltung (Schwanzhaltung locker herabhängend, evtl. wedelnd, Ohren nach vorne oder seitlich gerichtet) direkt dem auslösenden Reiz (z.B. Testperson, künstliche Hand) zu, blickt diese an und verringert die Distanz zu diesem.
Beschnüffeln (mod. nach EISFELD, 1966, ZIMEN, 1971, 1990, SCHÖNE, 2004)	Der Hund beriecht einen auslösenden Reiz (z.B. künstliche Hand) und stupst diesen eventuell mit der Schnauze an.
Belecken (mod. nach ZIMEN, 1971 und 1990)	Der Hund leckt mit der Zunge über Teile eines auslösenden Reizes (z.B. Testperson, künstliche Hand).
Spielbewegungen (nach Feddersen-Petersen, 2008)	Der Hund führt übertriebene Bewegungen aus, um einen Reiz (z.B. die Testperson, die künstliche Hand) mit der Schnauze anzustoßen, mit ausholenden Vorderbeinbewegungen zu packen, mit schnellen und ausholenden Bewegungen des Kopfes zu verfolgen oder mit den Zähnen zu beknabbern oder zu fangen und

	festzuhalten; der Schwanz ist aufgerichtet und wedelt eventuell, die Ohren sind aufgestellt und das Gesicht entspannt; vorher und dabei werden keine agonistischen Verhaltensweisen gezeigt.
Agonistisches Verhalten	
Abwenden (mod. nach FEDDERSEN, 1978)	Der Hund bewegt seinen Kopf oder seinen Vorderkörper *ohne das Objekt* (z.B. Futter, Rinderpansen) von einem auslösenden Reiz (z.B. Testperson, künstliche Hand) weg, befindet sich aber (ab Testsituation 6.) noch in Reichweite der Testperson bzw. der künstlichen Hand.
Erstarren (mod. nach DONALDSON, 2002; VAN DEN BERG et al., 2003)	Der Hund hält in der Bewegung inne und verharrt regungslos.
Drohfixieren (nach EISFELD, 1966; VAN DEN BERG et al., 2003, VALSECCHI et al., 2011)	Der Hund richtet seinen Blick starr in die Augen eines Gegners (z.B. Testperson), oft in Verbindung mit Erstarren.
Knurren (nach FEDDERSEN-PETERSEN 2004; VAN DEN BERG et al., 2003)	Der Hund gibt einen atonalen, tief-frequenten Kehllaut von sich.
Zähneblecken (nach ZIMEN, 1971, 1990; VAN DEN BERG et al., 2003)	Der Hund zieht die Ober- bzw. die Unterlippe nach oben bzw. nach unten, die Zähne und das Zahnfleisch sind mehr oder weniger sichtbar, der Nasenrücken ist gerunzelt.

Schnappen (mod. nach ZIMEN, 1971, 1990; VAN DEN BERG et al., 2003)	Der Hund führt Beißbewegungen in der Luft aus, die gegen einen auslösenden Reiz (z.B. Testperson, künstliche Hand) gerichtet sind, dabei berühren die Zähne des Hundes nicht den auslösenden Reiz und das Aufeinanderschlagen der Zähne kann hörbar sein.
Beißen (mod. nach VAN DER BORG et al., 2010)	Der Hund führt Beißbewegungen gegen ein Objekt (z.B. künstliche Hand) aus, wobei es zum Kontakt des Objekts mit den Zähnen des Ober- und Unterkiefer des Hundes kommt; außer bei Spielbewegungen (der Hund führt übertriebene Bewegungen aus, um den Reiz (z.B. die Testperson, die künstliche Hand) mit der Schnauze anzustoßen, mit ausholenden Vorderbeinbewegungen zu packen, mit schnellen und ausholenden Bewegungen des Kopfes zu verfolgen oder mit den Zähnen zu beknabbern oder zu fangen und festzuhalten; der Schwanz ist aufgerichtet und wedelt eventuell, die Ohren sind aufgestellt und das Gesicht entspannt; vorher und dabei werden keine agonistischen Verhaltensweisen gezeigt (nach Feddersen-Petersen, 2008, S. 223)).
Stresssignale	
Gähnen (nach EISFELF, 1966; RUGAAS, 2001)	Der Hund öffnet das Maul weit und atmet dabei tief ein und kurz aus.
Pföteln (mod. nach VAN DEN BERG et al., 2003)	Der Hund hebt eine Vorderpfote und/oder legt sie einem Objekt (z.B. künstliche Hand) auf.
Maullecken (nach FEDDERSEN-PETERSEN, 2004; VAN DEN BERG et al., 2003)	Der Hund leckt über seine eigene Schnauze oder führt Intentionsbewegungen zum Lecken aus.

Winseln (nach FEDDERSEN-PETERSEN, 2008; VAN DEN BERG et al., 2003)	Der Hund gibt aneinandergereihte, hohe, langgezogene tonale Laute von sich.
Schütteln (mod. nach VAN DER BORG et al., 2010)	Der Hund führt schnelle rotierende Seitwärtsbewegungen mit dem Kopf und/oder dem Körper aus.
Anderes Verhalten	
Außerhalb Reichweite (mod. nach FEDDERSEN, 1978)	Der Hund *bewegt sich oder befindet sich* außerhalb der Reichweite der Testperson oder der künstlichen Hand.
Kein agonistisches Verhalten	Keine der oben beschriebenen, agonistischen Verhaltensweisen beobachtet.

9.5 Auswertungstabellen (Prätest)

9.5.1 Prätest 1

Hund: PRÄTEST 1 Datum:

Verhalten/Testsituation	1	2	3	4	5	6	7	8	9	10
Sozioposit. Verhalten										
Freundl. Annäherung										
Beschnüffeln										
Belecken										
Spielbewegungen										
Agonist. Verhalten										
0 Abwenden										
3 Erstarren										
4 Drohfixieren										
5 Knurren										
5 Zähneblecken										
6 Schnappen										
7 Beißen										
Kein agon./soziop. V.										

Außerhalb Reichweite										
Stresssignale										
Gähnen										
Pföteln										
Maullecken										
Winseln										
Schütteln										
PUNKTE										
Gesamt Prätest 1										

9.5.2 Prätest 2

Hund: PRÄTEST 2 Datum:

Verhalten/Testsituation	1	2	3	4	5	6	7	8	9	10
Sozioposit. Verhalten										
Freundl. Annäherung										
Beschnüffeln										
Belecken										
Spielbewegungen										
Agonist. Verhalten										
0 Abwenden										
3 Erstarren										
4 Drohfixieren										
5 Knurren										
5 Zähneblecken										
6 Schnappen										
7 Beißen										
Kein agon./soziop. V.										
Außerhalb Reichweite										
Stresssignale										
Gähnen										
Pföteln										
Maullecken										
Winseln										
Schütteln										
PUNKTE										
Gesamt Prätest 2										

DANKSAGUNG

Ich danke meinen Eltern für ihre große Unterstützung und den Mut, den sie mir immer wieder gemacht haben.

Ich danke meinem Mann Tino für seine Geduld, seine guten Ideen, seine Aufmunterungen und die Zeit, die er sich für mich genommen hat.

Ich danke Herrn Prof. Hackbarth für seine Anregungen und Hilfe bei der Umsetzung sowie bei der Korrektur der Arbeit.

Ich danke Frau Dr. Moesta für ihre fachliche Weisung und Unterstützung bei der Planung sowie beim Schreiben der Arbeit.

Ich danke Frau Schürmann für ihre Teilnahme an der Studie, ihre große Unterstützung bei der Durchführung der Tests und die Zeit, die sie sich für mich genommen hat. Außerdem bedanke ich mich bei ihren Hunden, welche an der Studie teilgenommen haben.

Ich danke Frau Dr. Thiesen-Moussa für ihre Auswertung der Videos und die Zeit, die sie sich genommen hat.

Ich danke Herrn Dr. Rohn für die Hilfe bei den statistischen Berechnungen.

Ich danke weiterhin Frau Zimmermann, Frau Dr. Bohnet und allen Mitarbeitern des Instituts für Ihre Unterstützung und freundlichen Worte, die sie für mich hatten.